2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

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

Prepared for the 2012 Methodological Resource Book

RTI Project No. 0212800.001.107.005
Contract No. HHSS283201000003C
Deliverable No. 41

Authors:

Matthew Westlake
Patrick Chen
Harper Gordek

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

February 2014

DISCLAIMER

SAMHSA provides links to other Internet sites as a service to its users and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

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

Prepared for the 2012 Methodological Resource Book

RTI Project No. 0212800.001.107.005
Contract No. HHSS283201000003C
Deliverable No. 41

Authors:

Matthew Westlake
Patrick Chen
Harper Gordek

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

February 2014

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 on Drug Use and Health (NSDUH) data from 2012. 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, which 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. (2014), 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. Census Bureau 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 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 on estimation in the presence of multiplicities and extreme weights (Chromy & Singh, 2001) and on GEM calibration of pair weights (Penne, Chen, & Singh, 2001). 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), Center for Behavioral Health Statistics and Quality (CBHSQ), by RTI International (a trade name of Research Triangle Institute), Research Triangle Park, North Carolina, under Contract No. HHSS283201000003C. The authors would like to take this opportunity to thank a number of individuals for useful discussions and suggestions: Joe Gfroerer and Art Hughes of SAMHSA and Jim Chromy 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**

List of Terms and Abbreviations……..xi

1. Introduction 1

2. Questionnaire Dwelling Unit and Pair Selection Probabilities 7
   - 2.1 Pair Selection Probability 8
     - 2.1.1 Case I: DUs with $S \geq 2$ 8
     - 2.1.2 Case II: DUs with $S < 2$ 8
   - 2.2 Questionnaire Dwelling Unit Selection Probability 10

3. Brief Description of the Generalized Exponential Model 13

4. Predictor Variables for the Questionnaire Dwelling Unit and Pair Weight Calibration via the Generalized Exponential Model 15
   - 4.1 Questionnaire Dwelling Unit Weight Calibration 15
   - 4.2 Pair Weight Calibration 16

5. Definition of Extreme Weights 21
   - 5.1 Questionnaire Dwelling Unit Extreme Weight Definition 21
   - 5.2 Person Pair Extreme Weight Definition 22

6. Weight Calibration at Questionnaire Dwelling Unit and Pair Levels 25
   - 6.1 Phase I SDU-Level Weight Components 30
   - 6.2 QDU Weight Components 31
     - 6.2.1 QDU Weight Component #11: Inverse of Selection Probability of at Least One Person in the Dwelling Unit 31
     - 6.2.2 QDU Weight Component #12: Selected QDU Poststratification to SDU-Based Control Totals 31
     - 6.2.3 QDU Weight Component #13: Respondent QDU Nonresponse Adjustment 31
     - 6.2.4 QDU Weight Component #14: Respondent QDU Poststratification to SDU-Based Control Totals 32
     - 6.2.5 QDU Weight Component #15: Respondent QDU Extreme Value Adjustment 32
   - 6.3 Pair-Level Weight Components 32
     - 6.3.1 Pair Weight Component #11: Inverse of Selection Probability of a Person Pair in the Dwelling Unit 32
     - 6.3.2 Pair Weight Component #12: Selected Pair Poststratification to SDU-Based Control Totals 32
     - 6.3.3 Pair Weight Component #13: Respondent Pair Nonresponse Adjustment 32
     - 6.3.4 Pair Weight Component #14: Respondent Pair Poststratification to SDU-Based Control Totals 33
     - 6.3.5 Pair Weight Component #15: Respondent Pair Extreme Weight Adjustment 33

7. Evaluation of Calibration Weights 35
7.1 Response Rates 35
7.2 Proportions of Extreme Values and Outwinsors 35
7.3 Slippage Rates 36
7.4 Weight Adjustment Summary Statistics 36
7.5 Sensitivity Analysis of Drug Use Estimates 37
References 49

Appendix **Page**

A Technical Details about the Generalized Exponential Model A-1
B Derivation of Poststratification Control Totals B-1
C GEM Modeling Summary for the Questionnaire Dwelling Unit Weights C-1
C.1: Model Group 1: Northeast… C-11
C.2: Model Group 2: Midwest… C-19
C.3: Model Group 3: South… C-27
C.4: Model Group 4: West… C-35
D Evaluation of Calibration Weights: Questionnaire Dwelling Unit-Level Response Rates D-1
E Evaluation of Calibration Weights: Questionnaire Dwelling Unit-Level Proportions of Extreme Values and Outwinsors E-1
F Evaluation of Calibration Weights: Questionnaire Dwelling Unit-Level Slippage Rates… F-1
G Evaluation of Calibration Weights: Questionnaire Dwelling Unit-Level Weight Summary Statistics G-1
H GEM Modeling Summary for the Pair Weights H-1
H.1: Model Group 1: Northeast and South… H-5
H.2: Model Group 2: Midwest and West… H-13
I Evaluation of Calibration Weights: Pair-Level Response Rates I-1
J Evaluation of Calibration Weights: Pair-Level Proportions of Extreme Values and Outwinsors J-1
K Evaluation of Calibration Weights: Pair-Level Slippage Rates K-1
L Evaluation of Calibration Weights: Pair-Level Weight Summary Statistics L-1
M Pair Analysis Manual Excerpt M-1

List of Tables

Table		Page
1.1	2008–2012 NSDUH Sample Sizes	2
1.2	Pair Domains	3
2.1	Building Blocks of the QDU and Person Pair Samples: Dwelling Units and Persons in the 2008–2012 NSDUHs	11
6.1	Sample Size, by Model Group at QDU and Pair Levels	30
7.1	Estimates of Totals and SEs for Domains of Interest Based on QDU Sample: 2012	38
7.2a	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: 2012	39
7.2b	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: 2012	40
7.3a	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: 2012	41
7.3b	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: 2012	42
7.4	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: 2012	43
7.5	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: 2012	44
7.6a	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: 2012	45
7.6b	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: 2012	46
7.7a	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: 2012	47
7.7b	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: 2012	48

C.1a 2012 QDU Weight GEM Modeling Summary (Model Group 1: Northeast) C-13
C.1b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Northeast) C-14
C.2a 2012 QDU Weight GEM Modeling Summary (Model Group 2: Midwest)..... C-21
C.2b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 2: Midwest)..... C-22
C.3a 2012 QDU Weight GEM Modeling Summary (Model Group 3: South) C-29
C.3b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 3: South)..... C-30
C.4a 2012 QDU Weight GEM Modeling Summary (Model Group 4: West) C-37
C.4b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 4: West)..... C-38
D.1 2012 NSDUH QDU-Level Response Rates D-1
E.1 2012 NSDUH Selected QDU-Level Proportions of Extreme Values and Outwinsors... ..... E-1
E.2 2012 NSDUH Respondent QDU-Level Proportions of Extreme Values and Outwinsors... ..... E-3
F.1 2012 NSDUH QDU-Level Slippage Rates..... F-1
G.1 2012 NSDUH Selected QDU-Level Weight Summary Statistics G-1
G.2 2012 NSDUH Respondent QDU-Level Weight Summary Statistics... ..... G-3
H.1a 2012 Pair Weight GEM Modeling Summary (Model Group 1: Northeast and South)..... H-7
H.1b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Northeast and South) H-8
H.2a 2012 Pair Weight GEM Modeling Summary (Model Group 2: Midwest and West)..... H-17
H.2b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 2: Midwest and West)..... H-18
I.1 2012 NSDUH Person Pair-Level Response Rates..... I-1
J.1 2012 NSDUH Selected Pair-Level Proportions of Extreme Values and Outwinsors... ..... J-1
J.2 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (res.pr.nr)... ..... J-3
J.3 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (res.pr.ps and res.pr.ev) J-5
K.1 2012 NSDUH Respondent Pair-Level Slippage Rates K-1
L.1 2012 NSDUH Selected Pair-Level Weight Summary Statistics..... L-1
L.2 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.nr) L-4
L.3 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.ps and res.pr.ev)..... L-6

List of Exhibits

Exhibit		Page
1.1	QDU and Pair Sampling Weight Calibration Steps	4
4.1	Definitions of Levels for QDU-Level Calibration Modeling Variables	17
4.2	Definitions of Levels for Pair-Level Calibration Modeling Variables	18
6.1	Summary of 2012 NSDUH QDU Sample Weight Components	27
6.2	Summary of 2012 NSDUH Person Pair Sample Weight Components	28
6.3	U.S. Census Bureau Regions/Model Groups	29
C.1	Definitions of Levels for QDU-Level Calibration Modeling Variables	C-2
C.2	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights	C-9
C.1.1	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 1: Northeast	C-16
C.1.2	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 1: Northeast	C-17
C.1.3	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 1: Northeast	C-18
C.2.1	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 2: Midwest	C-24
C.2.2	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 2: Midwest	C-25
C.2.3	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 2: Midwest	C-26
C.3.1	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 3: South	C-32
C.3.2	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 3: South	C-33
C.3.3	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 3: South	C-34
C.4.1	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 4: West	C-40
C.4.2	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 4: West	C-41
C.4.3	Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 4: West	C-42
H.1	Definitions of Levels for Pair-Level Calibration Modeling Variables	H-2
H.2	Covariates for 2012 NSDUH Pair Weights	H-4
H.1.1	Covariates for 2012 NSDUH Pair Weights (sel.pr.ps) Model Group 1: Northeast and South	H-10
H.1.2	Covariates for 2012 NSDUH Pair Weights (res.pr.nr) Model Group 1: Northeast and South	H-11
H.1.3	Covariates for 2012 NSDUH Pair Weights (res.pr.ps) Model Group 1: Northeast and South	H-12
H.1.4	Covariates for 2012 NSDUH Pair Weights (res.pr.ev) Model Group 1: Northeast and South	H-13

H.2.1 Covariates for 2012 NSDUH Pair Weights (sel.pr.ps) Model Group 2: Midwest and West..... H-20
H.2.2 Covariates for 2012 NSDUH Pair Weights (res.pr.nr) Model Group 2: Midwest and West..... H-21
H.2.3 Covariates for 2012 NSDUH Pair Weights (res.pr.ps) Model Group 2: Midwest and West..... H-22
H.2.4 Covariates for 2012 NSDUH Pair Weights (res.pr.ev) Model Group 2: Midwest and West..... H-23

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.
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 11 in the editing and imputation report (Carpenter et al., 2014) 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 Carpenter et al. (2014) 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).
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 6.3.3 for more detail.
res.qdu.nr	Respondent questionnaire dwelling unit nonresponse adjustment step. See Section 6.2.3 for more detail.
res.pr.ev	Respondent pair extreme value adjustment step. See Section 6.3.5 for more detail.
res.qdu.ev	Respondent questionnaire dwelling unit extreme value adjustment step. See Section 6.2.5 for more detail.

res.pr.ps	Respondent pair poststratification adjustment step. See Section 6.3.4 for more detail.
res.qdu.ps	Respondent questionnaire dwelling unit poststratification adjustment step. See Section 6.2.4 for more detail.
SDU	Screener dwelling unit: a household where screener information is available.
sel.pr.ps	Selected person pair poststratification adjustment step. See Section 6.3.2 for more detail.
sel.qdu.ps	Selected questionnaire dwelling unit poststratification adjustment step. See Section 6.2.2 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 generally avoided since persons from the same household often exhibit the same or similar characteristics and behavioral patterns. The intra-class correlation found among members of the same household leads to a clustering effect on the variance of estimates resulting in less precise estimates compared with estimates of the same sample size from a simple random sample. Selecting only one person per household avoids this 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 67,500 persons selected from the civilian, noninstitutionalized 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 42 smaller States and the District of Columbia were partitioned into 12 State sampling (SS) regions, whereas the eight largest 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 within census tracts, and dwelling units (DUs) within segments were each selected using probability proportional to size sampling. NSDUH is sometimes referred to as a two-phase sample where the first phase consisted of a large number of screener dwelling units (SDUs, about 200,000) 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 second phase. 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. Census Bureau-based controls were not available. The second phase consisted of

[2] This report presents information from the 2012 National Survey on Drug Use and Health (NSDUH). Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

the selection of zero, one, or two persons from each selected SDU using a modification of Brewer's method such that prescribed sampling rates for the five age groups in 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, pairs, and persons for each of the 5 years (2008–2012). 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 adopted solutions.

Table 1.1 2008–2012 NSDUH Sample Sizes

Sample Unit		2008	2009	2010	2011	2012
SDU	Selected	160,114	161,377	166,532	179,293	178,586
	Completed	142,159	142,933	147,010	156,048	153,873
QDU	Selected	58,942	58,288	58,702	61,441	60,621
	Completed	48,180	48,088	48,113	50,133	48,850
Pair	Selected	26,769	26,497	26,295	27,095	27,035
	Completed	19,748	19,919	19,691	19,976	19,459
Person	Selected	85,711	84,785	84,997	88,536	87,656
	Completed	67,928	68,007	67,804	70,109	68,309

Note: The 2008–2010 sample sizes reflect the removal of falsified cases found in Pennsylvania and Maryland. The 2011 and 2012 samples were not affected. For additional information, see Section B.3.5 in Appendix B of the 2011 and 2012 national findings reports (Center for Behavioral Health Statistics and Quality, 2012; 2013).

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, For example, analysts are interested in an outcome at the person level, the proportion of children who use drugs and whose parents report talking to them about drugs, where the focus is on the child in a parent-child pair. Several parent-child pairs in the household could be associated with the same child. 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 (Carpenter et al., 2014). 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

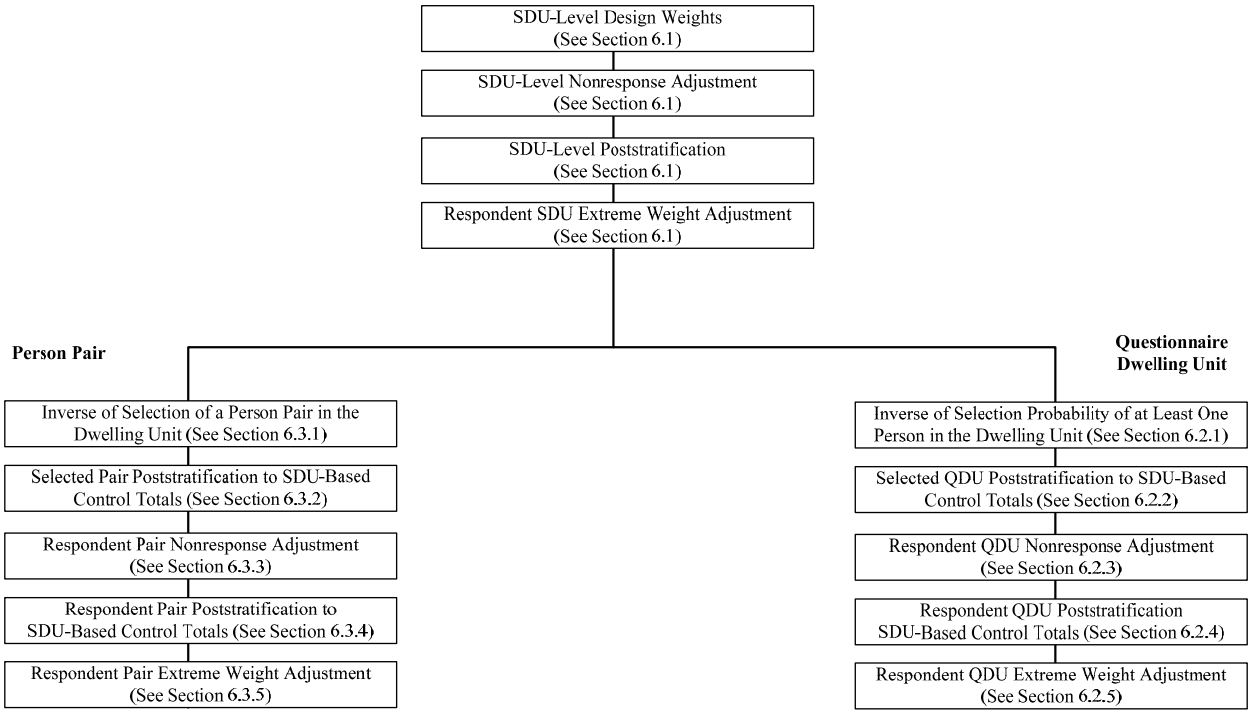
Pair Relationship	Focus
Parent-child: parent, child aged 12-14	Parent
Parent-child: parent, child aged 12-14	Child
Parent-child: parent, child aged 12-17	Parent
Parent-child: parent, child aged 12-17	Child
Parent-child: parent, child aged 12-20	Parent
Parent-child: parent, child aged 12-20	Child
Parent-child: parent, child aged 15-17	Parent
Parent-child: parent, 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 or partner-partner, with or without children	No multiplicity necessary
Spouse-spouse or 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. Prior editions of this report contained a chapter on editing and imputation of pair relationships, multiplicity factors, and household-level person counts for poststratification, but this information has been removed in the interest of consolidating the imputation documentation and can now be found in the editing and imputation report (Carpenter et al., 2014). Special consideration is required for analysis of pair-level data, and details can be found in *How To Prepare and Analyze Pair Data in the National Survey on Drug Use and Health* (Frechtel, Warren, & Porter, in press). See Appendix M for an excerpt from this manual on inferential population and multiplicities.

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 paragraphs.

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 make the sample representative of the 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 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 (e.g., 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 2012 survey was the same as the one used in the 1999–2011 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.

[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.

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

[5] Pair age in this case should not 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 2002–2011 National Surveys on Drug Use and Health (NSDUHs),[6] the 2012 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, Martin, Shook-Sa, Chromy, & Hirsch, 2013).

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 three dummy persons so that the sum of the person probabilities was made to equal 2. Brewer's method was then applied to select a person pair. 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.

Starting with the 2002 NSDUH and continuing through 2012, the pair-sampling algorithm was modified to increase the number of pairs selected in the sample. Dwelling units with the sum of person selection probabilities greater than or equal to 2 were treated the same as in previous survey years. However, DUs where the sum of person-level selection probabilities was less than 2 received a slightly different treatment that increased the chance for selecting a pair of real persons. Section 2.1 describes the selection process for both types of DUs.

Table 2.1 provides a summary of these NSDUH sampling units: eligible and completed screening dwelling units (SDUs), selected and completed questionnaire dwelling units (QDUs), selected and completed person interviews, and selected and completed person pairs, as well as their response rates. Using Brewer's method, zero, one, or two individuals were selected per household. Those SDUs where at least one person was selected were counted as the selected QDUs. A QDU where two persons were selected and both had completed interviews was considered to be a completed person pair. The table provides a breakdown by age group at the person level and age group by selection group (none, single, or pair) at the person pair level.

[6] This report presents information from the 2012 National Survey on Drug Use and Health (NSDUH). Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

2.1 Pair Selection Probability

2.1.1 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

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

by setting K at

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

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

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

where age category 1 corresponds 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.

Note the above scheme always selects a pair of two eligible persons.

2.1.2 Case II: DUs with $S < 2$

If the sum S of person-level selection probabilities was less than 2, the method used in survey years 1991 to 2001 consisted of dividing $2 - S$ equally among the three dummy persons

added to the household, and then used Brewer's method to select a pair, as in Case I. However, if the household had two or more persons, we preferred a pair of real persons to have a greater chance of being 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

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

where $T(\lambda) = S + \lambda(2 - S)$ and λ is set to 0.5. Note that if λ is chosen as 0, then $F_s = 1$ and the selection scheme would follow that of Case I. The individual person probabilities are scaled upward by the factor F_s so they either sum to 2 or sum as close to 2 as possible. Denote S' as the sum of the selection probability after scale adjustment by F_s. If S' is exactly 2, then dummy persons are not needed. If S' is less than 2, then three dummy persons are added to the DU.

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

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

by setting K' at

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

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),

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

where age category 0 corresponds to dummy persons, and categories 1 to 5 are defined as in Case I.

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 Brewer's method as follows. First, draw a random number, R, from a uniform (0,1) distribution. 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, $P_{h(i)}$. Unlike Case I, where a pair of eligible persons was always selected, this adjusted selection scheme allows for zero, one, or two persons to be selected from a DU.

2.2 Questionnaire Dwelling Unit Selection Probability

A dwelling unit was considered a selected 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 an unadjusted 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 Building Blocks of the QDU and Person Pair Samples: Dwelling Units and Persons in the 2008–2012 NSDUHs

Domain	2008			2009			2010			2011			2012		
	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	160,114	142,159	88.79	161,377	142,933	88.57	166,532	147,010	88.28	179,293	156,048	87.04	178,586	153,873	86.16
QDUs															
Total QDUs	58,942	48,180	81.74	58,288	48,088	82.50	58,702	48,113	81.96	61,441	50,133	81.60	60,621	48,850	80.58
Persons															
Total Persons	85,711	67,928	79.25	84,785	68,007	80.21	84,997	67,804	79.77	88,536	70,109	79.19	87,656	68,309	77.93
12-17	26,228	22,250	84.83	26,157	22,398	85.63	25,908	21,960	84.76	27,911	23,510	84.23	27,147	22,473	82.78
18-25	28,793	22,875	79.45	28,158	22,686	80.57	28,164	22,793	80.93	28,589	22,876	80.02	28,639	22,529	78.67
26-34	8,337	6,560	78.69	8,242	6,591	79.97	8,545	6,780	79.34	8,323	6,543	78.61	8,304	6,484	78.08
35-49	12,995	9,674	74.44	12,855	9,616	74.80	12,979	9,668	74.49	12,220	9,149	74.87	12,364	9,076	73.41
50+	9,358	6,569	70.20	9,373	6,716	71.65	9,401	6,603	70.24	11,493	8,031	69.88	11,202	7,747	69.16
Pairs															
Total Pairs[4]	26,769	19,748	73.77	26,497	19,919	75.17	26,295	19,691	74.88	27,095	19,976	73.73	27,035	19,459	71.98
0,0[5]	83,217	N/A	N/A	84,645	N/A	N/A	88,308	N/A	N/A	94,607	N/A	N/A	93,252	N/A	N/A
0, 12-17	8,634	8,094	93.75	8,432	7,936	94.12	8,595	7,906	91.98	9,402	8,651	92.01	9,017	8,277	91.79
0, 18-25	9,932	9,213	92.76	9,870	9,081	92.01	10,093	9,270	91.85	10,306	9,497	92.15	10,325	9,461	91.63
0, 26-34	3,771	3,303	87.59	3,798	3,318	87.36	3,914	3,418	87.33	3,930	3,386	86.16	3,856	3,327	86.28
0, 35-49	4,601	3,835	83.35	4,565	3,810	83.46	4,659	3,843	82.49	4,431	3,704	83.59	4,368	3,645	83.45
0, 50+	5,235	4,013	76.66	5,126	4,042	78.85	5,146	3,998	77.69	6,277	4,919	78.37	6,020	4,681	77.76
12-17, 12-17	4,319	3,654	84.60	4,288	3,648	85.07	4,160	3,525	84.74	4,649	3,885	83.57	4,507	3,668	81.38
12-17, 18-25	3,449	2,668	77.36	3,595	2,852	79.33	3,444	2,718	78.92	3,756	2,921	77.77	3,627	2,759	76.07
12-17, 26-34	897	690	76.92	872	724	83.03	922	752	81.56	834	685	82.13	825	658	79.76
12-17, 35-49	3,944	3,012	76.37	3,979	3,061	76.93	3,948	3,044	77.10	3,855	2,918	75.69	3,813	2,812	73.75
12-17, 50+	666	482	72.37	703	532	75.68	679	493	72.61	766	565	73.76	851	631	74.15
18-25, 18-25	5,872	4,261	72.56	5,588	4,168	74.59	5,502	4,165	75.70	5,476	4,015	73.32	5,476	3,901	71.24
18-25, 26-34	1,103	821	74.43	1,102	820	74.41	1,140	851	74.65	1,049	768	73.21	1,079	794	73.59
18-25, 35-49	1,706	1,112	65.18	1,576	1,059	67.20	1,639	1,098	66.99	1,469	994	67.67	1,582	1,053	66.56
18-25, 50+	859	546	63.56	839	547	65.20	844	537	63.63	1,057	666	63.01	1,074	660	61.45
26-34, 26-34	890	614	68.99	886	635	71.67	903	621	68.77	858	599	69.81	880	604	68.64
26-34, 35-49	545	380	69.72	447	312	69.80	512	354	69.14	492	314	63.82	469	320	68.23
26-34, 50+	241	145	60.17	251	149	59.36	251	162	64.54	302	192	63.58	315	177	56.19
35-49, 35-49	873	535	61.28	917	569	62.05	886	542	61.17	748	474	63.37	833	487	58.46
35-49, 50+	453	269	59.38	454	239	52.64	449	245	54.57	477	271	56.81	466	272	58.37
50+, 50+	952	559	58.72	1,000	604	60.40	1,016	584	57.48	1,307	709	54.25	1,238	663	53.55

DU = dwelling unit; N/A = not applicable; QDU = questionnaire dwelling unit.

Note: The 2008–2010 sample sizes reflect the removal of falsified cases found in Pennsylvania and Maryland. The 2011 and 2012 samples were not affected. For additional information, see Section B.3.5 in Appendix B of the 2011 and 2012 national findings reports (Center for Behavioral Health Statistics and Quality, 2012; 2013).

[1] Selected pairs are based on the screener age.
[2] Respondent pairs are based on the questionnaire age and are comprised only of respondent persons.
[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.
[5] Among the completed screening dwelling units, no person was selected in this dwelling unit.

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) via weighting classes, 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 to allow 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 7.

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. (2014).

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. Census Bureau. A way around this 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., 2014).

4.1 Questionnaire Dwelling Unit 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: Non-Group 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-Occupied)[a]
1: 50-100%,[1] 2: 10-<50%, 3: 0-<10%

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

Percentage of Segments That Are Hispanic or Latino[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: Two or More Races

Race/Ethnicity of Householder[a]
1: Hispanic or Latino White,[1] 2: Hispanic or Latino Black or African American, 3: Hispanic or Latino Other, 4: Non-Hispanic or Latino White, 5: Non-Hispanic or Latino Black or African American, 6: Non-Hispanic or Latino Other

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 States 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: Non-Group Quarter[1]

Household Size
1: DU with 2 Persons,[1] 2: DU with 3 Persons, 3: 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: Two or More Races 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: Two or More Races 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
1: 50-100%, 2: 10-<50%, 3: 0-<10%[1]

Percentage of Segments That Are Hispanic or Latino
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 Other, 4: Non-Hispanic or Latino White, 5: Non-Hispanic or Latino Black or African American, 6: Non-Hispanic or Latino Other

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. Sampling weights are often classified as extreme (high or low) if they fall outside the interval, median ± 3 × interquartile range (IQR). The interval is set for prespecified domains defined usually by design variables corresponding to deep stratification.[7] Similar to previous NSDUHs, for the GEM modeling used in the 2012 NSDUH, a more conservative (narrower) interval was defined, median ± 2.5 × IQR. The narrower interval better prevents the adjusted weights from crossing the standard interval boundaries by treating weights near but not outside the commonly used boundaries (i.e., those that have the most potential to become extreme) as extreme as well.

Denote the interval boundaries (or critical values) for low and high extreme values by $b_{k(l)}$ and $b_{k(u)}$, respectively. 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-year-olds, young adult 18- to 25-year-olds, and adult 26 years or older.

[7] Deep stratification refers to the stratification that was used in the sample design. In the case of the 2012 survey, deep stratification refers to the cross-classification of State sampling region by age group.

a. Youth, Young Adult, Adult;
b. Youth, Young Adult;
c. Youth, Adult;
d. Young Adult, Adult;
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 particular selected dwelling unit (DU) from the 2012 survey. This DU gave rise to a selected pair of respondents, both aged 50 or older. The selection probability in this DU was 0.10207 for a respondent aged 50 or older. Using the formula (2.6) in Chapter 2, the pair selection probability was computed to be 0.000227082. Therefore, the inverse of the selection probability, the pair-level design weight, was 4403.70. 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 1999–2012 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.

[8] Pair age in this case should not be confused with the modeling term, which has a finer level breakdown.

The hierarchy was used to satisfy the minimum of 30 observations for defining the boundaries for extreme values. If this sample size requirement was not met at the lower level, then the next level up in the hierarchy was 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 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 alleviate this problem, States were grouped into a small number of clusters, such as 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 was insufficient in some cases, so the most general level of the hierarchy, the national level, was required. Furthermore, 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. Weight Calibration at Questionnaire Dwelling Unit and Pair Levels

The 2012 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. Census Bureau blocks) within 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 2012 NSDUH sample design report (Morton, Martin, Shook-Sa, Chromy, & Hirsch, 2013).

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 2012 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 2012 survey sample are the product of all weight components for each type of sample, illustrated in Exhibits 6.1 and 6.2.

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 6.3 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 2008–2012 surveys in Table 6.1.

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 2002–2012 survey pair data, the built-in ev control feature was used for each adjustment step.

Exhibit 6.1 Summary of 2012 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., 2014). For computational feasibility, all weight adjustments were done using the four model groups based on census regions defined in Exhibit 6.3.

Exhibit 6.2 Summary of 2012 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., 2014). For computational feasibility, all weight adjustments were done using the four model groups based on census regions defined in Exhibit 6.3.

Exhibit 6.3 U.S. Census Bureau 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 6.1 Sample Size, by Model Group at QDU and Pair Levels

Model Group	2008		2009		2010		2011		2012	
	Selected QDUs	Completed QDUs	Selected QDUs	Completed QDUs	Selected QDUs	Completed QDUs	Selected QDUs	Completed QDUs	Selected QDUs	Completed QDUs
QDU										
Northeast	11,490	9,216	11,605	9,340	11,627	9,339	11,997	9,456	12,616	9,917
South	17,703	14,747	17,756	14,909	17,880	14,857	19,690	16,487	18,345	15,019
Midwest	16,763	13,650	16,382	13,491	16,670	13,686	17,045	13,752	16,984	13,687
West	12,986	10,567	12,545	10,348	12,525	10,231	12,709	10,438	12,676	10,227
Total	58,942	48,180	58,288	48,088	58,702	48,113	61,441	50,133	60,621	48,850
Model Group	**2008**		**2009**		**2010**		**2011**		**2012**	
	Selected Pairs	**Completed Pairs**	**Selected Pairs**	**Completed Pairs**	**Selected Pairs**	**Completed Pairs**	**Selected Pairs**	**Completed Pairs**	**Selected Pairs**	**Completed Pairs**
Pair										
Northeast + South	13,060	9,700	13,058	9,806	12,872	9,590	13,686	10,127	13,619	9,723
Midwest + West	13,709	10,048	13,439	10,113	13,423	10,101	13,409	9,849	13,416	9,736
Total	26,769	19,748	26,497	19,919	26,295	19,691	27,095	19,976	27,035	19,459

QDU = questionnaire dwelling unit.

Note: The 2008–2010 sample sizes reflect the removal of falsified cases found in Pennsylvania and Maryland. The 2011 and 2012 samples were not affected. For additional information, see Section B.3.5 in Appendix B of the 2011 and 2012 national findings reports (Center for Behavioral Health Statistics and Quality, 2012; 2013).

6.1 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 2009 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 2012 NSDUH sample design report (Morton et al., 2013), and for more detail on Weight Components #3 and #8 through #10, see the 2012 person-level sampling weight calibration report (Chen et al., 2014).

Note that from 2008 to 2012, there was an occasional second subsegmentation step when the initial partitioning of segments was insufficient due to out-of-date census counts or the segment was still too large to list after the original subsegmentation. This second partitioning was not accounted for in the weighting over these survey years. A comparison was done to evaluate the effect of this omission, and it was determined that the missing second subsegmenting factor in the analysis weight had minimal impact on estimates. Therefore, weights for these years were not reproduced. Additional detail can be found in the 2012 NSDUH sample design report (Morton et al., 2013).

6.2 QDU Weight Components

6.2.1 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).

6.2.2 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. (2014).

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

6.2.3 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.

6.2.4 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.

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

The extreme weight proportions for the final poststratified weights were acceptably low, eliminating the need for the extreme value adjustment. Weight Component #15 was set to one for each responding QDU.

6.3 Pair-Level Weight Components

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

6.3.1 Pair Weight Component #11: Inverse of Selection Probability of a Person Pair in the Dwelling Unit

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 either formula (2.1) or formula (2.6) in Chapter 2. This weight component is the inverse of the selection probability discussed above.

6.3.2 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.

6.3.3 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.

6.3.4 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 11 in the editing and imputation report (Carpenter et al., 2014) for more detail on the 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.

6.3.5 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.

7. 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.

7.1 Response Rates

Table D.1 in Appendix D displays the final selected and responding questionnaire dwelling unit (QDU) sample sizes from the 2012 National Survey on Drug Use and Health (NSDUH) for various national domains. This table also shows the weighted response rates. Most domains reflect the overall 75.87 percent response rate, with most rates relatively close to 80 percent, although the highest response rate is 93.20 percent, from the Group level of the Group Quarters category. The lowest response rate came from Race/Ethnicity of Householder Non-Hispanic or Latino Other, with 72.15 percent.

Table I.1 in Appendix I displays the final selected and responding pair-level sample sizes from the 2012 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 63.76 percent. The response rates range from a low of 37.40 percent in the pair race/ethnicity Black or African American and Other category to a high of 81.51 percent from the Pair Age Group 12-17,12-17. 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.

7.2 Proportions of Extreme Values and Outwinsors

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 proportions 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 6.3.5. 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.

7.3 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. Census Bureau population estimates (Chen et al., 2014). 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 2012 QDU domain household size of two, indicates a sample size of 16,974 with a total design-based weight of 54,585,381 and a census total of 54,556,667 with an initial slippage rate of -0.07 percent, which would imply a common weight adjustment approximately equal to 1.000681, 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, 18-25 has a sample size of 3,901, a design-based weight of 12,270,700, and a census total of 12,545,793, showing an initial slippage of -2.69 percent. The resultant required adjustment would be approximately equal to 1.027695, 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.

7.4 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 7.2—were used as guidelines for determining model adequacy.

7.5 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 in the household, whether the household had health insurance coverage, and number of elders living in the household. The estimates and SEs are displayed in Table 7.1. 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 7.2a to 7.7b.

As seen in Table 7.1, 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 7.1 Estimates of Totals and SEs for Domains of Interest Based on QDU Sample: 2012

Domain	2012				
	n	Baseline (B)[1]	Final (F)[2]	(B-F)/F% (Estimate)	(B-F)/F% (SE)
Households with Family Income					
$0 - < $10,000	5,078	10,460,715 (329,056)	10,444,004 (329,693)	0.16	-0.19
$10,000 - < $20,000	6,555	15,622,561 (398,493)	15,638,194 (400,681)	-0.10	-0.55
$20,000 - < $30,000	5,748	14,195,740 (359,770)	14,214,304 (360,747)	-0.13	-0.27
$30,000 - < $40,000	5,331	13,379,640 (327,509)	13,364,505 (327,477)	0.11	0.01
$40,000 - < $50,000	5,080	12,604,296 (332,618)	12,606,932 (331,778)	-0.02	0.25
$50,000 - < $75,000	7,703	19,169,185 (415,719)	19,159,643 (415,892)	0.05	-0.04
$75,000 - < $100,000	5,265	13,710,548 (383,622)	13,716,657 (385,188)	-0.04	-0.41
$100,000+	8,090	20,005,297 (507,343)	20,003,744 (507,248)	0.01	0.02
Households with Number of Youths (< 18)					
0	20,419	77,402,125 (1,069,348)	77,400,559 (1,071,586)	0.00	-0.21
1	11,659	17,372,185 (274,586)	17,379,411 (276,588)	-0.04	-0.72
2	9,719	15,069,648 (275,565)	15,063,588 (274,382)	0.04	0.43
3	4,603	6,422,304 (162,360)	6,404,101 (161,754)	0.28	0.37
4+	2,450	2,881,721 (96,317)	2,900,323 (99,755)	-0.64	-3.45
Households with Insurance Coverage					
Yes	39,384	99,535,440 (1,147,242)	99,551,648 (1,148,703)	-0.02	-0.13
No	9,466	19,612,543 (381,660)	19,596,335 (381,456)	0.08	0.05
Households with Number of Older Adults (65+)					
0	43,406	89,339,074 (983,199)	89,341,651 (986,485)	-0.00	-0.33
1	3,721	19,513,672 (525,590)	19,511,535 (525,200)	0.01	0.07
2	1,676	10,100,864 (375,736)	10,094,422 (376,016)	0.06	-0.07
3+	47	194,373 (42,011)	200,375 (44,298)	-3.00	-5.16

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 7.2a 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: 2012

Drug	Mother User	2012		
		n	Baseline[1]	Final[2]
Alcohol				
Lifetime	Yes	1,976	30.36 (1.74)	30.03 (1.77)
	No	288	22.64 (3.93)	22.65 (3.96)
	Overall	2,264	29.16 (1.61)	28.89 (1.63)
Past Year	Yes	1,636	26.27 (1.85)	26.01 (1.89)
	No	628	14.28 (2.06)	14.25 (2.07)
	Overall	2,264	22.73 (1.47)	22.52 (1.49)
Past Month	Yes	1,232	13.56 (1.59)	13.54 (1.68)
	No	1,032	9.24 (1.33)	9.25 (1.33)
	Overall	2,264	11.50 (1.06)	11.49 (1.09)
Cigarettes				
Lifetime	Yes	1,458	16.07 (1.51)	15.80 (1.51)
	No	806	10.55 (1.52)	10.57 (1.54)
	Overall	2,264	13.75 (1.14)	13.59 (1.14)
Past Year	Yes	597	15.79 (2.05)	15.34 (2.03)
	No	1,667	7.88 (1.05)	7.84 (1.05)
	Overall	2,264	9.63 (0.94)	9.49 (0.94)
Past Month	Yes	532	9.61 (1.79)	9.50 (1.79)
	No	1,732	4.43 (0.81)	4.29 (0.79)
	Overall	2,264	5.44 (0.75)	5.29 (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 7.2b 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: 2012

		2012		
Drug	**Father User**	***n***	**Baseline[1]**	**Final[2]**
Alcohol				
Lifetime	Yes	1,385	31.37 (2.19)	31.58 (2.28)
	No	99	15.24 (4.48)	15.06 (4.48)
	Overall	1,484	30.19 (2.08)	30.30 (2.16)
Past Year	Yes	1,178	26.45 (2.22)	26.59 (2.33)
	No	306	11.82 (2.29)	11.44 (2.29)
	Overall	1,484	23.51 (1.85)	23.44 (1.92)
Past Month	Yes	971	15.55 (2.21)	15.62 (2.37)
	No	513	6.45 (1.28)	6.24 (1.25)
	Overall	1,484	12.38 (1.53)	12.32 (1.62)
Cigarettes				
Lifetime	Yes	1,059	20.53 (2.20)	20.29 (2.27)
	No	425	10.73 (2.76)	10.87 (3.02)
	Overall	1,484	17.11 (1.75)	16.96 (1.83)
Past Year	Yes	428	16.48 (3.02)	16.40 (3.14)
	No	1,056	10.97 (1.88)	10.95 (1.93)
	Overall	1,484	12.28 (1.60)	12.24 (1.64)
Past Month	Yes	377	13.01 (3.27)	13.55 (3.47)
	No	1,107	3.51 (0.89)	3.43 (0.89)
	Overall	1,484	5.41 (1.00)	5.43 (1.03)

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 7.3a 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: 2012

		2012		
Drug	**Mother User**	***n***	**Baseline[1]**	**Final[2]**
Any Illicit				
Lifetime	Yes	1,204	25.17 (2.08)	25.25 (2.09)
	No	1,060	15.88 (1.72)	16.01 (1.78)
	Overall	2,264	20.42 (1.36)	20.49 (1.38)
Past Year	Yes	229	21.82 (4.03)	21.65 (4.05)
	No	2,035	13.82 (1.19)	13.89 (1.21)
	Overall	2,264	14.48 (1.15)	14.51 (1.17)
Past Month	Yes	103	14.57 (4.77)	15.02 (5.03)
	No	2,161	6.98 (0.90)	7.01 (0.92)
	Overall	2,264	7.28 (0.89)	7.32 (0.91)
Marijuana				
Lifetime	Yes	1,080	17.40 (1.86)	17.45 (1.87)
	No	1,184	11.27 (1.40)	11.25 (1.41)
	Overall	2,264	13.92 (1.12)	13.89 (1.12)
Past Year	Yes	134	20.63 (5.11)	19.63 (4.91)
	No	2,130	10.60 (1.03)	10.60 (1.04)
	Overall	2,264	11.02 (1.01)	10.96 (1.01)
Past Month	Yes	68	12.38 (4.47)	12.18 (4.35)
	No	2,196	5.22 (0.75)	5.25 (0.76)
	Overall	2,264	5.38 (0.74)	5.41 (0.75)

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 7.3b 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: 2012

		2012		
Drug	**Father User**	***n***	**Baseline[1]**	**Final[2]**
Any Illicit				
Lifetime	Yes	898	25.51 (2.48)	26.17 (2.69)
	No	586	19.18 (2.94)	18.99 (2.95)
	Overall	1,484	22.98 (1.90)	23.28 (2.01)
Past Year	Yes	178	24.24 (6.13)	24.57 (6.55)
	No	1,306	15.95 (1.79)	16.02 (1.89)
	Overall	1,484	16.89 (1.74)	16.99 (1.84)
Past Month	Yes	116	23.46 (8.39)	23.69 (8.99)
	No	1,368	8.96 (1.35)	8.96 (1.43)
	Overall	1,484	10.07 (1.43)	10.09 (1.52)
Marijuana				
Lifetime	Yes	816	18.40 (2.32)	18.88 (2.55)
	No	668	10.93 (2.45)	10.93 (2.45)
	Overall	1,484	15.04 (1.68)	15.27 (1.79)
Past Year	Yes	135	25.14 (7.83)	25.84 (8.57)
	No	1,349	11.56 (1.63)	11.71 (1.73)
	Overall	1,484	12.59 (1.62)	12.79 (1.74)
Past Month	Yes	97	23.24 (10.27)	24.22 (11.35)
	No	1,387	6.98 (1.24)	7.05 (1.34)
	Overall	1,484	7.81 (1.32)	7.93 (1.44)

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 7.4 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: 2012

Drug	Parent Talked about Dangers with Child	2012		
		n	Baseline[1]	Final[2]
Alcohol				
Lifetime	Yes	2,227	30.76 (1.62)	30.52 (1.65)
	No	1,468	29.21 (2.26)	29.06 (2.29)
	Overall	3,695	30.12 (1.33)	29.92 (1.35)
Past Year	Yes	2,227	24.91 (1.47)	24.72 (1.51)
	No	1,468	21.28 (1.92)	20.94 (1.91)
	Overall	3,695	23.41 (1.18)	23.17 (1.20)
Past Month	Yes	2,227	12.92 (1.19)	12.88 (1.25)
	No	1,468	11.08 (1.52)	10.88 (1.48)
	Overall	3,695	12.16 (0.94)	12.06 (0.96)
Cigarettes				
Lifetime	Yes	2,227	14.91 (1.24)	14.75 (1.29)
	No	1,468	16.94 (1.86)	16.61 (1.83)
	Overall	3,695	15.75 (1.05)	15.52 (1.06)
Past Year	Yes	2,227	10.73 (1.09)	10.58 (1.11)
	No	1,468	11.97 (1.67)	11.76 (1.63)
	Overall	3,695	11.24 (0.94)	11.07 (0.94)
Past Month	Yes	2,227	5.56 (0.76)	5.48 (0.76)
	No	1,468	6.03 (1.08)	5.86 (1.05)
	Overall	3,695	5.75 (0.63)	5.64 (0.63)

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 7.5 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: 2012

Drug	Parent Talked about Dangers with Child	2012		
		n	Baseline[1]	Final[2]
Any Illicit				
Lifetime	Yes	2,227	20.16 (1.35)	20.25 (1.41)
	No	1,468	24.77 (2.20)	24.87 (2.21)
	Overall	3,695	22.06 (1.23)	22.15 (1.25)
Past Year	Yes	2,227	15.29 (1.23)	15.28 (1.28)
	No	1,468	17.02 (1.93)	17.07 (1.92)
	Overall	3,695	16.00 (1.09)	16.01 (1.11)
Past Month	Yes	2,227	8.46 (1.00)	8.51 (1.06)
	No	1,468	8.64 (1.37)	8.65 (1.39)
	Overall	3,695	8.53 (0.83)	8.56 (0.87)
Marijuana				
Lifetime	Yes	2,227	14.16 (1.17)	14.23 (1.22)
	No	1,468	15.90 (1.81)	15.83 (1.79)
	Overall	3,695	14.88 (1.01)	14.89 (1.03)
Past Year	Yes	2,227	11.56 (1.11)	11.56 (1.17)
	No	1,468	12.65 (1.73)	12.60 (1.71)
	Overall	3,695	12.01 (0.96)	11.99 (0.98)
Past Month	Yes	2,227	6.42 (0.90)	6.49 (0.97)
	No	1,468	6.38 (1.09)	6.38 (1.10)
	Overall	3,695	6.40 (0.70)	6.45 (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 7.6a 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: 2012

Drug	Mother Talked about Dangers with Child	2012		
		n	Baseline[1]	Final[2]
Alcohol				
Lifetime	0 times	171	23.88 (5.10)	24.15 (5.25)
	1-2 times	310	24.65 (3.85)	24.16 (3.91)
	A few times	546	25.42 (3.53)	24.67 (3.50)
	Many times	1,147	33.49 (2.25)	33.44 (2.30)
	Overall	2,174	29.36 (1.65)	29.14 (1.68)
Past Year	0 times	171	15.35 (3.74)	15.15 (3.78)
	1-2 times	310	18.83 (3.49)	18.70 (3.60)
	A few times	546	22.43 (3.49)	21.91 (3.45)
	Many times	1,147	26.16 (2.00)	26.07 (2.06)
	Overall	2,174	23.17 (1.53)	22.99 (1.56)
Past Month	0 times	171	7.53 (2.63)	7.77 (2.81)
	1-2 times	310	8.17 (2.15)	7.91 (2.07)
	A few times	546	10.60 (2.76)	10.15 (2.67)
	Many times	1,147	13.75 (1.54)	13.95 (1.65)
	Overall	2,174	11.58 (1.10)	11.58 (1.13)
Cigarettes				
Lifetime	0 times	171	11.53 (4.89)	10.97 (4.87)
	1-2 times	310	11.01 (2.48)	10.62 (2.38)
	A few times	546	11.05 (2.55)	11.14 (2.59)
	Many times	1,147	16.32 (1.51)	16.20 (1.52)
	Overall	2,174	13.83 (1.16)	13.70 (1.17)
Past Year	0 times	171	5.50 (2.14)	4.86 (2.00)
	1-2 times	310	7.70 (2.25)	7.27 (2.11)
	A few times	546	9.09 (2.50)	9.19 (2.53)
	Many times	1,147	11.52 (1.23)	11.44 (1.24)
	Overall	2,174	9.81 (0.97)	9.68 (0.97)
Past Month	0 times	171	4.12 (1.99)	3.62 (1.87)
	1-2 times	310	5.36 (1.96)	5.10 (1.85)
	A few times	546	5.14 (2.06)	5.00 (2.01)
	Many times	1,147	6.12 (0.94)	6.06 (0.94)
	Overall	2,174	5.59 (0.78)	5.45 (0.76)

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 7.6b 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: 2012

Drug	Father Talked about Dangers with Child	2012		
		n	Baseline[1]	Final[2]
Alcohol				
Lifetime	0 times	174	20.01 (5.17)	21.59 (5.66)
	1-2 times	247	23.73 (4.69)	23.00 (4.75)
	A few times	459	34.16 (4.00)	34.70 (4.24)
	Many times	497	30.66 (3.24)	30.67 (3.30)
	Overall	1,377	29.29 (2.11)	29.53 (2.19)
Past Year	0 times	174	13.01 (4.46)	13.61 (4.84)
	1-2 times	247	21.30 (4.61)	20.68 (4.67)
	A few times	459	24.11 (3.49)	24.44 (3.79)
	Many times	497	24.52 (2.84)	24.34 (2.85)
	Overall	1,377	22.43 (1.83)	22.45 (1.91)
Past Month	0 times	174	3.04 (1.48)	2.91 (1.45)
	1-2 times	247	10.85 (2.56)	10.10 (2.45)
	A few times	459	12.73 (3.03)	13.32 (3.39)
	Many times	497	12.55 (2.37)	12.33 (2.43)
	Overall	1,377	11.18 (1.42)	11.16 (1.52)
Cigarettes				
Lifetime	0 times	174	13.46 (4.95)	14.89 (5.44)
	1-2 times	247	14.25 (4.31)	13.89 (4.37)
	A few times	459	16.12 (3.12)	16.25 (3.42)
	Many times	497	18.51 (2.61)	17.85 (2.61)
	Overall	1,377	16.31 (1.70)	16.23 (1.79)
Past Year	0 times	174	10.70 (4.82)	11.79 (5.33)
	1-2 times	247	10.94 (4.16)	11.04 (4.28)
	A few times	459	9.63 (2.41)	9.40 (2.35)
	Many times	497	14.42 (2.49)	13.92 (2.50)
	Overall	1,377	11.71 (1.55)	11.62 (1.57)
Past Month	0 times	174	6.61 (4.10)	7.29 (4.54)
	1-2 times	247	3.58 (1.15)	3.54 (1.16)
	A few times	459	4.13 (1.28)	4.13 (1.31)
	Many times	497	6.48 (1.89)	6.41 (1.89)
	Overall	1,377	5.16 (0.97)	5.21 (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 7.7a 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: 2012

Drug	Mother Talked about Dangers with Child	2012		
		n	Baseline[1]	Final[2]
Any Illicit				
Lifetime	0 times	171	16.81 (5.15)	16.94 (5.22)
	1-2 times	310	19.66 (3.65)	19.03 (3.54)
	A few times	546	19.04 (2.99)	19.41 (3.12)
	Many times	1,147	22.10 (1.82)	22.29 (1.86)
	Overall	2,174	20.52 (1.39)	20.64 (1.42)
Past Year	0 times	171	5.67 (1.77)	5.74 (1.76)
	1-2 times	310	11.75 (2.64)	11.38 (2.57)
	A few times	546	13.90 (2.70)	13.97 (2.76)
	Many times	1,147	17.25 (1.71)	17.45 (1.75)
	Overall	2,174	14.55 (1.19)	14.65 (1.20)
Past Month	0 times	171	2.73 (1.28)	2.52 (1.13)
	1-2 times	310	6.27 (1.88)	6.20 (1.86)
	A few times	546	7.58 (2.39)	7.72 (2.44)
	Many times	1,147	8.34 (1.28)	8.44 (1.32)
	Overall	2,174	7.32 (0.93)	7.39 (0.95)
Marijuana				
Lifetime	0 times	171	11.81 (4.81)	11.89 (4.83)
	1-2 times	310	11.66 (2.72)	11.39 (2.66)
	A few times	546	13.64 (2.62)	13.78 (2.68)
	Many times	1,147	15.11 (1.46)	15.15 (1.49)
	Overall	2,174	13.93 (1.15)	13.96 (1.16)
Past Year	0 times	171	4.60 (1.62)	4.51 (1.53)
	1-2 times	310	9.29 (2.55)	9.03 (2.47)
	A few times	546	10.87 (2.49)	10.89 (2.52)
	Many times	1,147	12.66 (1.38)	12.69 (1.40)
	Overall	2,174	10.98 (1.03)	10.97 (1.03)
Past Month	0 times	171	1.42 (0.69)	1.41 (0.66)
	1-2 times	310	5.14 (1.82)	5.13 (1.80)
	A few times	546	5.13 (2.12)	5.25 (2.15)
	Many times	1,147	6.22 (0.94)	6.23 (0.97)
	Overall	2,174	5.36 (0.76)	5.40 (0.77)

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 7.7b 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: 2012

Drug	Father Talked about Dangers with Child	2012		
		n	Baseline[1]	Final[2]
Any Illicit				
Lifetime	0 times	174	17.12 (4.72)	18.72 (5.16)
	1-2 times	247	21.40 (4.75)	21.62 (4.90)
	A few times	459	19.79 (3.31)	20.46 (3.62)
	Many times	497	27.18 (3.10)	26.96 (3.24)
	Overall	1,377	22.43 (1.91)	22.83 (2.03)
Past Year	0 times	174	13.47 (4.54)	15.02 (4.99)
	1-2 times	247	15.00 (4.36)	14.70 (4.45)
	A few times	459	15.20 (3.09)	15.84 (3.43)
	Many times	497	19.45 (2.67)	18.69 (2.65)
	Overall	1,377	16.48 (1.72)	16.56 (1.82)
Past Month	0 times	174	7.52 (3.71)	7.71 (3.71)
	1-2 times	247	5.29 (1.57)	4.99 (1.48)
	A few times	459	9.20 (2.61)	9.74 (3.02)
	Many times	497	12.46 (2.30)	11.90 (2.24)
	Overall	1,377	9.43 (1.33)	9.39 (1.42)
Marijuana				
Lifetime	0 times	174	6.97 (3.05)	8.19 (3.79)
	1-2 times	247	14.49 (4.33)	14.92 (4.50)
	A few times	459	14.38 (3.09)	14.96 (3.43)
	Many times	497	16.33 (2.54)	15.99 (2.58)
	Overall	1,377	14.24 (1.63)	14.54 (1.75)
Past Year	0 times	174	6.55 (3.04)	7.76 (3.78)
	1-2 times	247	12.46 (4.28)	12.28 (4.35)
	A few times	459	12.65 (3.05)	13.35 (3.40)
	Many times	497	13.15 (2.39)	12.65 (2.39)
	Overall	1,377	12.08 (1.58)	12.25 (1.69)
Past Month	0 times	174	2.57 (1.47)	2.62 (1.55)
	1-2 times	247	5.00 (1.54)	4.70 (1.46)
	A few times	459	8.32 (2.59)	8.90 (3.01)
	Many times	497	8.53 (2.03)	8.27 (2.01)
	Overall	1,377	7.10 (1.19)	7.16 (1.30)

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

Carpenter, L., Edwards, S., Frechtel, P., Gulledge, K., Handley, W., King, S., Kroutil, L., Laufenberg, J., Martin, P., Moore, A., & Scott, V. (2014). Editing and imputation report. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Deliverable No. 41, RTI/0212800.001.107.006.007). Research Triangle Park, NC: RTI International.

Center for Behavioral Health Statistics and Quality. (2012). *Results from the 2011 National Survey on Drug Use and Health: Summary of national findings* (HHS Publication No. SMA 12-4713, NSDUH Series H-44). Rockville, MD: Substance Abuse and Mental Health Services Administration. Available at http://www.samhsa.gov/data/NSDUH.aspx

Center for Behavioral Health Statistics and Quality. (2013). *Results from the 2012 National Survey on Drug Use and Health: Summary of national findings* (HHS Publication No. SMA 13-4795, NSDUH Series H-46). Rockville, MD: Substance Abuse and Mental Health Services Administration. Available at http://www.samhsa.gov/data/NSDUH.aspx

Chen, P., Cribb, D., Dai, L., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (2014). Person-level sampling weight calibration. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Phase I, Deliverable No. 41, RTI/0212800.001.107.004). Research Triangle Park, NC: RTI International.

Chen, P., Dai, L., Gordek, H., Grau, E., Shi, W., & Westlake, M. (2006). Person-level sampling weight calibration. In *2004 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-03-9028, Deliverable No. 26, RTI/0208726.174.003). Research Triangle Park, NC: RTI International.

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 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 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 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.

Frechtel, P., Warren, L. K., & Porter, J. (in press). *How To Prepare and Analyze Pair Data in the National Survey on Drug Use and Health* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, RTI/0212800). Research Triangle Park, NC: RTI International.

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.

Morton, K. B., Martin, P. C., Shook-Sa, B. E., Chromy, J. R., & Hirsch, E. L. (2013). Sample design report. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Phase I, Deliverable No. 8, RTI/0212800.001.103). Research Triangle Park, NC: RTI International.

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 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.

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* being the sample selected. The distance function minimized under the generalized exponential model (GEM), subject to calibration constraints, is given by

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

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

$$\sum_{k\in s} x_k d_k a_k = T_x. \tag{A.1.2}$$

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

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

Note that the number of parameters in the GEM should be $\le n$, 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

$$\ell_k < a_k < u_k,\ k=1,\ldots,n. \tag{A.1.4}$$

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

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

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

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

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: extreme value treatment, nonresponse, and poststratification. For extreme value treatment via winsorization, denote the winsorized weights by $\{b_k\}$, where $b_k = d_k$ if d_k is not an extreme weight, and $b_k = \text{med}\{d_k\} \pm 3 * \text{IQR}$ if d_k is an extreme weight, where IQR denotes the interquartile range, and the median and 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: the nonextreme weight subsample and 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, $\ell_k = \ell_1 m_k, c_k = \rho^{-1} m_k,$ 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, the following weights are set: for nonextreme weights, $\ell_k = \ell_2, \; c_k = c_2 = 1,$ and $u_k = u_2;$ for high extreme weights, $\ell_k = \ell_1 m_k, c_k = m_k,$ 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 the 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 $n \times p$ 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 the Newton-Raphson iteration v, the value of the p-vector λ is adjusted as

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

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

The convergence criterion is based on the Euclidean distance $\left\|T_x - \hat{T}_x^{(v)}\right\|$, which is defined as $\sqrt{\left(T_x - \hat{T}_x^{(v)}\right)'\left(T_x - \hat{T}_x^{(v)}\right)}$. At each iteration, it is checked to determine whether it is decreasing or not. If not, a half step[1] is used in the iteration increment.

A.4 Scaled Constrained Exponential Model

In National Household Surveys on Drug Abuse (NHSDAs)[2] prior to 1999, constrained exponential models (CEMs) were used for poststratification, and scaled CEMs were used for nonresponse adjustments. The CEM refers to the logit model of Deville and Särndal (1992), in which lower and upper bounds do not vary with k; that is, $\ell_k = \ell$, $u_k = u$, and $c_k = c = 1$, such that $\ell < 1 < u$. Thus, the CEM is a special case of the GEM. For the nonresponse adjustment, Folsom and Witt (1994) modified the CEM 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 the CEM as ρ ensures that the scaled adjustment factor for nonresponse is at least one.

[1] A half step refers to halving the increment in the Newton-Raphson iterative process for fitting GEM.

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

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), Hispanicity (2), race (6), gender (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), gender (2), age (11), and number of States[1] 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.

[1] 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 Chapter 11 in Carpenter et al. (2014) for details on the multiplicity counts and 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

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 appendix provides a model-specific overview of weight calibration, as opposed to a State- or domain-specific one.

For 2012, 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 extreme value 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 4.1 (included here as Exhibit C.1 for easy reference) in Chapter 4. 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: Non-Group 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-Occupied)[a]
1: 50-100%,[1] 2: 10-<50%, 3: 0-<10%

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

Percentage of Segments That Are Hispanic or Latino[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: Two or More Races

Race/Ethnicity of Householder[a]
1: Hispanic or Latino White,[1] 2: Hispanic or Latino Black or African American, 3: Hispanic or Latino Other, 4: Non-Hispanic or Latino White, 5: Non-Hispanic or Latino Black or African American, 6: Non-Hispanic or Latino Other

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 percentage 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

This glossary provides a list of general terms. Certain other terms are sometimes used within a particular section.

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.

Conv. If model is not convergent, dropping or collapsing of variables is performed.

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

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.

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

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

Hier. One or more of the factor effects in a higher order interaction is collapsed or dropped in an interaction at a lower order and the hierarchical effect carries up, either eliminating or combining factors of higher order interactions with that effect.

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.

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 (Chen et al., 2006).

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. Coll. (1,4,2) & (1,4,3); conv. Drop (3,4,2); sing. Drop (3,*,*); conv. Coll. (5,1,2) & (5,1,3); conv. Repeat for all levels of age in State (5).

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,[1] 5: California

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

Race (3 Levels)
1: White,[1] 2: Black or African American, 3: Other

3. Construct the cross-classification table.

For example, the initial proposed set of covariates in Race (4 Levels) is defined this way:

Race (4 Levels)	White	Black or African American	Asian	American Indian or Alaska Native

Shading indicates the reference-level set.

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

This is the cross-classification table for the initial proposed set of covariates in State × Race (4 Levels):

State × Race (4 Levels)	White	Black or African American	Asian	American Indian or Alaska Native
AK				
HI				
OR				
WA				
CA				

Shading indicates the reference-level set.

The cross-classification table of interest for the initial proposed set of covariates in 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+			

Shading indicates the reference-level set.

The number of respondents in the class State × Age × Race (3 Levels) 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.

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)	Coll. (1,3) & (1,4). Do the same for all other States except (2). Coll. (2,2), (2,3), & (2,4).
Age × Race (3 Levels)	All levels present.

The reason for the hier. instruction in the three-factor effect directions 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 we would then have to work with is as follows:

State × Race (4 Levels)	**White**	**Black or African American**	**Asian**	**American Indian or Alaska Native**
AK				
HI				
OR				
WA				
CA				

Shading 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. Coll. (1,4,2) & (1,4,3); conv. Drop (3,4,2); sing. Drop (3,*,*); conv. Coll. (5,1,2) & (5,1,3); conv. Repeat for all levels of age in State (5).

Construct the complete table, and 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 resulting post-modeling 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+			

Shading indicates the reference-level set.

Exhibit C.2 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights

Variables	Binary	Counting	Level	Proposed
One-Factor Effects		**76**	**76**	
Intercept	Yes		1	1
Population Density	Yes		4	3
Group Quarter	Yes		3	2
Race/Ethnicity of Householder	Yes		6	5
Rent/Housing	Yes		5	4
Segment % Black or African American	Yes		3	2
Segment % Hispanic or Latino	Yes		3	2
Segment % Owner-Occupied	Yes		3	2
Household Type	Yes		7	6
State	Yes	Yes	Model-specific	
Quarter	Yes	Yes	4	3
Age Group		Yes	5	4
Race		Yes	5	4
Hispanicity		Yes	2	1
Gender		Yes	2	1
Household Size		Yes	1	1
Two-Factor Effects				
Age × Race (3 Levels)		Yes	5 × 3	8
Age × Hispanicity		Yes	5 × 2	4
Age × Gender		Yes	5 × 2	4
Race (3 Levels) × Hispanicity		Yes	3 × 2	2
Race (3 Levels) × Gender		Yes	3 × 2	2
Hispanicity × Gender		Yes	2 × 2	1
State × Age		Yes	Model-specific	
State × Race (5 Levels)		Yes	Model-specific	
State × Gender		Yes	Model-specific	
State × Hispanicity		Yes	Model-specific	
% Black or African American × % Owner-Occupied	Yes		3 × 3	4
% Black or African American × Rent/Housing		Yes	3 × 5	8
% Hispanicity × % Owner-Occupied		Yes	3 × 3	4
% Hispanicity × Rent/Housing		Yes	3 × 5	8
% Owner × Rent/Housing	Yes		3 × 5	8
Three-Factor Effects				
Race (3 Levels) × Age × Gender		Yes	8	8
State/Region × Age × Gender		Yes		
State/Region × Age × Hispanicity		Yes		
State/Region × Age × Race (3 Levels)		Yes		
State/Region × Hispanicity × Gender		Yes		
State/Region × Race (3 Levels) × Hispanicity		Yes		
State/Region × Race (3 Levels) × Gender		Yes		

Appendix C.1: Model Group 1: Northeast

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

Table C.1a 2012 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	2.33	3.79	1.09	2.9132	243	(0.59, 2.20)	(0.61, 2.20)
	2.15	4.48	1.06	2.9101	242	(0.34, 4.19)	(0.36, 4.19)
						(0.90, 3.03)	(0.90, 3.03)
res.qdu.nr	2.33	4.92	1.27	3.0250	243	(1.00, 2.50)	(1.00, 2.50)
	1.92	4.56	0.86	3.3407	236	(1.00, 4.61)	(1.00, 4.58)
						(1.00, 5.00)	(1.00, 1.17)
res.qdu.ps	1.92	4.56	0.86	3.3407	243	(0.81, 2.90)	(0.82, 2.90)
	2.06	4.21	0.68	3.3402	239	(0.58, 2.90)	(0.60, 2.70)
						(0.90, 1.03)	(0.90, 1.02)

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

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

[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 2012 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	18	1.00	18	0.29	8	0.38	8	0.54	6
1%	80	1.00	89	0.58	89	1.00	98	0.73	96
5%	120	1.00	162	0.74	164	1.01	179	0.93	173
10%	189	1.00	233	0.81	229	1.04	260	0.96	258
25%	305	1.00	503	0.91	483	1.12	521	0.99	513
Median	603	1.17	856	0.98	850	1.23	988	1.00	985
75%	834	3.71	1,954	1.07	1,931	1.37	2,287	1.01	2,308
90%	1,226	6.88	4,304	1.20	4,249	1.54	5,546	1.03	5,531
95%	1,576	7.89	6,266	1.31	6,144	1.65	8,446	1.05	8,442
99%	2,573	12.47	11,326	1.72	12,088	2.10	17,320	1.13	17,225
Maximum	7,128	15.74	30,862	7.32	29,264	4.58	47,848	2.86	45,340
n	12,616	-	12,616	-	12,616	-	9,917	-	9,917
Mean	679	2.62	1,718	1.01	1,705	1.27	2,168	1.00	2,168
Max/Mean	11	-	18	7.27	17	3.59	22	2.86	21

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

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

Model Group 1 Overview

Selected Questionnaire Dwelling Unit-Level Poststratification

Out of 243 proposed effects, 242 were kept in the model, with the exception of State by Race, which combined American Indian or Alaska Native and Asian for New Hampshire.

Respondent Questionnaire Dwelling Unit-Level Nonresponse

Out of 243 proposed effects, 236 were kept in the model. Two-factor effects were modified for State by Race, combining American Indian or Alaska Native and Asian for each of the following States: New Jersey, Connecticut, New York, and Rhode Island. Three-factor effects for State/Region by Race by Hispanicity combined Black or African American with Other.

Respondent Questionnaire Dwelling Unit-Level Poststratification

Out of 243 proposed effects, 239 were kept in the model. Two-factor effects were modified for State by Race, combining American Indian or Alaska Native and Asian for each of the following States: New Jersey, Connecticut, Vermont, and New Hampshire. All main and three-factor effects were retained at proposed levels.

Exhibit C.1.1 Covariates for 2012 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 × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	9 × 5	32	32	All levels present.
State × Race	9 × 5	32	31	Coll. (4,3) & (4,4); conv.
State × Gender	9 × 2	8	8	All levels present.
State × Hispanicity	9 × 2	8	8	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Race (3 Levels)	3 × 5 × 3	16	16	All levels present.
State/Region × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Hispanicity	3 × 3 × 2	4	4	All levels present.
State/Region × Race (3 Levels) × Gender	3 × 3 × 2	4	4	All levels present.
Total		**243**	**242**	

Exhibit C.1.2 Covariates for 2012 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	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**	**127**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	9 × 5	32	32	All levels present.
State × Race	9 × 5	32	26	Coll. (6,3) & (6,4), (1,3) & (1,4), (5,3) & (5,4) & (5,5), (8,3) & (8,4) & (8,5); conv.
State × Gender	9 × 2	8	8	All levels present.
State × Hispanicity	9 × 2	8	8	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**50**	**49**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Race (3 Levels)	3 × 5 × 3	16	16	All levels present.
State/Region × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Hispanicity	3 × 3 × 2	4	3	Coll. (1,2,1) & (1,3,1); conv.
State/Region × Race (3 Levels) × Gender	3 × 3 × 2	4	4	All levels present.
Total		**243**	**236**	

Exhibit C.1.3 Covariates for 2012 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**	**129**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	9 × 5	32	32	All levels present.
State × Race	9 × 5	32	28	Coll. (1,3) & (1,4), (4,3) & (4,4), (5,3) & (5,4), (9,3) & (9,4); conv.
State × Gender	9 × 2	8	8	All levels present.
State × Hispanicity	9 × 2	8	8	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Race (3 Levels)	3 × 5 × 3	16	16	All levels present.
State/Region × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Hispanicity	3 × 3 × 2	4	3	All levels present.
State/Region × Race (3 Levels) × Gender	3 × 3 × 2	4	4	All levels present.
Total		**243**	**239**	

Appendix C.2: Model Group 2: Midwest

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

Table C.2a 2012 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	1.78	1.86	0.42	2.6793	300	(0.42, 1.10)	(0.42, 1.10)
	1.10	0.98	0.16	2.7109	300	(0.20, 3.27)	(0.20, 3.27)
						(0.90, 2.07)	(0.90, 2.07)
res.qdu.nr	1.25	1.26	0.20	2.7756	300	(1.00, 1.30)	(1.00, 1.30)
	0.76	0.90	0.12	2.9433	299	(1.00, 3.43)	(1.00, 3.33)
						(1.10, 1.56)	(1.10, 1.55)
res.qdu.ps	0.76	0.90	0.12	2.9433	300	(0.91, 1.20)	(0.92, 1.20)
	0.81	0.97	0.05	2.9433	300	(0.77, 1.21)	(0.78, 1.21)
						(0.90, 1.03)	(0.90, 1.03)

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

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

[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 2012 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	33	1.00	40	0.19	23	0.43	30	0.56	27
1%	97	1.00	107	0.61	109	1.00	128	0.96	128
5%	147	1.00	217	0.78	210	1.02	242	0.98	241
10%	268	1.00	373	0.85	350	1.05	385	0.99	382
25%	452	1.00	513	0.94	507	1.13	600	1.00	601
Median	541	1.13	737	1.00	762	1.21	911	1.00	913
75%	707	3.64	1,773	1.07	1,751	1.32	2,041	1.00	2,047
90%	1,150	6.65	3,754	1.16	3,805	1.44	4,917	1.01	4,913
95%	1,385	7.56	5,434	1.25	5,460	1.53	6,977	1.02	6,977
99%	1,832	11.84	9,852	1.59	10,371	1.81	14,335	1.05	14,315
Maximum	6,654	13.41	43,516	3.56	25,191	3.33	35,530	1.29	35,530
n	16,984	-	16,984	-	16,984	-	13,687	-	13,687
Mean	633	2.50	1,544	1.01	1,558	1.24	1,933	1.00	1,933
Max/Mean	11	-	28	-	16	-	18	-	18

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

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

Model Group 2 Overview

Selected Questionnaire Dwelling Unit-Level Poststratification

All 300 proposed effects were kept in the model.

Respondent Questionnaire Dwelling Unit-Level Nonresponse

All main and three-factor effects were maintained at proposed levels. Two-factor effects were modified for State by Race, combining American Indian or Alaska Native and Asian for North Dakota.

Respondent Questionnaire Dwelling Unit-Level Poststratification

All 300 proposed effects were kept in the model.

Exhibit C.2.1 Covariates for 2012 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 × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	12 × 5	44	44	All levels present.
State × Race	12 × 5	44	44	All levels present.
State × Gender	12 × 2	11	11	All levels present.
State × Hispanicity	12 × 2	11	11	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity or Latino × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity or Latino × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	4 × 5 × 2	12	12	All levels present.
State/Region × Age × Hispanicity	4 × 5 × 2	12	12	All levels present.
State/Region × Age × Race (3 Levels)	4 × 5 × 3	24	24	All levels present.
State/Region × Hispanicity × Gender	4 × 2 × 2	3	3	All levels present.
State/Region × Race (3 Levels) × Hispanicity	4 × 3 × 2	6	6	All levels present.
State/Region × Race (3 Levels) × Gender	4 × 3 × 2	6	6	All levels present.
Total		**300**	**300**	

Exhibit C.2.2 Covariates for 2012 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**	**162**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	12 × 5	44	44	All levels present.
State × Race	12 × 5	44	43	Coll. (9,3) & (9,4); conv.
State × Gender	12 × 2	11	11	All levels present.
State × Hispanicity	12 × 2	11	11	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	3	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	4 × 5 × 2	12	12	All levels present.
State/Region × Age × Hispanicity	4 × 5 × 2	12	12	All levels present.
State/Region × Age × Race (3 Levels)	4 × 5 × 3	24	24	All levels present.
State/Region × Hispanicity × Gender	4 × 2 × 2	3	3	All levels present.
State/Region × Race (3 Levels) × Hispanicity	4 × 3 × 2	6	6	All levels present.
State/Region × Race (3 Levels) × Gender	4 × 3 × 2	6	6	All levels present.
Total		**300**	**299**	

Exhibit C.2.3 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 2: Midwest

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**66**	**66**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/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 × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	12 × 5	44	44	All levels present.
State × Race	12 × 5	44	44	All levels present.
State × Gender	12 × 2	11	11	All levels present.
State × Hispanicity	12 × 2	11	11	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	4 × 5 × 2	12	12	All levels present.
State/Region × Age × Hispanicity	4 × 5 × 2	12	12	All levels present.
State/Region × Age × Race (3 Levels)	4 × 5 × 3	24	24	All levels present.
State/Region × Hispanicity × Gender	4 × 2 × 2	3	3	All levels present.
State/Region × Race (3 Levels) × Hispanicity	4 × 3 × 2	6	6	All levels present.
State/Region × Race (3 Levels) × Gender	4 × 3 × 2	6	6	All levels present.
Total		**300**	**300**	

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 2012 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	1.50	1.91	0.38	2.5352	339	(0.42, 1.10)	(0.42, 1.10)
	1.17	1.51	0.17	2.5423	338	(0.46, 4.94)	(0.49, 4.93)
						(0.90, 4.99)	(0.90, 4.99)
res.qdu.nr	1.24	1.57	0.20	2.5925	339	(1.00, 1.70)	(1.00, 1.70)
	1.10	1.49	0.24	2.8943	339	(1.00, 5.00)	(1.00, 5.00)
						(1.30, 4.97)	(1.30, 4.79)
res.qdu.ps	1.10	1.49	0.24	2.8943	339	(0.20, 1.50)	(0.94, 1.50)
	1.09	1.51	0.12	2.8950	338	(0.20, 5.00)	(0.79, 2.79)
						(0..90, 5.00)	(0.95, 1.07)

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

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

[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 2012 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	9	1.00	12	0.25	10	0.52	20	0.49	25
1%	59	1.00	75	0.65	71	1.00	75	0.94	75
5%	112	1.00	216	0.80	213	1.02	219	0.98	220
10%	253	1.00	425	0.85	412	1.05	449	0.99	444
25%	641	1.00	817	0.92	813	1.10	911	1.00	912
Median	923	1.15	1,284	1.00	1,300	1.18	1,479	1.00	1,482
75%	1,313	3.50	2,761	1.09	2,787	1.29	2,980	1.00	2,975
90%	1,716	6.54	6,014	1.19	6,214	1.43	7,732	1.01	7,712
95%	2,144	7.92	8,523	1.27	8,924	1.53	11,901	1.02	11,911
99%	3,118	12.02	14,217	1.59	14,457	1.82	19,919	1.06	19,913
Maximum	8,236	21.12	45,296	5.31	44,158	5.00	64,497	2.79	63,870
n	18,345	-	18,345	-	18,345	-	15,019	-	15,019
Mean	1,013	2.55	2,405	1.02	2,444	1.22	2,985	1.00	2,985
Max/Mean	8	-	19	-	18	-	22	-	21

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

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

Model Group 3 Overview

Selected Questionnaire Dwelling Unit-Level Poststratification

Out of 339 proposed effects, 338 were kept in the model. Two-factor effects were modified for State by Race, combining American Indian or Alaska Native and Asian for Delaware.

Respondent Questionnaire Dwelling Unit-Level Nonresponse

All 339 proposed effects were kept in the model.

Respondent Questionnaire Dwelling Unit-Level Poststratification

Out of 339 proposed effects, 338 were kept in the model. Two-factor effects were modified for State by Race, combining American Indian or Alaska Native and Asian for Delaware.

Exhibit C.3.1 Covariates for 2012 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 × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	17 × 5	64	64	All levels present.
State × Race	17 × 5	64	63	Coll. (3,4) & (3,5); conv.
State × Gender	17 × 2	16	16	All levels present.
State × Hispanicity	17 × 2	16	16	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Race (3 Levels)	3 × 5 × 3	16	16	All levels present.
State/Region × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Hispanicity	3 × 3 × 2	4	4	All levels present.
State/Region × Race (3 Levels) × Gender	3 × 3 × 2	4	4	All levels present.
Total		**339**	**338**	

Exhibit C.3.2 Covariates for 2012 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**	**213**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	17 × 5	64	64	All levels present.
State × Race	17 × 5	64	64	All levels present.
State × Gender	17 × 2	16	16	All levels present.
State × Hispanicity	17 × 2	16	16	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Race (3 Levels)	3 × 5 × 3	16	16	All levels present.
State/Region × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Hispanicity	3 × 3 × 2	4	4	All levels present.
State/Region × Race (3 levels) × Gender	3 × 3 × 2	4	4	All levels present.
Total		**339**	**339**	

Exhibit C.3.3 Covariates for 2012 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**	**212**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	17 × 5	64	64	All levels present.
State × Race	17 × 5	64	63	Coll. (3,3) & (3,4); conv.
State × Gender	17 × 2	16	16	All levels present.
State × Hispanicity	17 × 2	16	16	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Race (3 Levels)	3 × 5 × 3	16	16	All levels present.
State/Region × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Hispanicity	3 × 3 × 2	4	4	All levels present.
State/Region × Race (3 levels) × Gender	3 × 3 × 2	4	4	All levels present.
Total		**339**	**338**	

Appendix C.4: Model Group 4: West

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

Table C.4a 2012 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	2.09	3.52	0.99	3.1857	270	(0.49, 2.50)	(0.49, 2.50)
	1.44	2.60	0.62	3.1282	268	(0.34, 3.27)	(0.34, 3.27)
						(0.90, 1.66)	(0.90, 1.66)
res.qdu.nr	1.51	3.11	0.72	3.1508	270	(1.00, 2.90)	(1.00, 2.90)
	1.10	3.36	0.66	3.5391	267	(1.00, 5.00)	(1.00, 5.00)
						(1.30, 1.34)	(1.30, 1.34)
res.qdu.ps	1.10	3.36	0.66	3.5391	270	(0.82, 2.70)	(0.84, 2.70)
	0.91	2.98	0.57	3.5469	268	(0.71, 2.70)	(0.75, 1.71)
						(0.90, 1.02)	(1.02, 1.02)

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

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

[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 2012 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	20	1.00	20	0.22	13	0.45	41	0.38	44
1%	87	1.00	96	0.61	92	1.00	98	0.87	101
5%	112	1.00	141	0.74	132	1.02	155	0.96	155
10%	142	1.00	182	0.79	178	1.05	208	0.98	209
25%	251	1.00	426	0.88	416	1.11	468	0.99	469
Median	707	1.16	1,103	0.98	1,083	1.20	1,228	1.00	1,222
75%	1,480	3.32	2,176	1.09	2,200	1.32	2,645	1.01	2,627
90%	1,952	6.05	5,347	1.21	5,195	1.47	6,245	1.02	6,246
95%	2,211	7.90	8,100	1.31	8,010	1.57	10,398	1.04	10,482
99%	2,994	12.50	15,867	1.62	15,947	1.92	21,382	1.18	21,060
Maximum	8,444	14.78	48,558	3.77	35,038	5.00	53,491	1.71	59,335
n	12,676	-	12,676	-	12,676	-	10,227	-	10,227
Mean	918	2.43	2,109	1.00	2,080	1.24	2,578	1.00	2,578
Max/Mean	9	-	23	-	17	-	21	-	23

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

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

Model Group 4 Overview

Selected Questionnaire Dwelling Unit-Level Poststratification

Out of 270 proposed effects, 268 were kept in the model. All main effects were maintained in full. Two-factor effects were modified for Rent/Housing by Percent Black or African American, combining 50-100% and 10-<50% for the first and the fourth quintiles.

Respondent Questionnaire Dwelling Unit-Level Nonresponse

Out of 270 proposed effects, 267 were kept in the model. All main effects were maintained in full. Two-factor effects were modified for Rent/Housing by Percent Black or African American, combining 50-100% and 10-<50% for the first and the fourth quintiles. Three-factor effects were modified for State by Race by Hispanicity, combining Black or African American with Other for California.

Respondent Questionnaire Dwelling Unit-Level Poststratification

This step used the same set of effects as the selected questionnaire dwelling unit-level poststratification.

Exhibit C.4.1 Covariates for 2012 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**	**171**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	13 × 5	48	48	All levels present.
State × Race	13 × 5	48	48	All levels present.
State × Gender	13 × 2	12	12	All levels present.
State × Hispanicity	13 × 2	12	12	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	6	Coll. (1,1) & (2,1); zero. Coll. (1,4) & (2,4); sing.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**29**	**29**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	2 × 5 × 2	4	4	All levels present.
State/Region × Age × Hispanicity	2 × 5 × 2	4	4	All levels present.
State/Region × Age × Race (3 Levels)	2 × 5 × 3	8	8	All levels present.
State/Region × Hispanicity × Gender	2 × 2 × 2	1	1	All levels present.
State/Region × Race (3 Levels) × Hispanicity	2 × 3 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Gender	2 × 3 × 2	2	2	All levels present.
Total		**270**	**268**	

Exhibit C.4.2 Covariates for 2012 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**	**171**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	13 × 5	48	48	All levels present.
State × Race	13 × 5	48	48	All levels present.
State × Gender	13 × 2	12	12	All levels present.
State × Hispanicity	13 × 2	12	12	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	6	Coll. (1,1) & (2,1); zero. Coll. (1,4) & (2,4); sing.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**29**	**28**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	2 × 5 × 2	4	4	All levels present.
State/Region × Age × Hispanicity	2 × 5 × 2	4	4	All levels present.
State/Region × Age × Race (3 Levels)	2 × 5 × 3	8	8	All levels present.
State/Region × Hispanicity × Gender	2 × 2 × 2	1	1	All levels present.
State/Region × Race (3 Levels) × Hispanicity	2 × 3 × 2	2	1	Coll. (3,2,1) & (3,3,1); conv.
State/Region × Race (3 Levels) × Gender	2 × 3 × 2	2	2	All levels present.
Total		**270**	**267**	

Exhibit C.4.3 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 4: West

This step used the same set of covariates as the selected questionnaire dwelling unit poststratification.

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

Table D.1 2012 NSDUH QDU-Level Response Rates

Domain	Selected QDU	Respondent QDU	% Interview Response Rate[1]
Total	60,621	48,850	75.87
Census Region			
Northeast	12,616	9,917	73.10
South	18,345	15,019	76.39
Midwest	16,984	13,687	77.25
West	12,676	10,227	75.87
Quarter			
Quarter 1	14,387	11,628	75.80
Quarter 2	16,164	12,975	76.14
Quarter 3	15,762	12,743	76.00
Quarter 4	14,308	11,504	75.55
Household Type			
12-17, 18-25, 26+	5,811	5,004	85.32
12-17, 18-25	88	72	78.98
12-17, 26+	17,350	14,625	84.38
18-25, 26+	12,807	10,290	79.79
12-17	29	24	81.94
18-25	6,367	5,447	85.41
26+	18,169	13,388	72.47
Race/Ethnicity of Householder			
Hispanic or Latino White	7,920	6,574	78.41
Hispanic or Latino Black or African American	187	165	78.73
Hispanic or Latino Other	415	355	83.80
Non-Hispanic or Latino White	40,460	32,098	74.78
Non-Hispanic or Latino Black or African American	7,135	6,092	81.01
Non-Hispanic or Latino Other	4,504	3,566	72.15
% Hispanic or Latino in Segment			
50-100%	4,137	3,397	78.21
10-<50%	11,129	9,003	75.80
<10%	45,355	36,450	75.66
% Black or African American in Segment			
50-100%	4,381	3,749	80.55
10-<50%	9,185	7,524	76.96
<10%	47,055	37,577	75.17
% Owner-Occupied DUs in Segment			
50-100%	45,901	36,704	75.09
10-<50%	11,429	9,421	77.84
<10%	3,291	2,725	79.98
Combined Median Rent/Housing Value			
1st Quintile	9,920	8,288	78.90
2nd Quintile	12,960	10,661	77.97
3rd Quintile	13,713	11,009	75.45
4th Quintile	13,171	10,428	74.99
5th Quintile	10,857	8,464	73.22
Population Density			
Large MSA	25,508	20,223	74.24
Medium to Small MSA	30,307	24,705	77.53
Non-MSA, Urban	1,249	1,028	80.00
Non-MSA, Rural	3,557	2,894	77.06
Group Quarters			
Group	774	729	93.20
Non-Group	59,847	48,121	75.76
Household Size			
One	7,063	5,567	75.11
Two	21,700	16,974	73.66
Three	17,330	14,071	78.45
Four or More	14,528	12,238	82.73

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 WT1*...*WT9*DUWT10*DUWT11.

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

Table E.1 2012 NSDUH Selected QDU-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)			Before sel.qdu.ps[1] (SDUWT*DUWT10)			After sel.qdu.ps[1] (SDUWT*DUWT10*DUWT11)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Total	60,621	2.62	5.38	1.44	1.88	2.60	0.66	1.41	2.17	0.43
Census Region										
Northeast	12,616	2.85	6.54	2.01	2.33	3.79	1.09	2.15	4.48	1.06
South	18,345	2.15	4.78	1.13	1.50	1.91	0.38	1.17	1.51	0.17
Midwest	16,984	2.47	4.31	0.91	1.78	1.86	0.42	1.10	0.98	0.16
West	12,676	3.26	6.46	1.98	2.09	3.52	0.99	1.44	2.60	0.62
Quarter										
Quarter 1	14,387	2.91	5.41	1.48	1.97	2.64	0.70	1.46	2.27	0.50
Quarter 2	16,164	2.36	5.15	1.41	1.68	2.28	0.59	1.39	2.35	0.49
Quarter 3	15,762	2.33	4.78	1.12	1.76	2.49	0.54	1.21	1.95	0.33
Quarter 4	14,308	2.93	6.19	1.75	2.12	3.00	0.81	1.60	2.10	0.39
Household Type										
12-17, 18-25, 26+	5,811	2.63	5.43	1.52	2.63	5.43	1.52	2.27	5.20	1.09
12-17, 18-25	88	4.55	7.61	1.34	0.00	0.00	0.00	0.00	0.00	0.00
12-17, 26+	17,350	2.66	5.20	1.42	2.64	5.18	1.42	1.98	4.21	0.89
18-25, 26+	12,807	2.96	6.18	1.63	2.46	5.57	1.47	1.56	4.39	1.12
12-17	29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18-25	6,367	2.42	4.70	0.96	2.18	4.44	0.91	1.66	3.42	0.61
26+	18,169	2.39	5.12	1.43	0.40	1.35	0.32	0.40	1.16	0.18
Race/Ethnicity of Householder										
Hispanic or Latino White	7,920	2.45	5.03	1.41	1.93	2.74	0.80	1.40	2.28	0.41
Hispanic or Latino Black or African American	187	57.75	82.43	35.31	48.13	60.13	21.66	40.64	61.93	18.69
Hispanic or Latino Other	415	34.70	62.52	23.51	27.95	44.43	11.95	17.83	31.05	6.77
Non-Hispanic or Latino White	40,460	1.26	2.31	0.43	0.80	1.01	0.18	0.49	0.68	0.08
Non-Hispanic or Latino Black or African American	7,135	3.90	7.48	1.82	3.10	4.13	1.04	3.27	4.20	0.80
Non-Hispanic or Latino Other	4,504	7.88	11.84	2.62	5.20	5.19	1.23	3.60	4.59	0.96

Table E.1 2012 NSDUH Selected QDU-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)			Before sel.qdu.ps[1] (SDUWT*DUWT10)			After sel.qdu.ps[1] (SDUWT*DUWT10*DUWT11)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
% Hispanic or Latino in Segment										
50-100%	4,137	2.13	5.12	1.88	1.84	3.56	1.13	1.55	3.46	1.00
10-<50%	11,129	3.80	8.36	2.68	2.95	4.25	1.26	2.40	3.64	0.81
<10%	45,355	2.37	4.37	0.94	1.62	2.01	0.43	1.15	1.59	0.25
% Black or African American in Segment										
50-100%	4,381	3.54	8.02	2.26	3.15	4.93	1.21	3.99	6.13	1.37
10-<50%	9,185	2.77	6.22	1.97	2.19	3.75	1.15	2.01	3.58	0.79
<10%	47,055	2.50	4.89	1.22	1.70	2.13	0.50	1.05	1.47	0.26
% Owner-Occupied DUs in Segment										
50-100%	45,901	2.27	4.70	1.25	1.59	2.10	0.52	1.06	1.55	0.27
10-<50%	11,429	3.73	7.01	1.84	2.80	3.95	1.00	1.95	3.05	0.74
<10%	3,291	3.65	8.34	2.33	2.61	4.85	1.40	4.35	7.78	1.53
Combined Median Rent/Housing Value										
1st Quintile	9,920	2.46	4.70	1.26	1.88	2.29	0.60	1.76	2.24	0.42
2nd Quintile	12,960	2.35	5.03	1.55	1.87	2.70	0.66	1.40	1.94	0.47
3rd Quintile	13,713	2.68	5.38	1.47	1.90	2.65	0.65	1.24	2.07	0.39
4th Quintile	13,171	2.73	5.46	1.50	1.90	2.55	0.70	1.31	2.30	0.50
5th Quintile	10,857	2.88	6.09	1.33	1.83	2.72	0.66	1.43	2.28	0.35
Population Density										
Large MSA[1]	25,508	2.93	6.67	1.86	2.16	3.43	0.90	1.81	2.91	0.62
Medium to Small MSA[1]	30,307	2.47	4.22	1.06	1.75	1.84	0.46	1.17	1.53	0.24
Non-MSA,[1] Urban	1,249	1.28	1.23	0.16	0.96	0.41	0.06	0.88	0.54	0.09
Non-MSA,[1] Rural	3,557	2.08	2.19	0.38	1.27	1.37	0.12	0.79	0.54	0.12
Group Quarters										
Group	774	4.01	7.46	0.51	2.71	3.60	0.36	2.97	6.11	1.34
Non-Group	59,847	2.60	5.35	1.45	1.87	2.60	0.66	1.39	2.14	0.42
Household Size										
One	7,063	2.10	5.10	1.33	0.93	1.17	0.28	0.72	1.02	0.14
Two	21,700	2.35	5.09	1.38	1.35	2.17	0.51	1.01	1.70	0.32
Three	17,330	2.80	5.40	1.50	2.44	4.60	1.28	1.51	3.49	0.80
Four or More	14,528	3.06	5.85	1.47	2.46	4.72	1.21	2.21	4.72	1.00

[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 2012 NSDUH Respondent QDU-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DUWT10*DUWT11)			After res.qdu.nr[1] (SDUWT*DUWT10*...*DUWT12)			Final Weight: After res.qdu.ps[1] (SDUWT*DUWT10*...*DUWT13)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Total	48,850	1.52	2.42	0.50	1.17	2.32	0.42	1.17	2.20	0.31
Census Region										
Northeast	9,917	2.33	4.92	1.27	1.92	4.56	0.86	2.06	4.21	0.68
South	15,019	1.24	1.57	0.20	1.10	1.49	0.24	1.09	1.51	0.12
Midwest	13,687	1.25	1.26	0.20	0.76	0.90	0.12	0.81	0.97	0.05
West	10,227	1.51	3.11	0.72	1.10	3.36	0.66	0.91	2.98	0.57
Quarter										
Quarter 1	11,628	1.49	2.55	0.56	1.18	2.58	0.43	1.22	2.49	0.32
Quarter 2	12,975	1.53	2.81	0.58	1.21	2.43	0.51	1.16	2.24	0.40
Quarter 3	12,743	1.37	2.08	0.38	0.94	2.01	0.33	0.97	1.91	0.22
Quarter 4	11,504	1.70	2.24	0.47	1.36	2.28	0.40	1.36	2.16	0.29
Household Type										
12-17, 18-25, 26+	5,004	2.16	5.05	1.03	1.60	4.54	1.05	1.58	4.10	0.75
12-17, 18-25	72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12-17, 26+	14,625	2.18	4.63	0.99	1.42	3.71	0.86	1.41	3.42	0.67
18-25, 26+	10,290	1.73	4.53	1.18	1.62	4.91	1.17	1.66	4.67	0.91
12-17	24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18-25	5,447	1.43	3.37	0.69	1.19	2.40	0.34	1.10	2.14	0.30
26+	13,388	0.44	1.31	0.22	0.39	1.50	0.18	0.41	1.46	0.11
Race/Ethnicity of Householder										
Hispanic or Latino White	6,574	1.55	2.76	0.40	1.11	2.30	0.43	1.22	2.36	0.34
Hispanic or Latino Black or African American	165	40.61	65.50	19.97	38.18	57.52	16.88	40.00	55.69	15.87
Hispanic or Latino Other	355	16.90	30.32	7.01	10.14	22.63	4.43	9.30	16.59	2.82
Non-Hispanic or Latino White	32,098	0.53	0.66	0.09	0.50	0.81	0.07	0.40	0.68	0.02
Non-Hispanic or Latino Black or African American	6,092	3.33	4.69	0.95	1.63	3.26	0.63	1.89	3.32	0.48
Non-Hispanic or Latino Other	3,566	3.95	5.81	1.33	3.93	8.90	1.64	4.21	9.25	1.14
% Hispanic or Latino in Segment										
50-100%	3,397	1.65	3.36	0.97	1.32	2.70	0.77	1.47	2.70	0.65
10-<50%	9,003	2.63	4.27	0.95	1.79	3.98	0.82	1.88	3.77	0.61
<10%	36,450	1.23	1.76	0.31	1.00	1.78	0.26	0.97	1.68	0.18

Table E.2 2012 NSDUH Respondent QDU-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DUWT10*DUWT11)			After res.qdu.nr[1] (SDUWT*DUWT10*...*DUWT12)			Final Weight: After res.qdu.ps[1] (SDUWT*DUWT10*...*DUWT13)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
% Black or African American in Segment										
50-100%	3,749	4.27	7.07	1.64	2.37	6.14	1.03	2.69	5.77	0.77
10-<50%	7,524	2.21	3.91	0.87	1.66	3.50	0.80	1.70	3.25	0.65
<10%	37,577	1.11	1.60	0.30	0.95	1.69	0.27	0.91	1.62	0.18
% Owner-Occupied DUs in Segment										
50-100%	36,704	1.18	1.83	0.35	1.01	2.00	0.34	0.98	1.91	0.24
10-<50%	9,421	1.98	3.19	0.77	1.33	2.82	0.56	1.54	2.79	0.46
<10%	2,725	4.48	7.59	1.61	2.79	5.11	0.98	2.46	4.33	0.74
Combined Median Rent/Housing Value										
1st Quintile	8,288	1.91	2.67	0.44	0.98	1.47	0.32	0.92	1.37	0.25
2nd Quintile	10,661	1.52	2.19	0.52	0.92	1.77	0.43	1.01	1.76	0.42
3rd Quintile	11,009	1.38	2.24	0.48	1.15	2.40	0.38	1.13	2.09	0.20
4th Quintile	10,428	1.44	2.74	0.61	1.18	2.74	0.47	1.23	2.70	0.38
5th Quintile	8,464	1.42	2.30	0.41	1.68	2.90	0.45	1.59	2.76	0.25
Population Density										
Large MSA[1]	20,223	1.96	3.19	0.72	1.71	3.52	0.63	1.67	3.26	0.46
Medium to Small MSA[1]	24,705	1.27	1.83	0.31	0.81	1.03	0.20	0.86	1.08	0.15
Non-MSA,[1] Urban	1,028	0.78	0.53	0.10	0.58	0.33	0.05	0.49	0.32	0.05
Non-MSA,[1] Rural	2,894	0.83	0.60	0.16	0.66	1.80	0.20	0.59	1.64	0.07
Group Quarters										
Group	729	2.88	6.48	1.42	1.51	2.75	0.15	1.37	3.82	1.11
Non-Group	48,121	1.50	2.39	0.49	1.16	2.32	0.42	1.17	2.19	0.30
Household Size										
One	5,567	0.70	1.13	0.17	0.79	1.43	0.15	0.65	1.30	0.11
Two	16,974	1.10	2.01	0.42	0.95	2.00	0.31	0.98	1.89	0.21
Three	14,071	1.69	3.67	0.78	1.17	2.95	0.72	1.28	3.06	0.65
Four or More	12,238	2.27	4.85	1.07	1.64	4.72	1.04	1.54	4.26	0.65

[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 2012 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	48,850	119,147,983	119,147,983	119,147,983	0.00	-0.00
Census Region						
Northeast	9,917	21,504,846	21,504,846	21,504,846	0.00	-0.00
South	15,019	44,826,838	44,826,838	44,826,838	0.00	-0.00
Midwest	13,687	26,453,818	26,453,818	26,453,818	0.00	-0.00
West	10,227	26,362,481	26,362,481	26,362,481	0.00	-0.00
Quarter						
Quarter 1	11,628	29,829,432	29,829,432	29,829,432	0.00	-0.00
Quarter 2	12,975	29,634,723	29,634,723	29,634,723	0.00	-0.00
Quarter 3	12,743	29,744,017	29,744,017	29,744,017	0.00	-0.00
Quarter 4	11,504	29,939,811	29,939,811	29,939,811	0.00	-0.00
Household Type						
12-17, 18-25, 26+	5,004	5,062,357	5,062,357	5,062,357	0.00	0.00
12-17, 18-25	72	66,238	66,238	66,238	0.00	0.00
12-17, 26+	14,625	13,691,239	13,691,239	13,691,239	0.00	-0.00
18-25, 26+	10,290	13,956,293	13,956,293	13,956,293	0.00	-0.00
12-17	24	22,258	22,258	22,258	0.00	0.00
18-25	5,447	5,779,323	5,779,323	5,779,323	0.00	-0.00
26+	13,388	80,570,274	80,570,274	80,570,274	0.00	-0.00
Race/Ethnicity of Householder						
Hispanic or Latino White	6,574	13,508,511	13,508,511	13,508,511	-0.00	-0.00
Hispanic or Latino Black or African American	165	725,656	725,656	725,656	-0.00	-0.00
Hispanic or Latino Other	355	1,005,979	1,005,979	1,005,979	0.00	-0.00
Non-Hispanic or Latino White	32,098	81,939,462	81,939,462	81,939,462	0.00	-0.00
Non-Hispanic or Latino Black or African American	6,092	14,332,726	14,332,726	14,332,726	0.00	-0.00
Non-Hispanic or Latino Other	3,566	7,635,648	7,635,648	7,635,648	0.00	-0.00
% Hispanic or Latino in Segment						
50-100%	3,397	8,634,013	8,634,013	8,634,013	0.00	-0.00
10-<50%	9,003	25,715,926	25,715,926	25,715,926	0.00	-0.00
<10%	36,450	84,798,044	84,798,044	84,798,044	0.00	-0.00
% Black or African American in Segment						
50-100%	3,749	8,810,957	8,810,957	8,810,957	0.00	-0.00
10-<50%	7,524	20,110,346	20,110,346	20,110,346	-0.00	-0.00
<10%	37,577	90,226,680	90,226,679	90,226,679	0.00	-0.00
% Owner-Occupied DUs in Segment						
50-100%	36,704	90,409,242	90,409,242	90,409,242	0.00	-0.00
10-<50%	9,421	22,264,244	22,264,244	22,264,244	0.00	-0.00
<10%	2,725	6,474,497	6,474,497	6,474,497	0.00	-0.00

Table F.1 2012 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	8,288	17,328,372	17,328,372	17,328,372	0.00	-0.00
2nd Quintile	10,661	23,902,320	23,902,320	23,902,320	0.00	-0.00
3rd Quintile	11,009	24,983,991	24,983,991	24,983,991	0.00	-0.00
4th Quintile	10,428	27,292,895	27,292,895	27,292,895	0.00	-0.00
5th Quintile	8,464	25,640,405	25,640,405	25,640,405	0.00	-0.00
Population Density						
Large MSA	20,223	60,608,068	60,608,068	60,608,068	0.00	-0.00
Medium to Small MSA	24,705	50,803,819	50,803,818	50,803,818	0.00	-0.00
Non-MSA, Urban	1,028	2,013,232	2,013,232	2,013,232	0.00	-0.00
Non-MSA, Rural	2,894	5,722,865	5,722,865	5,722,865	0.00	-0.00
Group Quarters						
Group	729	766,775	766,775	766,775	0.00	-0.00
Non-Group	48,121	118,381,208	118,381,208	118,381,208	0.00	-0.00
Household Size						
One	5,567	31,354,646	31,358,657	31,327,228	0.09	0.10
Two	16,974	54,585,381	54,556,667	54,622,573	-0.07	-0.12
Three	14,071	19,065,921	19,111,984	19,137,226	-0.37	-0.13
Four or More	12,238	14,142,035	14,120,675	14,060,955	0.58	0.42

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

[1] WT1*...*WT9*DUWT10*...*DUWT12 (before QDU poststratification).

[2] WT1*...*WT9*DUWT10*...*DUWT13 (after QDU poststratification).

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

Table G.1 2012 NSDUH Selected QDU-Level Weight Summary Statistics

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)						Before sel.qdu.ps[1] (SDUWT*DUWT10)						After sel.qdu.ps[1] (SDUWT*DUWT10*DUWT11)					
		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	60,621	9	436	664	1,098	8,444	1.56	12	552	1,013	2,115	48,558	2.86	8	542	1,015	2,137	44,158	2.85
Census Region																			
Northeast	12,616	18	305	603	834	7,128	1.57	18	503	856	1,954	30,862	2.91	8	483	850	1,931	29,264	2.91
South	18,345	9	641	923	1,313	8,236	1.39	12	817	1,284	2,761	45,296	2.54	10	813	1,300	2,787	44,158	2.54
Midwest	16,984	33	452	541	707	6,654	1.35	40	513	737	1,773	43,516	2.68	23	507	762	1,751	25,191	2.71
West	12,676	20	251	707	1,480	8,444	1.75	20	426	1,103	2,176	48,558	3.19	13	416	1,083	2,200	35,038	3.13
Quarter																			
Quarter 1	14,387	21	495	717	1,140	8,082	1.54	21	610	1,079	2,216	45,473	2.88	10	597	1,065	2,231	35,038	2.85
Quarter 2	16,164	18	402	601	1,045	8,236	1.56	18	505	951	1,961	30,862	2.80	8	498	955	1,980	27,818	2.84
Quarter 3	15,762	9	431	617	1,087	7,138	1.57	12	520	981	2,068	45,296	2.85	10	510	975	2,077	44,158	2.85
Quarter 4	14,308	16	471	699	1,111	8,444	1.57	16	590	1,035	2,274	48,558	2.88	13	585	1,062	2,280	30,041	2.84
Household Type																			
12-17, 18-25, 26+	5,811	18	463	679	1,156	8,444	1.59	18	463	679	1,156	8,446	1.59	13	454	680	1,164	9,750	1.59
12-17, 18-25	88	42	357	592	958	2,459	1.53	42	357	592	958	2,460	1.53	27	350	559	1,099	2,529	1.61
12-17, 26+	17,350	16	412	626	1,059	8,351	1.55	16	414	628	1,065	8,379	1.56	10	400	631	1,080	8,263	1.56
18-25, 26+	12,807	9	486	709	1,188	8,236	1.57	12	584	891	1,410	9,795	1.53	8	569	905	1,441	11,391	1.54
12-17	29	83	319	641	951	2,460	1.52	84	322	648	957	2,485	1.52	116	322	579	1,205	2,291	1.56
18-25	6,367	21	364	652	1,069	6,940	1.58	21	424	784	1,237	7,503	1.52	30	408	768	1,226	6,586	1.56
26+	18,169	20	435	656	1,080	8,082	1.53	74	1,893	3,509	5,773	48,558	1.72	53	1,865	3,491	5,815	44,158	1.71
Race/Ethnicity of Householder																			
Hispanic or Latino White	7,920	19	505	809	1,317	8,444	1.44	20	596	1,102	1,839	27,837	2.38	8	589	1,110	1,858	26,638	2.47
Hispanic or Latino Black or African American	187	18	807	1,559	2,869	8,351	1.79	18	947	2,014	4,659	37,254	2.79	20	896	1,913	4,342	29,264	2.76
Hispanic or Latino Other	415	9	189	615	1,448	7,653	2.68	12	269	935	2,525	45,473	4.61	10	298	964	2,404	33,561	4.15
Non-Hispanic or Latino White	40,460	18	416	625	1,016	8,082	1.50	18	537	992	2,214	48,558	2.86	10	530	985	2,217	44,158	2.87
Non-Hispanic or Latino Black or African American	7,135	29	549	757	1,191	7,579	1.50	29	667	1,046	2,111	43,516	2.80	13	656	1,078	2,186	31,554	2.69
Non-Hispanic or Latino Other	4,504	20	244	595	1,222	8,359	1.79	20	383	924	2,043	39,281	3.04	21	362	896	1,942	35,038	3.03
% Hispanic or Latino in Segment																			
50-100%	4,137	26	590	1,011	1,441	6,622	1.34	26	769	1,331	2,091	22,955	2.35	33	763	1,348	2,139	26,638	2.43
10-<50%	11,129	9	565	876	1,420	8,444	1.49	12	712	1,317	2,491	43,516	2.57	10	712	1,330	2,557	30,041	2.56
<10%	45,355	16	384	605	963	8,082	1.57	16	512	916	2,005	48,558	3.00	8	503	913	1,999	44,158	2.98

Table G.1 2012 NSDUH Selected QDU-Level Weight Summary Statistics (continued)

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)						Before sel.qdu.ps[1] (SDUWT*DUWT10)						After sel.qdu.ps[1] (SDUWT*DUWT10*DUWT11)					
		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%	4,381	9	530	735	1,115	7,874	1.57	12	648	1,002	2,038	27,138	2.99	10	650	1,050	2,184	31,554	2.87
10-<50%	9,185	22	533	802	1,286	8,236	1.48	22	688	1,192	2,412	43,516	2.65	13	680	1,197	2,432	30,807	2.59
<10%	47,055	18	400	627	1,055	8,444	1.57	18	520	975	2,070	48,558	2.89	8	511	970	2,071	44,158	2.91
% Owner-Occupied DUs[1] in Segment																			
50-100%	45,901	18	434	648	1,074	8,444	1.55	18	546	1,005	2,131	48,558	2.85	8	540	1,011	2,143	44,158	2.84
10-<50%	11,429	16	433	698	1,115	8,236	1.57	16	567	1,006	2,055	43,516	2.98	13	541	1,000	2,040	35,038	2.97
<10%	3,291	9	473	810	1,317	8,351	1.59	12	599	1,135	2,145	37,254	2.64	10	585	1,150	2,315	27,818	2.64
Combined Median Rent/Housing Value																			
1st Quintile	9,920	19	385	594	903	8,351	1.58	19	507	876	1,879	26,477	2.82	8	502	900	1,909	31,554	2.84
2nd Quintile	12,960	18	402	622	1,024	8,444	1.63	18	512	944	1,995	30,624	2.90	10	514	948	2,013	29,325	2.85
3rd Quintile	13,713	9	370	612	1,008	8,082	1.61	12	502	920	1,928	43,136	3.04	10	494	912	1,899	30,807	3.05
4th Quintile	13,171	16	474	732	1,182	7,853	1.53	16	609	1,110	2,291	45,296	2.81	21	585	1,083	2,279	44,158	2.81
5th Quintile	10,857	25	534	816	1,257	8,359	1.42	25	666	1,227	2,577	48,558	2.64	19	676	1,255	2,624	33,561	2.63
Population Density																			
Large MSA[1]	25,508	9	578	849	1,343	8,359	1.41	12	731	1,298	2,664	48,558	2.55	10	728	1,311	2,650	35,038	2.54
Medium to Small MSA[1]	30,307	18	315	558	886	8,444	1.62	18	458	838	1,765	43,516	3.06	8	450	843	1,770	30,041	3.06
Non-MSA,[1] Urban	1,249	20	207	512	791	3,628	1.63	20	382	770	1,684	16,479	2.97	45	359	792	1,711	17,924	3.05
Non-MSA,[1] Rural	3,557	18	199	469	802	4,220	1.71	19	335	772	1,692	45,296	3.44	23	330	759	1,648	44,158	3.51
Group Quarters																			
Group	774	25	219	459	986	4,098	1.90	25	275	592	1,269	30,862	4.92	27	264	571	1,225	20,208	3.38
Non-Group	59,847	9	439	665	1,099	8,444	1.56	12	556	1,019	2,136	48,558	2.85	8	548	1,021	2,153	44,158	2.84
Household Size																			
One	7,063	23	404	638	1,029	6,552	1.52	69	998	2,660	6,271	48,558	2.16	55	976	2,629	6,225	44,158	2.16
Two	21,700	19	428	649	1,055	8,236	1.54	20	706	1,490	3,530	45,473	2.21	14	704	1,481	3,501	33,561	2.20
Three	17,330	18	446	672	1,105	8,444	1.55	18	475	785	1,353	24,578	2.07	21	468	789	1,369	15,882	2.05
Four or More	14,528	9	442	694	1,185	8,359	1.60	12	454	729	1,268	14,597	1.78	8	439	728	1,262	13,876	1.79

[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 2012 NSDUH Respondent QDU-Level Weight Summary Statistics

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DUWT10*DUWT11)						After res.qdu.nr[1] (SDUWT*DUWT10*...*DUWT12)						Final Weight: After res.qdu.ps[1] (SDUWT*DUWT10*...*DUWT13)					
		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	48,850	8	527	971	1,975	44,158	2.90	8	627	1,175	2,485	64,497	3.21	6	626	1,178	2,487	63,870	3.22
Census Region																			
Northeast	9,917	8	450	803	1,716	29,264	3.02	8	521	988	2,287	47,848	3.34	6	513	985	2,308	45,340	3.34
South	15,019	10	788	1,254	2,452	44,158	2.59	20	911	1,479	2,980	64,497	2.89	25	912	1,482	2,975	63,870	2.89
Midwest	13,687	23	502	739	1,634	25,191	2.78	30	600	911	2,041	35,530	2.94	27	601	913	2,047	35,530	2.94
West	10,227	41	397	1,024	2,093	35,038	3.15	41	468	1,228	2,645	53,491	3.54	44	469	1,222	2,627	59,335	3.55
Quarter																			
Quarter 1	11,628	10	581	1,024	2,051	35,038	2.93	10	693	1,249	2,586	53,491	3.22	8	693	1,251	2,566	59,335	3.22
Quarter 2	12,975	8	484	914	1,850	27,818	2.91	8	577	1,102	2,336	37,706	3.18	6	577	1,106	2,336	37,778	3.19
Quarter 3	12,743	10	498	930	1,935	44,158	2.86	15	587	1,106	2,410	64,497	3.20	13	588	1,108	2,415	63,870	3.20
Quarter 4	11,504	13	571	1,020	2,096	30,014	2.88	14	681	1,230	2,660	47,848	3.21	10	678	1,231	2,658	45,340	3.21
Household Type																			
12-17, 18-25, 26+	5,004	13	451	680	1,156	9,750	1.57	14	508	781	1,340	9,750	1.61	10	506	781	1,342	9,750	1.60
12-17, 18-25	72	27	301	536	1,099	2,529	1.69	27	329	701	1,375	3,411	1.68	22	321	697	1,349	3,856	1.74
12-17, 26+	14,625	10	398	632	1,081	8,263	1.57	10	458	747	1,266	10,398	1.60	8	457	748	1,267	12,788	1.61
18-25, 26+	10,290	8	558	903	1,430	9,930	1.54	8	673	1,112	1,778	15,051	1.60	6	673	1,115	1,775	20,359	1.62
12-17	24	116	279	611	1,208	1,725	1.48	124	295	836	1,446	2,127	1.46	121	301	847	1,435	2,135	1.47
18-25	5,447	30	403	761	1,221	6,586	1.57	33	455	892	1,434	6,277	1.57	34	454	889	1,432	12,046	1.59
26+	13,388	53	1,827	3,412	5,696	44,158	1.73	57	2,383	4,556	7,834	64,497	1.78	44	2,384	4,544	7,837	63,870	1.78
Race/Ethnicity of Householder																			
Hispanic or Latino White	6,574	8	576	1,077	1,743	26,638	2.49	8	648	1,251	2,105	35,972	2.86	6	648	1,254	2,098	36,038	2.88
Hispanic or Latino Black or African American	165	20	796	1,789	3,943	29,264	2.99	20	885	2,126	4,681	47,848	3.39	25	1,009	2,201	4,493	45,340	3.38
Hispanic or Latino Other	355	10	296	982	2,419	33,561	4.10	28	313	1,051	2,782	44,608	4.82	30	330	1,045	2,888	38,613	4.48
Non-Hispanic or Latino White	32,098	10	513	942	2,043	44,158	2.94	10	621	1,165	2,641	64,497	3.20	8	620	1,168	2,641	63,870	3.20
Non-Hispanic or Latino Black or African American	6,092	34	647	1,048	2,037	28,858	2.69	36	718	1,182	2,379	34,671	3.01	36	720	1,190	2,382	34,627	3.02
Non-Hispanic or Latino Other	3,566	23	344	833	1,782	35,038	3.05	27	408	1,040	2,373	53,491	3.63	22	397	1,028	2,369	59,335	3.69
% Hispanic or Latino in Segment																			
50-100%	3,397	33	731	1,300	2,006	26,638	2.48	33	845	1,510	2,440	35,972	2.78	35	850	1,502	2,433	36,038	2.79
10-<50%	9,003	10	689	1,277	2,377	29,264	2.58	20	801	1,528	3,041	47,848	2.89	23	801	1,536	3,041	45,340	2.89
<10%	36,450	8	488	880	1,851	44,158	3.04	8	580	1,069	2,333	64,497	3.35	6	580	1,069	2,333	63,870	3.35

Table G.2 2012 NSDUH Respondent QDU-Level Weight Summary Statistics (continued)

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DUWT10*DUWT11)						After res.qdu.nr[1] (SDUWT*DUWT10*...*DUWT12)						Final Weight: After res.qdu.ps[1] (SDUWT*DUWT10*...*DUWT13)					
		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%	3,749	10	639	1,022	1,988	28,858	2.89	27	707	1,156	2,281	40,699	3.28	30	702	1,161	2,277	38,613	3.27
10-<50%	7,524	20	665	1,139	2,230	29,264	2.64	20	775	1,363	2,762	47,848	2.97	23	775	1,365	2,758	45,340	2.97
<10%	37,577	8	494	929	1,922	44,158	2.96	8	590	1,139	2,442	64,497	3.26	6	590	1,140	2,444	63,870	3.26
% Owner-Occupied DUs[1] in Segment																			
50-100%	36,704	8	526	967	1,972	44,158	2.89	8	632	1,180	2,510	64,497	3.19	6	631	1,181	2,510	63,870	3.19
10-<50%	9,421	23	517	954	1,924	35,038	3.04	26	594	1,130	2,309	53,491	3.38	17	594	1,131	2,312	59,335	3.40
<10%	2,725	10	574	1,109	2,223	27,818	2.67	20	646	1,292	2,720	39,554	2.93	25	634	1,293	2,718	37,754	2.93
Combined Median Rent/Housing Value																			
1st Quintile	8,288	8	487	871	1,784	28,858	2.86	8	563	1,014	2,149	35,972	3.16	6	564	1,010	2,139	36,038	3.17
2nd Quintile	10,661	10	499	915	1,897	26,430	2.92	10	582	1,085	2,290	40,699	3.24	8	582	1,086	2,287	38,613	3.24
3rd Quintile	11,009	10	481	883	1,765	29,155	3.06	20	569	1,069	2,237	39,581	3.41	20	568	1,072	2,242	39,310	3.41
4th Quintile	10,428	23	567	1,044	2,104	44,158	2.89	30	678	1,271	2,683	64,497	3.15	26	673	1,271	2,689	63,870	3.16
5th Quintile	8,464	19	646	1,194	2,378	33,561	2.67	20	792	1,491	3,123	44,608	2.93	18	789	1,491	3,109	40,087	2.93
Population Density																			
Large MSA[1]	20,223	10	704	1,243	2,431	35,038	2.60	20	843	1,499	3,145	53,491	2.88	17	843	1,502	3,128	59,335	2.88
Medium to Small MSA[1]	24,705	8	438	816	1,667	29,155	3.10	8	514	975	2,054	39,581	3.39	6	513	977	2,056	39,310	3.39
Non-MSA,[1] Urban	1,028	45	341	777	1,603	17,924	3.10	50	382	893	1,944	22,067	3.35	50	382	898	1,954	22,001	3.35
Non-MSA,[1] Rural	2,894	23	322	711	1,560	44,158	3.51	23	370	865	1,942	64,497	3.87	20	372	865	1,944	63,870	3.86
Group Quarters																			
Group	729	27	267	561	1,225	20,208	3.45	27	283	590	1,343	20,208	3.57	22	279	572	1,290	19,652	3.69
Non-Group	48,121	8	531	979	1,991	44,158	2.89	8	633	1,184	2,516	64,497	3.20	6	632	1,187	2,515	63,870	3.20
Household Size																			
One	5,567	63	917	2,379	5,997	44,158	2.22	72	1,114	2,949	8,054	64,497	2.34	83	1,114	2,965	8,041	63,870	2.35
Two	16,974	14	670	1,361	3,229	33,561	2.25	15	805	1,679	4,247	44,608	2.44	13	803	1,678	4,234	38,613	2.43
Three	14,071	23	459	775	1,341	15,882	2.01	27	543	938	1,617	23,531	2.32	22	544	940	1,621	24,616	2.34
Four or More	12,238	8	431	721	1,245	13,876	1.77	8	497	845	1,459	24,188	2.02	6	493	844	1,461	20,300	2.03

[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

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 2012, 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: Non-Group 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: Two or More Races 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: Two or More Races 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

1: 50-100%, 2: 10-<50%, 3: 0-<10%[1]

Percentage of Segments That Are Hispanic or Latino

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 Other, 4: Non-Hispanic or Latino White, 5: Non-Hispanic or Latino Black or African American, 6: Non-Hispanic or Latino Other

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 percentage 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 2012 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) × Pair Age (6 Levels)	5 × 6	20
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8
Pair Gender × Pair Age (6 Levels)	3 × 6	10
State/Region × Pair Race/Ethnicity (5 Levels)	Model-specific	
State/Region × Pair Age (6 Levels)	Model-specific	
State/Region × Pair Gender	Model-specific	
Rent/Housing × % Black or African American	5 × 3	8
Rent/Housing × % Hispanic or Latino	5 × 3	8
Rent/Housing × % Owner-Occupied	5 × 3	8
% Owner-Occupied × % Black or African American	3 × 3	4
% Owner-Occupied × % Hispanic or Latino	3 × 3	4
Three-Factor Effects		
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 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 2012 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	5.38	23.39	12.45	67.5167	213	(0.26, 1.10)	(0.26, 1.10)
	1.86	7.18	1.41	8.5932	201	(0.21, 3.98)	(0.21, 3.96)
						(0.90, 1.60)	(0.90, 1.60)
res.pr.nr	1.74	6.26	1.31	8.9725	213	(1.00, 1.60)	(1.00, 1.60)
	2.09	9.12	1.85	9.3991	213	(1.00, 5.00)	(1.00, 5.00)
						N/A	N/A
res.pr.ps	2.09	9.39	1.81	9.3991	223	(0.39, 1.20)	(0.39, 1.20)
	1.47	5.44	0.55	9.6725	211	(0.25, 1.89)	(0.25, 1.89)
						N/A	N/A
res.pr.ev	1.47	5.44	0.55	9.6725	223	(0.94, 1.30)	(0.96, 1.30)
	0.81	5.70	0.19	9.6484	211	(0.80, 1.44)	(0.85, 1.28)
						N/A	N/A

GEM = generalized exponential model.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

[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 is for nonextreme values, and the third is for low extreme values.

Table H.1b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Northeast and South)

	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	9	1.02	20	0.01	8	0.41	10	0.17	4	0.61	3
1%	63	1.11	131	0.22	62	0.85	65	0.30	53	0.88	51
5%	127	1.20	355	0.31	216	1.00	229	0.49	189	0.95	184
10%	214	1.31	645	0.41	410	1.01	433	0.62	382	0.96	385
25%	498	1.51	1,307	0.63	1,075	1.04	1,145	0.84	1,091	0.98	1,085
Median	781	5.52	3,306	0.94	3,088	1.16	3,462	1.01	3,367	1.00	3,360
75%	1,185	11.85	8,863	1.30	8,660	1.42	10,165	1.15	9,887	1.01	9,890
90%	1,629	22.86	18,678	1.75	21,257	1.88	29,270	1.37	28,709	1.03	28,478
95%	2,096	28.63	28,096	2.13	33,961	2.33	51,092	1.52	51,237	1.04	51,790
99%	3,206	53.82	60,653	2.89	75,394	3.96	135,350	1.69	138,677	1.11	139,613
Maximum	7,138	4,403.70	6,068,764	3.96	1,053,229	5.00	1,323,156	1.89	1,234,015	1.28	1,237,337
n	13,619	-	13,619	-	13,619	-	9,723	-	9,723	-	9,723
Mean	905	10.01	8,700	1.03	8,841	1.34	12,384	1.00	12,384	1.00	12,384
Max/Mean	8	-	698	-	119	-	107	-	100	-	100

SDU = screener dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

Model Group 1 Overview

Selected Pair-Level Poststratification

In the selected pair-level poststratification step, 201 of 213 proposed factors were retained in the final model. All main and two-factor effects were retained. All three-factor effects were dropped due to convergence problems.

Respondent Pair-Level Nonresponse

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

Respondent Pair-Level Poststratification

In the respondent pair-level poststratification step, 211 of 223 proposed factors were retained in the final model. All main and two-factor effects were retained. All three-factor effects were dropped due to convergence problems.

Respondent Pair-Level Extreme Value Adjustment

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

Exhibit H.1.1 Covariates for 2012 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	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) × Pair Age (6 Levels)	5 × 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8	8	All levels present.
Pair Gender × Pair Age (6 Levels)	3 × 6	10	10	All levels present.
State/Region × Pair Race/Ethnicity (5 Levels)	6 × 5	20	20	All levels present.
State/Region × Pair Age (6 Levels)	6 × 6	25	25	All levels present.
State/Region × Pair Gender	6 × 3	10	10	All levels present.
Rent/Housing × % Black or African American	5 × 3	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	5 × 3	8	8	All levels present.
Rent/Housing × % Owner-Occupied	5 × 3	8	8	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12	0	Drop all; conv.
Total		**213**	**201**	

Exhibit H.1.2 Covariates for 2012 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) × Pair Age (6 Levels)	5 × 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8	8	All levels present.
Pair Gender × Pair Age (6 Levels)	3 × 6	10	10	All levels present.
State/Region × Pair Race/Ethnicity (5 Levels)	6 × 5	20	20	All levels present.
State/Region × Pair Age (6 Levels)	6 × 6	25	25	All levels present.
State/Region × Pair Gender	6 × 3	10	10	All levels present.
Rent/Housing × % Black or African American	5 × 3	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	5 × 3	8	8	All levels present.
Rent/Housing × % Owner-Occupied	5 × 3	8	8	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
Three-Factor Effects		**12**	**12**	
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12	12	All levels present.
Total		**213**	**213**	

Exhibit H.1.3 Covariates for 2012 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	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.
Pair Relationship	10	10	10	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) × Pair Age (6 Levels)	5 × 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8	8	All levels present.
Pair Gender × Pair Age (6 Levels)	3 × 6	10	10	All levels present.
State/Region × Pair Race/Ethnicity (5 Levels)	6 × 5	20	20	All levels present.
State/Region × Pair Age (6 Levels)	6 × 6	25	25	All levels present.
State/Region × Pair Gender	6 × 3	10	10	All levels present.
Rent/Housing × % Black or African American	5 × 3	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	5 × 3	8	8	All levels present.
Rent/Housing × % Owner-Occupied	5 × 3	8	8	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12	0	Drop all; conv.
Total		**223**	**211**	

Exhibit H.1.4 Covariates for 2012 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 2012 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	5.02	17.79	6.19	16.0574	212	(0.41, 2.00)	(0.42, 2.00)
	2.08	6.47	1.21	10.0819	200	(0.24, 3.46)	(0.25, 3.45)
						(0.90, 1.73)	(.0.90, 1.73)
res.pr.nr	2.31	9.13	1.90	10.7706	212	(1.01, 2.70)	(1.01, 2.70)
	2.87	11.04	2.21	11.1015	212	(1.00, 5.00)	(1.00, 5.00)
						N/A	N/A
res.pr.ps	2.83	12.68	3.78	11.1015	222	(0.56, 1.90)	(0.58, 1.90)
	2.25	12.52	2.14	9.2899	210	(0.43, 2.50)	(0.45, 2.49)
						N/A	N/A
res.pr.ev	2.25	12.52	2.14	9.2899	222	(0.80, 1.80)	(0.94, 1.77)
	1.04	7.73	0.48	8.7981	210	(0.70, 1.80)	(0.80, 1.30)
						N/A	N/A

GEM = generalized exponential model.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

[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 is for nonextreme values, and the third is for low extreme values.

Table H.2b 2012 Distribution of Weight Adjustment Factors and Weight Products (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	20	1.02	64	0.06	29	0.50	30	0.11	18	0.52	16
1%	92	1.10	161	0.31	132	0.99	153	0.50	124	0.83	118
5%	132	1.21	336	0.47	293	1.00	334	0.59	292	0.89	279
10%	175	1.32	550	0.58	470	1.01	518	0.66	474	0.92	462
25%	405	1.49	964	0.78	984	1.04	1,115	0.82	1,073	0.96	1,056
Median	577	5.26	2,692	1.01	2,564	1.15	2,913	0.97	2,888	0.99	2,861
75%	1,045	11.43	7,082	1.28	7,159	1.43	8,363	1.15	8,574	1.02	8,610
90%	1,691	21.68	15,760	1.59	17,689	1.93	22,988	1.45	23,090	1.06	23,367
95%	1,995	28.31	24,872	1.79	27,387	2.45	41,048	1.66	40,453	1.09	40,772
99%	2,664	54.34	57,015	2.42	67,482	4.09	118,207	2.03	122,494	1.14	121,848
Maximum	8,444	1,603.54	1,594,823	3.45	1,028,039	5.00	1,550,385	2.49	804,009	1.30	639,872
n	13,416	-	13,416	-	13,416	-	9,736	-	9,736	-	9,736
Mean	787	9.09	7,314	1.06	7,569	1.36	10,430	1.02	10,430	0.99	10,430
Max/Mean	11	-	218	-	136	-	149	-	77	-	61

SDU = screener dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

Model Group 2 Overview

Selected Pair-Level Poststratification

In the selected pair-level poststratification step, 200 of 212 proposed factors were retained in the final model. All main and two-factor effects were retained at proposed levels. None of the 12 three-factor effects were kept in the model due to convergence problems.

Respondent Pair-Level Nonresponse

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

Respondent Pair-Level Poststratification

In the respondent pair-level poststratification step, 210 of 222 proposed factors were retained in the final model, as in the selected pair-level poststratification step.

Respondent Pair-Level Extreme Value Adjustment

In the respondent pair-level extreme value adjustment step, 210 of 222 proposed factors were retained in the final model, as in the respondent pair-level poststratification step.

Exhibit H.2.1 Covariates for 2012 NSDUH Pair Weights (sel.pr.ps) Model Group 2: Midwest and West

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**75**	**75**	
Intercept	1	1	1	All levels present.
State	25	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.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) × Pair Age (6 Levels)	5 × 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8	8	All levels present.
Pair Gender × Pair Age (6 Levels)	3 × 6	10	10	All levels present.
State/Region × Pair Race/Ethnicity (5 Levels)	6 × 5	20	20	All levels present.
State/Region × Pair Age (6 Levels)	6 × 6	25	25	All levels present.
State/Region × Pair Gender	6 × 3	10	10	All levels present.
Rent/Housing × % Black or African American	5 × 3	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	5 × 3	8	8	All levels present.
Rent/Housing × % Owner-Occupied	5 × 3	8	8	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12	0	Drop all; conv.
Total		**212**	**200**	

Exhibit H.2.2 Covariates for 2012 NSDUH Pair Weights (res.pr.nr) Model Group 2: Midwest and West

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**75**	**75**	
Intercept	1	1	1	All levels present.
State	25	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.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) × Pair Age (6 Levels)	5 × 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8	8	All levels present.
Pair Gender × Pair Age (6 Levels)	3 × 6	10	10	All levels present.
State/Region × Pair Race/Ethnicity (5 Levels)	6 × 5	20	20	All levels present.
State/Region × Pair Age (6 Levels)	6 × 6	25	25	All levels present.
State/Region × Pair Gender	6 × 3	10	10	All levels present.
Rent/Housing × % Black or African American	5 × 3	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	5 × 3	8	8	All levels present.
Rent/Housing × % Owner-Occupied	5 × 3	8	8	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
Three-Factor Effects		**12**	**12**	
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12	12	All levels present.
Total		**212**	**212**	

Exhibit H.2.3 Covariates for 2012 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) × Pair Age (6 Levels)	5 × 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8	8	All levels present.
Pair Gender × Pair Age (6 Levels)	3 × 6	10	10	All levels present.
State/Region × Pair Race/Ethnicity (5 Levels)	6 × 5	20	20	All levels present.
State/Region × Pair Age (6 Levels)	6 × 6	25	25	All levels present.
State/Region × Pair Gender	6 × 3	10	10	All levels present.
Rent/Housing × % Black or African American	5 × 3	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	5 × 3	8	8	All levels present.
Rent/Housing × % Owner-Occupied	5 × 3	8	8	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12	0	Drop all; conv.
Total		**222**	**210**	

Exhibit H.2.4 Covariates for 2012 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 2012 NSDUH Person Pair-Level Response Rates

Domain	Selected Pairs	Respondent Pairs	% Interview Response Rate[1]
Total	27,035	19,459	63.76
Pair Age Group			
12-17, 12-17	4,507	3,666	81.51
12-17, 18-25	3,627	2,778	77.17
12-17, 26-34	825	653	77.42
12-17, 35-49	3,813	2,816	73.86
12-17, 50+	851	627	74.42
18-25, 18-25	5,476	3,976	73.18
18-25, 26-34	1,079	742	67.14
18-25, 35-49	1,582	1,058	65.84
18-25, 50+	1,074	650	58.92
26-34, 26-34	880	597	71.06
26-34, 35-49	469	305	52.70
26-34, 50+	315	178	53.89
35-49, 35-49	833	489	61.19
35-49, 50+	466	266	50.91
50+, 50+	1,238	658	52.22
Pair Race/Ethnicity			
Hispanic or Latino	4,176	3,084	69.40
Black or African American	2,760	2,210	71.36
White	16,039	11,284	61.92
Other	1,791	1,175	50.45
White & Black or African American	217	166	68.55
White & Hispanic or Latino	921	692	71.01
White & Other	737	552	63.97
Black or African American & Hispanic or Latino	123	91	69.54
Black or African American & Other	126	99	37.40
Hispanic or Latino & Other	145	106	58.42
Pair Gender			
Male, Male	5,928	4,173	60.83
Female, Female	5,813	4,425	67.33
Male, Female	15,294	10,861	63.58
Household Size			
Two	6,855	4,660	59.76
Three	7,398	5,248	60.52
Four or More	12,782	9,551	67.43

Table I.1 2012 NSDUH Person Pair-Level Response Rates (continued)

Domain	Selected Pairs	Respondent Pairs	% Interview Response Rate[1]
Census Region			
Northeast	5,685	3,856	58.61
South	7,934	5,867	65.95
Midwest	7,515	5,455	66.14
West	5,901	4,281	62.52
Quarter			
Quarter 1	6,334	4,613	64.66
Quarter 2	7,183	5,115	63.01
Quarter 3	7,012	5,093	64.85
Quarter 4	6,506	4,638	62.52
% Hispanic or Latino in Segment			
50-100%	2,105	1,537	65.52
10-<50%	5,112	3,726	64.88
<10%	19,818	14,196	63.07
% Black or African American in Segment			
50-100%	1,905	1,482	64.05
10-<50%	4,030	2,974	66.88
<10%	21,100	15,003	63.01
% Owner-Occupied DUs in Segment			
50-100%	20,778	14,866	63.43
10-<50%	4,889	3,600	64.61
<10%	1,368	993	71.03
Combined Median Rent/Housing Value			
1st Quintile	4,264	3,227	68.88
2nd Quintile	5,758	4,250	67.80
3rd Quintile	6,136	4,364	65.47
4th Quintile	6,013	4,276	58.78
5th Quintile	4,864	3,342	60.58
Population Density			
Large MSA	11,649	8,192	61.00
Medium to Small MSA	13,392	9,833	67.16
Non-MSA, Urban	511	380	71.70
Non-MSA, Rural	1,483	1,054	65.73
Group Quarters			
Group	386	312	79.80
Non-Group	26,649	19,147	63.70

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 WT1*…*WT9*PRWT10*PRWT11.

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

Table J.1 2012 NSDUH Selected Pair-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)			Before sel.pr.ps[1] (SDUWT*PRWT10)			After sel.pr.ps[1] (SDUWT*PRWT10*PRWT11)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Total	27,035	2.14	5.22	1.44	5.34	26.13	15.62	2.15	15.29	7.51
Pair Age Group										
12-17, 12-17	4,507	1.84	4.16	1.29	3.64	17.26	6.55	0.55	3.06	0.64
12-17, 18-25	3,627	2.21	5.76	1.79	8.08	28.38	11.52	1.96	7.45	1.53
12-17, 26-34	825	1.82	4.03	1.52	1.82	5.22	1.36	0.85	2.56	0.36
12-17, 35-49	3,813	1.60	3.70	0.89	2.12	9.32	2.46	0.87	3.65	0.67
12-17, 50+	851	2.12	6.13	1.48	2.23	10.42	3.28	0.59	0.89	0.17
18-25, 18-25	5,476	2.12	4.99	1.22	8.16	27.90	10.87	3.67	12.83	2.39
18-25, 26-34	1,079	3.61	7.88	1.97	3.15	13.23	4.97	2.32	8.97	1.53
18-25, 35-49	1,582	3.41	9.38	3.19	6.76	26.19	11.00	2.40	7.55	1.43
18-25, 50+	1,074	2.33	5.87	1.61	3.82	16.98	6.23	1.30	3.90	0.59
26-34, 26-34	880	2.73	5.15	1.11	3.64	21.67	12.88	2.84	21.62	10.40
26-34, 35-49	469	2.77	5.47	1.26	7.46	43.03	29.62	5.97	36.72	19.45
26-34, 50+	315	2.22	5.94	2.07	4.44	22.48	12.39	1.59	16.78	12.07
35-49, 35-49	833	2.04	6.54	2.01	6.24	44.99	30.94	4.80	45.93	30.71
35-49, 50+	466	1.50	4.26	0.59	1.93	20.54	15.31	2.15	21.08	13.47
50+, 50+	1,238	1.62	2.89	0.40	8.24	43.73	33.58	4.36	21.24	8.98
Pair Race/Ethnicity										
Hispanic or Latino	4,176	3.86	10.45	3.98	5.17	35.09	24.22	2.95	22.46	11.57
Black or African American	2,760	3.44	5.97	1.01	6.63	33.29	22.33	3.22	20.24	10.19
White	16,039	0.68	1.49	0.19	4.65	21.19	12.31	1.35	11.75	6.17
Other	1,791	6.03	12.30	2.98	7.76	28.26	10.20	3.35	15.98	5.08
White & Black or African American	217	6.45	9.80	1.92	9.68	16.84	6.21	13.82	19.84	3.28
White & Hispanic or Latino	921	2.71	5.44	1.05	6.84	21.82	7.40	3.04	11.85	3.66
White & Other	737	3.53	6.72	1.57	4.75	18.77	8.09	2.04	14.84	7.83
Black or African American & Hispanic or Latino	123	17.89	43.36	18.62	17.89	66.55	44.19	6.50	39.53	17.08
Black or African American & Other	126	5.56	12.08	2.35	5.56	40.75	27.57	1.59	34.09	21.52
Hispanic or Latino & Other	145	8.28	16.29	5.15	8.97	21.25	9.63	6.90	7.90	1.71
Pair Gender										
Male, Male	5,928	2.14	4.77	1.29	7.25	31.61	20.02	2.65	11.86	5.08
Female, Female	5,813	2.34	5.51	1.66	5.33	18.19	7.47	2.31	11.75	3.90
Male, Female	15,294	2.07	5.28	1.42	4.61	26.61	16.48	1.90	17.24	9.21
Household Size										
Two	6,855	1.74	4.34	1.11	0.83	2.47	0.67	0.44	2.29	0.31
Three	7,398	2.10	5.32	1.69	2.61	27.75	19.75	1.82	16.75	7.66
Four or More	12,782	2.39	5.59	1.47	9.35	37.59	21.18	3.25	20.99	11.01

Table J.1 2012 NSDUH Selected Pair-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)			Before sel.pr.ps[1] (SDUWT*PRWT10)			After sel.pr.ps[1] (SDUWT*PRWT10*PRWT11)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Census Region										
Northeast	5,685	2.62	6.98	2.28	5.82	21.94	9.44	2.36	15.05	6.33
South	7,934	1.78	4.49	1.05	5.29	30.24	20.63	1.79	15.23	7.69
Midwest	7,515	1.77	3.61	0.66	5.59	26.11	14.87	2.13	14.43	7.58
West	5,901	2.64	6.43	2.11	4.64	22.62	12.66	2.46	16.23	8.07
Quarter										
Quarter 1	6,334	2.07	4.60	1.28	5.32	29.27	20.03	2.68	13.34	5.68
Quarter 2	7,183	2.03	5.25	1.58	4.82	18.50	7.62	2.05	11.60	3.85
Quarter 3	7,012	2.05	4.87	1.15	5.03	25.81	14.78	1.64	19.10	10.98
Quarter 4	6,506	2.43	6.14	1.77	6.29	30.19	19.30	2.29	17.08	9.51
% Hispanic or Latino in Segment										
50-100%	2,105	2.28	5.26	1.89	4.42	36.21	27.31	2.23	17.79	8.80
10-<50%	5,112	3.66	9.15	3.09	5.97	33.28	22.09	3.01	18.47	9.83
<10%	19,818	1.74	3.74	0.76	5.28	21.35	10.80	1.92	13.74	6.47
% Black or African American in Segment										
50-100%	1,905	3.67	8.74	2.71	6.72	27.02	13.67	3.78	19.57	8.12
10-<50%	4,030	2.68	6.54	2.12	6.23	36.11	25.20	2.70	16.10	7.76
<10%	21,100	1.90	4.52	1.14	5.05	23.47	13.34	1.90	14.68	7.40
% Owner-Occupied DUs[1] in Segment										
50-100%	20,778	1.90	4.79	1.30	5.09	27.19	16.70	2.19	15.79	7.99
10-<50%	4,889	2.88	6.37	1.88	6.14	22.44	12.07	2.39	14.14	5.88
<10%	1,368	3.14	7.07	1.97	6.36	18.81	7.21	0.66	2.43	0.34
Combined Median Rent/Housing Value										
1st Quintile	4,264	1.74	4.47	1.35	5.42	39.11	29.10	2.11	20.67	12.05
2nd Quintile	5,758	2.15	5.91	1.99	5.45	22.10	9.42	2.57	16.93	7.54
3rd Quintile	6,136	2.22	4.62	1.18	5.30	23.81	13.29	1.65	14.94	8.36
4th Quintile	6,013	2.33	5.21	1.39	5.32	26.61	17.44	2.06	14.47	7.09
5th Quintile	4,864	2.16	5.64	1.28	5.24	20.76	10.26	2.43	11.49	4.17
Population Density										
Large MSA[1]	11,649	2.83	6.69	1.90	5.55	27.46	16.79	2.59	16.62	8.29
Medium to Small MSA[1]	13,392	1.68	3.67	0.97	5.39	19.31	7.73	1.92	12.88	5.52
Non-MSA,[1] Urban	511	0.78	0.68	0.13	3.33	72.80	70.94	1.37	23.83	19.61
Non-MSA,[1] Rural	1,483	1.35	2.73	0.51	3.98	26.19	19.15	1.01	17.82	12.43
Group Quarters										
Group	386	2.59	3.81	0.78	8.03	23.60	6.46	6.99	19.26	3.27
Non-Group	26,649	2.14	5.23	1.45	5.31	26.14	15.65	2.08	15.28	7.53

[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 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Before res.pr.nr[1] (SDUWT*PRWT10*PRWT11)			After res.pr.nr[1] (SDUWT*PRWT10*...*PRWT12)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Total	19,459	2.16	15.29	7.43	2.61	15.86	6.50
Pair Age Group							
12-17, 12-17	3,666	0.46	2.36	0.35	0.38	3.16	0.97
12-17, 18-25	2,778	2.12	8.17	1.63	1.73	7.94	2.34
12-17, 26-34	653	1.07	4.90	0.74	1.38	7.70	1.95
12-17, 35-49	2,816	0.96	4.78	0.78	1.07	2.89	0.55
12-17, 50+	627	0.32	0.72	0.22	0.64	3.03	0.66
18-25, 18-25	3,976	3.65	13.28	2.63	4.45	17.62	3.91
18-25, 26-34	742	2.96	11.93	1.64	4.04	17.04	3.72
18-25, 35-49	1,058	2.46	7.42	1.52	4.73	15.74	3.50
18-25, 50+	650	1.23	2.95	0.54	2.00	5.21	0.72
26-34, 26-34	597	3.85	25.89	13.06	2.68	27.33	11.77
26-34, 35-49	305	6.89	29.07	11.40	10.49	34.39	14.18
26-34, 50+	178	2.25	20.31	12.72	1.69	10.63	7.06
35-49, 35-49	489	5.32	49.78	35.62	7.77	48.74	30.19
35-49, 50+	266	3.38	26.20	12.42	4.14	16.72	6.22
50+, 50+	658	3.80	21.67	11.23	4.86	17.34	6.29
Pair Race/Ethnicity							
Hispanic or Latino	3,084	2.95	26.88	14.07	3.44	23.61	11.28
Black or African American	2,210	3.21	18.30	8.75	2.08	15.61	7.34
White	11,284	1.27	10.54	5.41	2.03	12.64	5.04
Other	1,175	3.32	18.31	5.76	6.47	26.06	7.61
White & Black or African American	166	14.46	20.29	4.46	10.84	15.25	3.46
White & Hispanic or Latino	692	3.76	13.37	4.64	0.87	7.53	4.25
White & Other	552	1.99	6.27	1.52	1.09	13.41	2.08
Black or African American & Hispanic or Latino	91	5.49	29.46	19.46	7.69	27.76	12.80
Black or African American & Other	99	2.02	8.26	0.30	10.10	20.77	7.12
Hispanic or Latino & Other	106	8.49	12.57	2.94	2.83	10.03	1.10
Pair Gender							
Male, Male	4,173	2.64	10.72	3.25	3.50	15.22	4.05
Female, Female	4,425	2.31	8.63	2.15	2.10	9.70	2.24
Male, Female	10,861	1.92	18.50	10.11	2.47	17.77	8.38
Household Size							
Two	4,660	0.45	2.58	0.43	0.67	4.13	1.44
Three	5,248	1.91	17.38	8.33	2.76	19.63	7.37
Four or More	9,551	3.14	19.92	10.09	3.47	19.72	8.56

Table J.2 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.pr.nr[1] (SDUWT*PRWT10*PRWT11)			After res.pr.nr[1] (SDUWT*PRWT10*...*PRWT12)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Census Region							
Northeast	3,856	2.26	14.23	6.16	2.52	17.44	6.83
South	5,867	1.62	13.45	7.09	2.01	13.98	5.98
Midwest	5,455	2.11	15.39	8.50	2.88	19.13	8.55
West	4,281	2.90	18.66	7.89	3.15	14.76	5.35
Quarter							
Quarter 1	4,613	2.64	13.92	5.48	3.14	12.82	4.46
Quarter 2	5,115	2.09	10.20	3.27	2.64	13.67	3.28
Quarter 3	5,093	1.45	18.44	11.32	1.65	18.64	9.28
Quarter 4	4,638	2.54	18.57	9.55	3.08	18.26	8.94
% Hispanic or Latino in Segment							
50-100%	1,537	2.93	21.09	10.38	2.80	17.33	7.53
10-<50%	3,726	3.46	21.35	10.55	3.65	19.43	8.57
<10%	14,196	1.74	12.08	5.77	2.31	14.34	5.58
% Black or African American in Segment							
50-100%	1,482	3.71	15.16	4.95	3.78	15.80	4.57
10-<50%	2,974	3.19	19.03	9.57	3.30	17.37	7.61
<10%	15,003	1.81	14.40	7.16	2.35	15.52	6.44
% Owner-Occupied DUs[1] in Segment							
50-100%	14,866	2.07	15.63	7.94	2.52	15.96	6.65
10-<50%	3,600	2.86	14.94	5.66	3.36	15.87	6.06
<10%	993	1.01	4.00	0.63	1.21	11.38	3.21
Combined Median Rent/Housing Value							
1st Quintile	3,227	2.17	22.21	14.19	2.14	23.88	13.52
2nd Quintile	4,250	2.49	18.39	8.55	2.49	18.83	7.49
3rd Quintile	4,364	1.81	16.57	8.97	2.47	17.25	7.64
4th Quintile	4,276	2.03	9.98	2.44	3.04	10.26	2.39
5th Quintile	3,342	2.36	11.42	5.10	2.81	12.86	4.56
Population Density							
Large MSA[1]	8,192	2.70	16.16	7.80	3.33	15.99	6.41
Medium to Small MSA[1]	9,833	1.88	13.18	5.37	2.16	14.08	5.06
Non-MSA,[1] Urban	380	1.58	33.21	27.29	1.84	33.42	22.73
Non-MSA,[1] Rural	1,054	0.85	18.99	16.13	1.42	25.72	16.42
Group Quarters							
Group	312	8.33	25.50	3.58	5.77	23.85	6.74
Non-Group	19,147	2.06	15.25	7.44	2.55	15.83	6.50

[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 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Before res.pr.ps[1] (SDUWT*PRWT10*...*PRWT12)			After res.pr.ps[1] (SDUWT*PRWT10*...*PRWT13)			Final Weight: After res.pr.ev[1] (SDUWT*PRWT10*...*PRWT14)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Total	19,459	2.46	10.90	2.71	1.86	8.68	1.28	0.93	5.01	0.33
Pair Age Group										
12-17, 12-17	3,668	0.41	3.31	1.07	0.41	1.82	0.25	0.27	1.52	0.16
12-17, 18-25	2,759	1.70	7.95	2.36	0.91	2.82	0.58	0.69	2.47	0.36
12-17, 26-34	658	1.37	7.58	1.92	1.22	3.40	0.26	0.61	1.83	0.12
12-17, 35-49	2,812	1.07	2.84	0.53	1.21	3.50	0.58	0.71	2.08	0.20
12-17, 50+	631	0.63	3.04	0.65	0.95	5.54	1.15	0.48	2.34	0.23
18-25, 18-25	3,901	4.43	17.14	3.78	3.05	10.07	1.19	0.95	3.40	0.24
18-25, 26-34	794	3.78	18.47	7.16	3.78	13.35	1.57	1.76	5.64	0.41
18-25, 35-49	1,053	5.03	16.30	3.45	5.13	14.84	2.08	2.28	6.87	0.62
18-25, 50+	660	1.97	5.13	0.75	2.58	10.37	1.51	1.52	3.63	0.23
26-34, 26-34	604	1.66	10.16	3.30	0.50	7.55	1.26	0.50	7.77	0.36
26-34, 35-49	320	8.13	15.90	5.30	1.25	3.16	0.33	0.94	2.71	0.12
26-34, 50+	177	2.26	13.69	2.88	0.56	5.68	0.28	0.00	0.00	0.00
35-49, 35-49	487	5.95	24.71	8.32	1.23	7.32	1.46	0.82	5.68	0.19
35-49, 50+	272	2.57	12.66	2.81	1.84	9.61	1.24	0.74	3.39	0.18
50+, 50+	663	4.37	9.62	1.32	5.28	14.60	2.27	4.07	11.55	0.69
Pair Race/Ethnicity										
Hispanic or Latino	3,134	3.06	16.96	3.48	1.79	9.82	1.35	1.40	5.41	0.52
Black or African American	2,139	2.10	7.08	1.49	1.92	6.00	0.63	0.70	1.36	0.14
White	11,006	1.92	8.98	2.34	1.33	6.65	1.04	0.31	3.11	0.15
Other	1,116	6.36	22.29	6.39	6.72	25.53	3.65	5.73	22.66	1.50
White & Black or African American	167	8.38	14.47	3.01	10.78	19.96	3.60	5.99	11.36	0.65
White & Hispanic or Latino	721	1.25	2.96	0.51	1.94	11.77	1.94	0.55	8.94	0.15
White & Other	715	0.98	2.51	0.70	0.28	0.80	0.02	0.00	0.00	0.00
Black or African American & Hispanic or Latino	109	9.17	15.51	7.09	5.50	6.74	1.02	7.34	7.51	0.66
Black or African American & Other	198	6.06	23.35	8.75	1.52	8.09	2.89	0.51	5.37	1.63
Hispanic or Latino & Other	154	2.60	1.72	0.79	0.65	1.00	0.16	0.00	0.00	0.00
Pair Gender										
Male, Male	4,162	3.41	13.03	3.17	2.33	7.66	1.10	0.96	3.83	0.25
Female, Female	4,430	2.19	9.64	2.19	1.81	7.70	1.10	0.93	4.16	0.44
Male, Female	10,867	2.21	10.65	2.73	1.70	9.23	1.38	0.91	5.57	0.31
Household Size										
Two	4,660	0.64	3.96	1.39	0.43	1.72	0.21	0.34	1.53	0.14
Three	5,248	2.69	15.06	4.04	2.02	13.08	2.13	1.28	9.73	0.61
Four or More	9,551	3.22	12.17	2.68	2.47	9.83	1.36	1.02	4.27	0.27

Table J.3 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PRWT10*...*PRWT12)			After res.pr.ps[1] (SDUWT*PRWT10*...*PRWT13)			Final Weight: After res.pr.ev[1] (SDUWT*PRWT10*...*PRWT14)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Census Region										
Northeast	3,856	2.54	12.15	2.44	1.92	7.94	0.95	1.32	5.11	0.39
South	5,867	1.79	7.96	1.49	1.18	4.14	0.35	0.48	1.46	0.09
Midwest	5,455	2.64	14.44	4.77	2.42	11.90	2.00	1.06	6.70	0.47
West	4,281	3.08	11.28	2.99	2.03	13.00	2.25	1.00	8.56	0.49
Quarter										
Quarter 1	4,613	2.99	8.43	2.00	1.99	6.66	1.02	0.89	4.69	0.30
Quarter 2	5,115	2.35	9.58	1.92	1.62	7.03	1.10	0.92	3.44	0.28
Quarter 3	5,093	1.49	9.89	3.06	1.57	12.63	1.87	0.98	7.23	0.46
Quarter 4	4,638	3.13	15.67	3.86	2.31	8.33	1.11	0.91	4.65	0.26
% Hispanic or Latino in Segment										
50-100%	1,537	2.47	8.16	1.91	1.63	7.42	1.05	1.11	4.54	0.30
10-<50%	3,726	3.38	14.96	3.61	2.74	13.61	1.92	1.83	8.96	0.67
<10%	14,196	2.22	9.86	2.52	1.66	7.09	1.08	0.67	3.65	0.21
% Black or African American in Segment										
50-100%	1,482	3.58	11.09	2.83	2.50	8.02	0.97	1.62	3.27	0.34
10-<50%	2,974	3.03	10.73	2.90	2.35	9.89	1.39	1.68	5.52	0.35
<10%	15,003	2.24	10.91	2.66	1.70	8.46	1.28	0.71	5.06	0.32
% Owner-Occupied DUs[1] in Segment										
50-100%	14,866	2.36	10.81	2.54	1.59	8.49	1.26	0.73	4.76	0.30
10-<50%	3,600	3.25	11.32	3.53	2.86	9.19	1.26	1.53	5.55	0.39
<10%	993	1.11	11.07	3.09	2.32	12.84	2.43	1.61	11.60	1.12
Combined Median Rent/Housing Value										
1st Quintile	3,227	2.05	16.76	5.92	1.77	12.66	2.44	0.90	6.50	0.54
2nd Quintile	4,250	2.24	7.56	1.39	1.53	6.50	0.80	0.73	4.36	0.27
3rd Quintile	4,364	2.29	11.09	2.09	1.60	6.66	0.83	0.69	2.28	0.15
4th Quintile	4,276	2.92	9.78	2.64	2.27	7.94	1.11	1.05	4.95	0.28
5th Quintile	3,342	2.78	11.32	2.57	2.18	10.99	1.62	1.35	7.41	0.46
Population Density										
Large MSA[1]	8,192	3.15	12.57	2.97	2.45	9.78	1.33	1.39	5.05	0.36
Medium to Small MSA[1]	9,833	2.02	7.23	1.47	1.38	6.92	1.07	0.59	4.69	0.28
Non-MSA,[1] Urban	380	1.84	4.24	1.68	1.05	0.63	0.21	0.26	0.09	0.00
Non-MSA,[1] Rural	1,054	1.42	25.72	11.58	1.99	13.15	2.95	0.66	9.05	0.39
Group Quarters										
Group	312	6.73	23.62	6.88	8.97	20.47	3.20	2.88	10.63	0.39
Non-Group	19,147	2.39	10.85	2.70	1.74	8.63	1.27	0.89	4.99	0.33

Table J.3 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PRWT10*...*PRWT12)			After res.pr.ps[1] (SDUWT*PRWT10*...*PRWT13)			Final Weight: After res.pr.ev[1] (SDUWT*PRWT10*...*PRWT14)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Pair Relationship Domain[4]										
Parent-Child (12-14)	1,981	0.91	3.04	0.71	1.31	3.74	0.58	0.56	1.53	0.11
Parent-Child (12-17)	3,748	0.96	3.49	0.74	1.17	3.83	0.57	0.59	1.89	0.18
Parent-Child (12-20)	4,443	1.67	6.67	1.28	2.21	8.89	1.33	1.04	3.48	0.34
Sibling (12-14)-Sibling (15-17)	2,191	0.46	4.07	1.43	0.46	2.39	0.36	0.32	2.01	0.24
Sibling (12-17)-Sibling (18-25)	2,477	1.53	7.04	1.82	0.89	2.81	0.54	0.61	2.11	0.31
Spouse-Spouse/Partner-Partner	3,664	2.07	9.93	2.85	1.83	8.06	1.30	0.98	6.75	0.31
Spouse-Spouse/Partner-Partner with Children (Younger Than 18)	1,640	1.77	14.25	3.18	2.87	8.20	1.34	0.91	5.10	0.15

[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 2012 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	19,459	221,952,167	221,952,167	221,952,167	-0.00	-0.00
Pair Age Group						
12-17, 12-17	3,668	7,255,585	7,213,926	7,213,926	0.58	-0.00
12-17, 18-25	2,759	8,114,464	8,172,406	8,172,406	-0.71	-0.00
12-17, 26-34	658	5,360,122	5,281,913	5,281,913	1.48	-0.00
12-17, 35-49	2,812	29,558,177	29,592,182	29,592,182	-0.11	-0.00
12-17, 50+	631	11,948,454	11,999,450	11,999,450	-0.42	-0.00
18-25, 18-25	3,901	12,207,700	12,545,793	12,545,793	-2.69	-0.00
18-25, 26-34	794	8,056,376	7,391,400	7,391,400	9.00	-0.00
18-25, 35-49	1,053	17,787,086	17,720,889	17,720,889	0.37	-0.00
18-25, 50+	660	18,580,624	18,321,739	18,321,739	1.41	-0.00
26-34, 26-34	604	9,996,036	10,662,196	10,662,196	-6.25	0.00
26-34, 35-49	320	8,782,458	8,567,386	8,567,386	2.51	0.00
26-34, 50+	177	10,685,258	11,157,395	11,157,395	-4.23	0.00
35-49, 35-49	487	18,497,856	18,578,012	18,578,012	-0.43	0.00
35-49, 50+	272	16,664,510	16,640,517	16,640,517	0.14	-0.00
50+, 50+	663	38,457,461	38,106,962	38,106,962	0.92	0.00
Pair Race/Ethnicity						
Hispanic or Latino	3,134	38,960,071	38,601,525	38,601,525	0.93	-0.00
Black or African American	2,139	24,049,201	24,291,619	24,291,619	-1.00	-0.00
White	11,006	121,400,147	123,391,386	123,391,386	-1.61	0.00
Other	1,116	15,936,917	16,500,923	16,500,923	-3.42	-0.00
White & Black or African American	167	1,811,631	2,028,598	2,028,598	-10.70	0.00
White & Hispanic or Latino	721	7,863,668	8,002,876	8,002,876	-1.74	0.00
White & Other	715	6,519,769	5,364,414	5,364,414	21.54	0.00
Black or African American & Hispanic or Latino	109	1,375,212	1,561,670	1,561,670	-11.94	-0.00
Black or African American & Other	198	2,236,103	1,038,323	1,038,323	115.36	-0.00
Hispanic or Latino & Other	154	1,799,447	1,170,834	1,170,834	53.69	0.00
Pair Gender						
Male, Male	4,162	39,706,191	39,891,754	39,891,754	-0.47	-0.00
Female, Female	4,430	39,862,298	39,874,794	39,874,794	-0.03	-0.00
Male, Female	10,867	142,383,679	142,185,618	142,185,618	0.14	-0.00
Pair Relationship Domain[3,4,5]						
Parent-Child (12-14)*	1,981	11,386,620	12,505,760	12,505,760	-8.95	-0.00
Parent-Child (12-17)*	3,748	23,617,276	25,045,406	25,045,406	-5.70	-0.00
Parent-Child (15-17)*	1,767	12,230,656	12,539,646	12,539,646	-2.46	-0.00
Parent-Child (12-20)*	4,443	31,454,617	34,378,755	34,378,755	-8.51	-0.00
Parent*-Child (12-14)	1,981	18,350,450	19,342,709	19,342,709	-5.13	-0.00
Parent*-Child (12-17)	3,748	30,953,375	32,235,193	32,235,193	-3.98	-0.00
Parent*-Child (15-17)	1,767	19,026,718	19,334,414	19,163,807	-0.72	0.89
Parent*-Child (12-20)	4,443	38,356,007	40,347,812	40,347,812	-4.94	-0.00
Sibling (12-14)-Sibling (15-17)*	2,191	3,929,975	4,012,028	4,012,028	-2.05	-0.00
Sibling (12-17)-Sibling (18-25)*	2,477	5,978,196	6,172,680	6,172,680	-3.15	-0.00
Spouse-Spouse/Partner-Partner	3,664	70,208,199	71,421,020	71,421,020	-1.70	-0.00
Spouse-Spouse/Partner-Partner with Children (Younger Than 18)	1,640	22,125,791	29,600,927	29,600,927	-25.25	-0.00

Table K.1 2012 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	4,660	54,622,573	54,622,573	54,622,573	-0.00	-0.00
Three	5,248	57,411,679	57,411,679	57,411,679	-0.00	-0.00
Four or More	9,551	109,917,914	109,917,914	109,917,914	-0.00	-0.00
Census Region						
Northeast	3,856	41,025,207	41,025,207	41,025,207	-0.00	-0.00
South	5,867	79,383,669	79,383,669	79,383,669	0.00	-0.00
Midwest	5,455	44,933,034	44,933,034	44,933,034	-0.00	-0.00
West	4,281	56,610,256	56,610,256	56,610,256	-0.00	-0.00
Quarter						
Quarter 1	4,613	54,726,611	54,726,611	54,726,611	-0.00	-0.00
Quarter 2	5,115	55,791,727	55,791,727	55,791,727	-0.00	-0.00
Quarter 3	5,093	55,933,469	55,933,469	55,933,469	-0.00	-0.00
Quarter 4	4,638	55,500,360	55,500,360	55,500,360	-0.00	-0.00
% Hispanic or Latino in Segment						
50-100%	1,537	23,192,230	23,192,230	23,192,230	-0.00	-0.00
10-<50%	3,726	52,879,054	52,879,055	52,879,055	-0.00	-0.00
<10%	14,196	145,880,883	145,880,883	145,880,883	-0.00	-0.00
% Black or African American in Segment						
50-100%	1,482	16,732,625	16,732,625	16,732,625	-0.00	-0.00
10-<50%	2,974	38,150,858	38,150,858	38,150,858	-0.00	-0.00
<10%	15,003	167,068,684	167,068,684	167,068,684	-0.00	-0.00
% Owner-Occupied DUs in Segment						
50-100%	14,866	182,125,104	182,125,104	182,125,104	-0.00	-0.00
10-<50%	3,600	35,952,245	35,952,245	35,952,245	-0.00	-0.00
<10%	993	3,874,817	3,874,817	3,874,817	-0.00	-0.00
Combined Median Rent/Housing Value						
1st Quintile	3,227	30,573,140	30,573,140	30,573,140	-0.00	-0.00
2nd Quintile	4,250	44,525,897	44,525,897	44,525,897	-0.00	-0.00
3rd Quintile	4,364	46,327,833	46,327,833	46,327,833	-0.00	-0.00
4th Quintile	4,276	53,795,269	53,795,269	53,795,269	-0.00	-0.00
5th Quintile	3,342	46,730,029	46,730,029	46,730,029	-0.00	-0.00
Population Density						
Large MSA	8,192	122,562,456	122,562,456	122,562,456	-0.00	-0.00
Medium to Small MSA	9,833	87,447,088	87,447,088	87,447,088	-0.00	-0.00
Non-MSA, Urban	380	2,868,028	2,868,028	2,868,028	0.00	0.00
Non-MSA, Rural	1,054	9,074,595	9,074,595	9,074,595	-0.00	-0.00
Group Quarters						
Group	312	798,082	798,082	798,082	0.00	0.00
Non-Group	19,147	221,154,085	221,154,085	221,154,085	-0.00	-0.00

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

[1] WT1*...*WT10*PRWT11*...*PRWT13 (before person pair poststratification).

[2] WT1*...*WT10*PRWT11*...*PRWT14 (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 2012 NSDUH Selected Pair-Level Weight Summary Statistics

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)						Before sel.pr.ps[1] (SDUWT*PRWT10)						After sel.pr.ps[1] (SDUWT*PRWT10*PRWT11)					
		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	27,035	9	441	678	1,134	8,444	1.58	20	1,136	2,941	7,922	6,068,764	46.74	8	1,023	2,791	7,871	1,053,229	9.27
Pair Age Group																			
12-17, 12-17	4,507	16	399	609	1,042	7,853	1.62	20	582	1,017	1,841	56,843	3.05	8	516	998	1,969	25,287	2.40
12-17, 18-25	3,627	18	476	694	1,182	8,444	1.60	26	744	1,363	2,544	52,622	2.84	15	709	1,451	2,854	31,529	2.22
12-17, 26-34	825	32	397	675	1,131	5,004	1.53	248	2,487	4,469	7,333	70,973	2.29	143	1,949	3,759	7,455	69,227	2.63
12-17, 35-49	3,813	20	435	626	1,064	8,351	1.52	141	2,952	5,274	8,974	168,153	2.41	114	2,321	4,657	9,259	117,119	2.55
12-17, 50+	851	56	472	727	1,164	7,097	1.56	878	5,875	10,166	17,187	219,397	2.32	229	4,637	9,115	17,385	156,917	2.33
18-25, 18-25	5,476	9	432	704	1,207	7,653	1.62	32	798	1,392	2,461	41,249	2.70	15	589	1,253	2,936	39,303	2.52
18-25, 26-34	1,079	19	477	738	1,208	5,421	1.57	164	2,798	4,588	7,820	179,640	3.04	105	1,819	3,421	8,319	139,322	3.17
18-25, 35-49	1,582	34	491	711	1,196	7,847	1.61	489	3,467	6,420	12,481	245,571	3.24	180	2,804	5,903	13,746	105,508	2.60
18-25, 50+	1,074	28	508	704	1,240	6,541	1.54	377	6,270	9,858	17,777	412,510	3.25	282	5,644	10,249	20,203	187,095	2.45
26-34, 26-34	880	20	442	711	1,134	5,065	1.49	334	5,246	8,233	13,604	789,953	7.71	167	3,867	6,865	12,454	313,383	5.05
26-34, 35-49	469	43	478	731	1,206	3,895	1.46	571	6,016	9,618	16,559	1,594,823	18.62	412	5,391	9,152	16,689	440,141	5.90
26-34, 50+	315	43	466	757	1,126	7,250	1.57	1,140	11,808	19,298	29,840	554,381	3.33	488	10,916	21,855	40,294	701,763	3.70
35-49, 35-49	833	36	445	649	1,142	6,883	1.59	553	5,602	9,452	15,385	1,317,081	13.17	439	5,576	10,612	18,055	1,053,229	11.42
35-49, 50+	466	59	446	666	1,098	3,202	1.48	1,471	9,255	15,553	27,455	858,997	5.46	1,037	9,888	20,007	39,458	868,188	4.56
50+, 50+	1,238	32	456	664	1,083	3,612	1.43	1,016	12,714	19,083	33,172	6,068,764	42.37	684	12,424	20,906	35,262	856,498	3.21
Pair Race/Ethnicity																			
Hispanic or Latino	4,176	9	493	815	1,390	8,444	1.61	23	1,375	3,155	8,373	6,068,764	102.68	16	1,168	3,013	8,416	1,053,229	11.77
Black or African American	2,760	42	579	796	1,251	5,724	1.45	59	1,341	3,285	8,689	1,594,823	23.31	17	1,240	3,266	8,011	586,052	9.23
White	16,039	38	428	636	1,031	4,310	1.50	54	1,066	2,834	7,640	5,298,149	39.50	28	988	2,654	7,599	1,028,039	9.12
Other	1,791	19	256	585	1,242	5,171	1.82	26	976	2,726	7,732	272,377	5.93	14	774	2,543	7,854	293,645	6.53
White & Black or African American	217	53	523	816	1,211	3,937	1.47	193	1,697	3,581	8,188	57,878	2.48	68	2,166	4,858	11,831	97,680	2.63
White & Hispanic or Latino	921	18	430	701	1,231	6,940	1.63	20	1,148	3,423	9,307	324,611	5.04	8	1,009	3,223	8,847	277,585	4.94
White & Other	737	20	255	535	981	5,026	1.72	91	1,014	2,737	8,205	306,037	6.09	29	777	2,428	7,257	453,420	8.99
Black or African American & Hispanic or Latino	123	38	696	1,166	1,904	7,579	1.81	39	1,931	5,048	10,354	789,953	18.15	27	1,805	4,597	9,866	285,003	7.52
Black or African American & Other	126	36	380	563	994	3,202	1.67	65	1,112	2,716	6,804	371,337	17.37	15	726	2,375	5,646	334,168	16.81
Hispanic or Latino & Other	145	20	263	559	1,137	5,421	1.91	85	942	2,697	7,865	72,609	3.69	120	1,026	3,407	8,538	88,347	3.58

Table L.1 2012 NSDUH Selected Pair-Level Weight Summary Statistics (continued)

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)						Before sel.pr.ps[1] (SDUWT*PRWT10)						After sel.pr.ps[1] (SDUWT*PRWT10*PRWT11)					
		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	5,928	18	446	705	1,137	7,154	1.56	20	1,020	2,442	6,430	5,298,149	99.54	8	995	2,517	6,644	856,498	8.36
Female, Female	5,813	9	426	651	1,126	8,359	1.61	23	1,032	2,588	6,877	348,424	5.05	15	924	2,455	6,648	527,688	6.63
Male, Female	15,294	18	447	676	1,136	8,444	1.58	31	1,242	3,295	9,094	6,068,764	40.82	14	1,078	3,126	9,128	1,053,229	9.71
Household Size																			
Two	6,855	20	428	663	1,087	6,940	1.54	38	1,320	4,305	11,582	105,073	2.48	15	862	2,867	10,535	192,611	3.36
Three	7,398	18	447	677	1,117	8,444	1.58	20	1,270	3,134	6,962	5,298,149	79.82	8	1,109	3,072	7,076	1,028,039	10.54
Four or More	12,782	9	442	689	1,176	8,359	1.60	23	1,014	2,371	7,069	6,068,764	52.53	14	1,069	2,633	7,458	1,053,229	11.36
Census Region																			
Northeast	5,685	18	335	605	843	7,128	1.61	20	998	2,548	7,060	451,843	6.11	8	784	2,357	6,775	586,033	8.38
South	7,934	9	659	958	1,348	7,138	1.40	23	1,545	4,149	10,607	6,068,764	79.32	15	1,332	3,723	10,044	1,053,229	8.39
Midwest	7,515	33	451	548	728	4,474	1.36	64	896	2,504	6,270	1,594,823	23.29	29	960	2,354	5,998	1,028,039	11.16
West	5,901	20	281	749	1,548	8,444	1.74	68	1,124	2,913	9,026	1,246,856	11.04	35	1,015	2,955	9,102	868,188	8.75
Quarter																			
Quarter1	6,334	21	496	733	1,173	7,847	1.54	26	1,205	2,948	7,968	6,068,764	90.25	14	1,182	3,093	8,706	674,933	6.79
Quarter2	7,183	18	409	619	1,092	7,853	1.58	31	1,069	2,743	7,529	789,953	6.39	16	966	2,669	7,602	456,537	5.56
Quarter3	7,012	9	434	631	1,141	7,138	1.59	20	1,077	2,911	7,852	1,246,856	12.56	8	895	2,545	7,289	868,188	12.34
Quarter4	6,506	16	476	711	1,145	8,444	1.60	23	1,179	3,230	8,450	5,298,149	63.92	15	1,071	2,967	8,090	1,053,229	12.19
% Hispanic or Latino in Segment																			
50-100%	2,105	55	583	1,046	1,488	6,365	1.36	85	1,810	4,352	10,864	6,068,764	114.14	42	1,465	3,748	10,433	674,933	7.42
10-<50%	5,112	9	570	895	1,480	8,444	1.52	23	1,571	3,777	10,099	5,298,149	57.95	16	1,455	3,901	10,200	1,053,229	9.93
<10%	19,818	16	389	613	987	7,853	1.58	20	1,008	2,647	7,213	1,594,823	11.23	8	914	2,467	7,090	1,028,039	8.91
% Black or African American in Segment																			
50-100%	1,905	9	556	763	1,172	7,128	1.59	23	1,312	3,254	8,310	371,337	6.18	15	1,180	3,228	8,308	527,688	7.76
10-<50%	4,030	22	541	832	1,337	7,847	1.50	31	1,436	3,529	9,224	6,068,764	95.73	27	1,237	3,357	9,426	800,116	8.03
<10%	21,100	18	408	639	1,089	8,444	1.58	20	1,071	2,821	7,657	5,298,149	33.09	8	979	2,674	7,618	1,053,229	9.72
% Owner-Occupied DUs[1] in Segment																			
50-100%	20,778	18	441	663	1,116	8,444	1.57	20	1,155	3,052	8,163	6,068,764	52.77	8	1,117	3,059	8,514	1,053,229	9.35
10-<50%	4,889	16	429	715	1,145	7,154	1.57	23	1,076	2,680	7,260	1,594,823	15.44	15	914	2,535	7,070	586,033	7.27
<10%	1,368	9	475	826	1,415	8,351	1.63	29	1,079	2,343	6,980	179,640	3.88	15	409	1,123	2,923	66,093	4.36
Combined Median Rent/Housing Value																			
1st Quintile	4,264	19	392	609	966	8,351	1.62	26	991	2,568	6,844	6,068,764	131.76	15	864	2,304	6,422	1,028,039	15.85
2nd Quintile	5,758	18	411	641	1,108	8,444	1.68	20	1,058	2,749	7,573	573,547	6.64	8	959	2,665	7,281	586,052	8.64
3rd Quintile	6,136	9	389	617	1,036	6,940	1.60	29	1,050	2,804	7,334	1,246,856	13.44	15	868	2,487	6,999	1,053,229	10.97
4th Quintile	6,013	16	478	750	1,210	7,853	1.54	23	1,266	3,248	8,839	5,298,149	65.38	14	1,161	3,121	8,853	868,188	8.61
5th Quintile	4,864	25	530	809	1,264	8,359	1.44	31	1,361	3,320	9,477	1,594,823	11.59	15	1,314	3,596	10,130	701,763	5.48

Table L.1 2012 NSDUH Selected Pair-Level Weight Summary Statistics (continued)

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)						Before sel.pr.ps[1] (SDUWT*PRWT10)						After sel.pr.ps[1] (SDUWT*PRWT10*PRWT11)					
		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]	11,649	9	586	869	1,395	8,359	1.42	23	1,593	3,974	10,354	5,298,149	33.22	15	1,538	3,988	10,441	1,053,229	8.13
Medium to Small MSA[1]	13,392	18	319	561	914	8,444	1.66	20	920	2,352	6,521	573,547	6.09	8	783	2,176	6,263	736,748	8.04
Non-MSA,[1] Urban	511	20	223	492	798	3,628	1.71	79	787	2,016	5,989	6,068,764	256.70	31	607	1,591	4,576	674,933	30.68
Non-MSA,[1] Rural	1,483	20	190	472	829	4,220	1.77	64	720	1,878	5,562	1,317,081	35.03	35	597	1,693	5,171	1,028,039	23.96
Group Quarters																			
Group	386	25	235	422	934	3,853	1.99	38	455	983	1,865	21,866	3.52	15	349	977	2,295	23,423	3.36
Non-Group	26,649	9	446	680	1,136	8,444	1.58	20	1,155	2,984	8,033	6,068,764	46.38	8	1,042	2,843	7,985	1,053,229	9.20

[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 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.nr)

Domain	*n*	Before res.pr.nr[1] (SDUWT*PRWT10*PRWT11)						After res.pr.nr[1] (SDUWT*PRWT10*...*PRWT12)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	19,459	8	965	2,556	7,018	1,053,229	9.79	10	1,129	3,171	9,270	1,550,385	10.18
Pair Age Group													
12-17, 12-17	3,666	8	515	1,007	1,973	17,520	2.32	10	590	1,169	2,375	34,664	2.58
12-17, 18-25	2,778	15	716	1,477	2,953	21,346	2.17	15	854	1,813	3,689	37,794	2.39
12-17, 26-34	653	143	1,901	3,700	7,161	69,227	2.69	143	2,053	4,448	8,869	151,613	3.23
12-17, 35-49	2,816	174	2,326	4,648	9,277	117,119	2.59	176	2,867	5,831	12,576	148,919	2.68
12-17, 50+	627	229	4,611	8,883	17,324	156,917	2.42	258	5,329	10,817	22,909	230,684	2.70
18-25, 18-25	3,976	15	587	1,252	2,936	39,303	2.56	17	645	1,528	3,893	44,826	2.84
18-25, 26-34	742	105	1,745	3,332	8,192	91,433	3.02	113	1,973	3,892	10,805	186,312	3.68
18-25, 35-49	1,058	180	2,675	5,658	13,746	105,508	2.64	183	3,407	8,114	21,008	184,727	2.83
18-25, 50+	650	282	5,534	9,839	19,855	148,367	2.33	289	7,718	15,800	34,492	296,605	2.53
26-34, 26-34	597	396	4,040	6,960	12,877	313,383	5.22	402	4,375	7,653	15,191	481,267	6.43
26-34, 35-49	305	412	4,837	8,543	15,131	382,959	4.50	412	7,194	13,271	29,726	776,131	5.09
26-34, 50+	178	989	10,601	21,015	38,170	701,763	4.02	991	12,816	31,414	78,144	919,088	3.14
35-49, 35-49	489	439	5,391	10,342	18,217	1,053,229	13.26	463	7,345	14,279	29,236	1,550,385	10.91
35-49, 50+	266	1,037	9,310	17,241	34,884	800,116	4.44	1,038	12,826	28,815	75,129	1,024,681	3.17
50+, 50+	658	753	13,107	20,833	34,426	674,933	3.29	1,612	20,499	36,877	69,712	876,058	2.68
Pair Race/Ethnicity													
Hispanic or Latino	3,084	22	1,110	2,795	7,446	1,053,229	14.92	25	1,237	3,235	9,954	1,323,156	13.93
Black or African American	2,210	21	1,175	3,050	7,271	586,052	8.76	21	1,270	3,630	8,954	876,058	9.62
White	11,284	28	927	2,403	6,635	1,028,039	8.77	38	1,107	3,054	8,977	1,550,385	9.94
Other	1,175	14	641	2,074	6,224	249,217	6.58	14	883	2,804	9,328	337,536	6.94
White & Black or African American	166	68	2,086	4,813	11,751	58,915	2.20	86	2,298	6,930	16,409	77,753	2.55
White & Hispanic or Latino	692	8	1,010	3,072	8,420	277,585	5.25	10	1,047	3,268	10,111	481,267	7.61
White & Other	552	29	760	2,335	6,975	118,112	4.53	29	816	2,564	8,087	355,652	8.32
Black or African American & Hispanic or Latino	91	27	1,805	4,597	9,157	285,003	8.23	27	1,965	5,172	12,415	285,060	6.03
Black or African American & Other	99	15	722	2,323	4,297	28,574	2.69	17	1,070	4,113	11,659	106,101	3.28
Hispanic or Latino & Other	106	120	1,132	3,393	6,819	37,041	2.69	126	1,363	4,407	10,976	104,183	3.84
Pair Gender													
Male, Male	4,173	8	941	2,356	5,891	342,223	5.71	10	1,130	2,898	8,040	488,930	7.38
Female, Female	4,425	15	882	2,362	6,170	249,217	4.74	17	1,024	2,850	7,945	285,780	6.13
Male, Female	10,861	14	1,012	2,769	7,844	1,053,229	11.30	14	1,170	3,457	10,446	1,550,385	11.15
Household Size													
Two	4,660	15	789	2,337	8,759	192,611	3.54	17	856	2,654	11,109	292,988	5.19
Three	5,248	8	992	2,765	6,318	1,028,039	11.78	10	1,189	3,534	8,651	1,550,385	12.87
Four or More	9,551	14	1,046	2,514	6,792	1,053,229	11.40	14	1,264	3,187	8,995	1,323,156	11.36

Table L.2 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.nr) (continued)

Domain	*n*	Before res.pr.nr[1] (SDUWT*PRWT10*PRWT11)						After res.pr.nr[1] (SDUWT*PRWT10*...*PRWT12)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Census Region													
Northeast	3,856	8	691	2,056	5,591	586,033	8.98	10	813	2,697	8,171	1,124,135	10.75
South	5,867	15	1,238	3,355	8,957	1,053,229	8.65	17	1,403	4,051	11,527	1,323,156	8.68
Midwest	5,455	29	922	2,184	5,400	1,028,039	13.98	30	1,083	2,645	6,808	1,550,385	14.22
West	4,281	35	986	2,738	8,023	800,116	8.29	36	1,161	3,392	10,802	1,024,681	8.68
Quarter													
Quarter1	4,613	14	1,138	2,868	7,787	674,933	6.90	14	1,351	3,582	10,437	854,597	7.22
Quarter2	5,115	22	912	2,501	6,737	285,003	4.94	25	1,071	3,131	8,922	417,070	6.38
Quarter3	5,093	8	830	2,311	6,395	800,116	12.88	10	963	2,769	8,219	1,024,681	12.79
Quarter4	4,638	15	1,000	2,635	7,081	1,053,229	14.19	15	1,189	3,328	9,797	1,550,385	14.07
% Hispanic or Latino in Segment													
50-100%	1,537	42	1,399	3,267	8,917	674,933	8.97	42	1,594	4,086	11,989	854,597	9.02
10-<50%	3,726	34	1,311	3,552	8,987	1,053,229	11.03	35	1,482	4,402	11,789	1,323,156	10.37
<10%	14,196	8	860	2,272	6,236	1,028,039	8.79	10	1,024	2,844	8,288	1,550,385	9.90
% Black or African American in Segment													
50-100%	1,482	15	1,076	2,881	7,396	342,223	5.53	17	1,232	3,602	10,074	488,930	6.27
10-<50%	2,974	29	1,178	3,074	8,215	800,116	9.87	29	1,287	3,799	10,451	1,024,681	9.57
<10%	15,003	8	923	2,428	6,752	1,053,229	10.14	10	1,089	3,042	8,915	1,550,385	10.70
% Owner-Occupied DUs[1] in Segment													
50-100%	14,866	8	1,055	2,758	7,528	1,053,229	9.93	10	1,248	3,505	10,121	1,550,385	9.97
10-<50%	3,600	15	855	2,336	6,170	586,033	7.50	17	991	2,731	8,123	1,124,135	9.53
<10%	993	15	419	1,065	2,810	66,093	4.60	17	458	1,220	3,258	133,448	6.99
Combined Median Rent/Housing Value													
1st Quintile	3,227	15	853	2,199	5,701	1,028,039	20.71	15	955	2,526	6,979	1,550,385	21.52
2nd Quintile	4,250	8	911	2,453	6,481	586,052	9.89	10	1,048	2,919	8,231	1,124,135	12.13
3rd Quintile	4,364	15	822	2,323	6,197	1,053,229	12.07	17	994	2,839	8,440	1,323,156	11.06
4th Quintile	4,276	14	1,077	2,772	7,763	313,383	4.68	14	1,285	3,558	10,977	355,652	6.06
5th Quintile	3,342	15	1,216	3,258	8,732	701,763	6.12	17	1,488	4,266	12,366	919,088	6.64
Population Density													
Large MSA[1]	8,192	15	1,418	3,593	8,988	1,053,229	8.36	17	1,717	4,648	12,779	1,323,156	8.06
Medium to Small MSA[1]	9,833	8	749	2,019	5,606	586,052	7.84	10	874	2,423	7,178	876,058	9.61
Non-MSA,[1] Urban	380	48	623	1,562	4,150	674,933	42.44	52	742	1,980	5,241	854,597	35.92
Non-MSA,[1] Rural	1,054	35	593	1,682	4,927	1,028,039	34.20	38	717	1,978	6,264	1,550,385	35.14
Group Quarters													
Group	312	15	349	872	1,924	23,423	3.63	17	429	1,061	2,898	34,734	4.01
Non-Group	19,147	8	983	2,606	7,093	1,053,229	9.72	10	1,150	3,234	9,414	1,550,385	10.09

[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 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.ps and res.pr.ev)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PRWT10*...*PRWT12)						After res.pr.ps[1] (SDUWT*PRWT10*...*PRWT13)						Final Weight: After res.pr.ev[1] (SDUWT*PRWT10*...*PRWT14)					
		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	19,459	10	1,129	3,171	9,270	1,550,385	10.18	4	1,082	3,123	9,214	1,234,015	9.58	3	1,068	3,112	9,208	1,237,337	9.36
Pair Age Group																			
12-17, 12-17	3,668	10	593	1,177	2,391	34,664	2.58	4	542	1,168	2,393	30,611	2.50	3	534	1,162	2,395	32,280	2.52
12-17, 18-25	2,759	15	847	1,791	3,683	37,794	2.42	6	815	1,821	3,816	25,541	2.31	6	813	1,829	3,856	24,364	2.31
12-17, 26-34	658	143	2,030	4,412	8,943	151,613	3.25	89	1,938	4,295	9,294	102,559	3.01	87	1,884	4,201	9,418	112,015	3.09
12-17, 35-49	2,812	176	2,886	5,847	12,501	148,919	2.69	130	2,567	5,540	12,475	142,220	2.83	105	2,514	5,526	12,426	143,455	2.82
12-17, 50+	631	258	5,228	10,765	22,907	230,684	2.72	184	4,418	10,190	23,400	196,486	2.77	179	4,462	10,151	23,711	192,088	2.71
18-25, 18-25	3,901	17	644	1,524	3,865	44,826	2.83	15	580	1,562	4,185	44,635	2.77	14	568	1,556	4,193	32,458	2.71
18-25, 26-34	794	52	1,764	3,722	10,598	323,710	4.78	40	1,516	3,814	9,782	153,541	3.94	36	1,448	3,764	9,723	151,757	3.98
18-25, 35-49	1,053	183	3,389	8,145	21,593	184,727	2.82	172	3,071	7,900	21,308	177,102	2.80	158	3,082	7,802	21,369	154,716	2.75
18-25, 50+	660	289	7,649	15,463	33,878	296,605	2.59	432	6,714	15,392	33,732	299,472	2.64	402	6,780	15,470	34,493	279,736	2.57
26-34, 26-34	604	207	3,971	7,400	14,503	481,267	6.58	232	3,844	8,467	16,566	727,153	7.05	216	3,645	8,417	16,631	624,327	6.46
26-34, 35-49	320	412	6,938	12,531	27,491	776,131	5.14	137	5,215	12,267	24,629	1,004,156	7.79	130	5,035	11,784	24,653	1,011,952	7.98
26-34, 50+	177	991	12,555	30,990	77,130	919,088	3.28	812	12,545	30,285	91,722	633,208	2.64	756	12,324	30,567	91,602	549,404	2.55
35-49, 35-49	487	463	7,217	14,279	29,236	1,550,385	10.96	164	6,271	15,001	32,702	1,234,015	8.51	155	6,301	15,391	32,738	1,237,337	8.28
35-49, 50+	272	1,038	12,849	27,776	71,681	1,024,681	3.21	671	10,431	27,472	74,724	760,439	2.95	620	10,194	26,379	74,748	612,413	2.84
50+, 50+	663	1,612	20,423	36,884	70,196	876,058	2.67	1,647	18,150	35,352	69,493	818,783	2.72	1,632	18,609	35,696	71,197	815,689	2.68
Pair Race/Ethnicity																			
Hispanic or Latino	3,134	25	1,251	3,295	10,003	1,323,156	13.84	15	1,218	3,280	9,992	1,234,015	13.21	15	1,201	3,234	9,988	1,237,337	12.94
Black or African American	2,139	21	1,286	3,663	9,282	876,058	9.71	15	1,253	3,637	9,153	818,783	9.44	15	1,247	3,647	9,107	815,689	9.47
White	11,006	29	1,102	3,054	8,995	1,550,385	10.05	23	1,093	3,116	9,136	1,010,083	8.72	21	1,078	3,108	9,117	1,011,952	8.50
Other	1,116	14	958	3,017	10,103	337,536	6.66	4	914	3,081	10,892	374,107	7.06	3	926	3,165	10,780	346,839	6.88
White & Black or African American	167	86	1,809	6,079	14,148	77,753	2.61	124	1,826	7,641	16,843	100,462	2.57	123	1,810	7,636	14,991	99,568	2.59
White & Hispanic or Latino	721	10	1,065	2,958	9,380	481,267	8.22	6	856	2,589	9,030	727,153	11.44	6	822	2,527	9,100	624,327	10.20
White & Other	715	16	881	2,689	7,897	240,232	6.02	8	608	1,904	6,289	347,902	8.03	8	568	1,854	6,261	370,874	8.47
Black or African American & Hispanic or Latino	109	98	1,474	4,559	9,426	140,505	4.00	166	1,473	4,575	10,998	165,865	4.66	145	1,364	4,448	10,885	180,929	4.82
Black or African American & Other	198	17	1,033	3,397	7,930	181,458	6.32	5	392	1,408	3,562	91,049	5.96	4	349	1,388	3,582	90,913	5.98
Hispanic or Latino & Other	154	55	1,030	3,551	9,841	206,329	6.36	32	539	2,016	6,085	109,860	6.18	30	521	1,973	6,135	109,119	6.22

Table L.3 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.ps and res.pr.ev) (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PRWT10*...*PRWT12)						After res.pr.ps[1] (SDUWT*PRWT10*...*PRWT13)						Final Weight: After res.pr.ev[1] (SDUWT*PRWT10*...*PRWT14)					
		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	4,162	10	1,127	2,892	8,027	488,930	7.41	6	1,078	2,843	8,046	501,086	7.64	6	1,057	2,848	8,133	502,540	7.60
Female, Female	4,430	17	1,023	2,852	7,945	285,780	6.13	13	965	2,752	7,804	374,107	6.57	12	961	2,741	7,772	346,839	6.53
Male, Female	10,867	14	1,170	3,460	10,465	1,550,385	11.13	4	1,141	3,386	10,381	1,234,015	10.20	3	1,121	3,357	10,468	1,237,337	9.92
Household Size																			
Two	4,660	17	856	2,654	11,109	292,988	5.19	16	774	2,609	11,083	280,634	5.16	14	743	2,532	11,015	288,931	5.26
Three	5,248	10	1,189	3,534	8,651	1,550,385	12.87	5	1,150	3,349	8,560	1,010,083	10.50	4	1,158	3,356	8,647	1,002,962	10.00
Four or More	9,551	14	1,264	3,187	8,995	1,323,156	11.36	4	1,226	3,180	8,946	1,234,015	11.35	3	1,224	3,186	8,955	1,237,337	11.11
Census Region																			
Northeast	3,856	10	813	2,697	8,171	1,124,135	10.75	4	754	2,559	8,056	1,010,083	11.36	3	746	2,522	8,063	1,011,952	11.29
South	5,867	17	1,403	4,051	11,527	1,323,156	8.68	5	1,362	3,928	11,269	1,234,015	8.81	4	1,364	3,931	11,279	1,237,337	8.81
Midwest	5,455	30	1,083	2,645	6,808	1,550,385	14.22	18	1,043	2,601	6,923	804,009	9.01	16	1,023	2,563	6,817	639,872	8.34
West	4,281	36	1,161	3,392	10,802	1,024,681	8.68	22	1,132	3,385	10,721	760,439	8.69	20	1,105	3,351	10,734	624,327	8.33
Quarter																			
Quarter1	4,613	14	1,351	3,582	10,437	854,597	7.22	4	1,274	3,548	10,107	805,478	7.47	3	1,247	3,494	10,168	806,173	7.45
Quarter2	5,115	25	1,071	3,131	8,922	417,070	6.38	9	1,021	3,015	8,778	506,343	6.85	8	1,001	3,000	8,776	515,372	6.86
Quarter3	5,093	10	963	2,769	8,219	1,024,681	12.79	6	917	2,717	8,257	884,682	12.27	6	910	2,724	8,293	890,858	11.63
Quarter4	4,638	15	1,189	3,328	9,797	1,550,385	14.07	6	1,170	3,336	9,836	1,234,015	11.60	6	1,151	3,312	9,789	1,237,337	11.37
% Hispanic or Latino in Segment																			
50-100%	1,537	42	1,594	4,086	11,989	854,597	9.02	34	1,508	4,019	12,166	884,682	9.65	33	1,482	3,980	12,144	890,858	9.73
10-<50%	3,726	35	1,482	4,402	11,789	1,323,156	10.37	9	1,434	4,263	11,836	1,234,015	9.54	8	1,399	4,222	11,854	1,237,337	9.27
<10%	14,196	10	1,024	2,844	8,288	1,550,385	9.90	4	968	2,790	8,241	1,010,083	9.16	3	960	2,779	8,206	1,011,952	8.90
% Black or African American in Segment																			
50-100%	1,482	17	1,232	3,602	10,074	488,930	6.27	5	1,175	3,589	10,217	501,086	6.35	4	1,165	3,573	10,398	502,540	6.32
10-<50%	2,974	29	1,287	3,799	10,451	1,024,681	9.57	18	1,257	3,738	10,228	805,478	8.61	16	1,267	3,750	10,205	806,173	8.21
<10%	15,003	10	1,089	3,042	8,915	1,550,385	10.70	4	1,034	2,978	8,935	1,234,015	10.13	3	1,018	2,949	8,924	1,237,337	9.94
% Owner-Occupied DUs[1] in Segment																			
50-100%	14,866	10	1,248	3,505	10,121	1,550,385	9.97	4	1,216	3,410	9,934	1,234,015	9.37	3	1,209	3,401	9,941	1,237,337	9.12
10-<50%	3,600	17	991	2,731	8,123	1,124,135	9.53	9	929	2,762	8,444	1,010,083	9.02	8	909	2,733	8,434	1,002,962	9.05
<10%	993	17	458	1,220	3,258	133,448	6.99	5	364	1,097	3,271	123,637	7.55	4	357	1,098	3,192	117,144	7.34
Combined Median Rent/Housing Value																			
1st Quintile	3,227	15	955	2,526	6,979	1,550,385	21.52	6	943	2,599	7,106	1,004,156	15.44	6	933	2,629	7,213	1,011,952	14.03
2nd Quintile	4,250	10	1,048	2,919	8,231	1,124,135	12.13	6	1,020	2,887	8,310	1,010,083	12.22	6	1,008	2,860	8,337	1,002,962	12.24
3rd Quintile	4,364	17	994	2,839	8,440	1,323,156	11.06	5	944	2,762	8,242	1,234,015	11.13	4	921	2,724	8,185	1,237,337	11.29
4th Quintile	4,276	14	1,285	3,558	10,977	355,652	6.06	4	1,220	3,461	10,809	399,094	6.23	3	1,209	3,424	10,736	413,094	6.20
5th Quintile	3,342	17	1,488	4,266	12,366	919,088	6.64	18	1,409	4,103	12,186	727,153	6.73	16	1,419	4,111	12,119	624,327	6.41

Table L.3 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.ps and res.pr.ev) (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PRWT10*...*PRWT12)						After res.pr.ps[1] (SDUWT*PRWT10*...*PRWT13)						Final Weight: After res.pr.ev[1] (SDUWT*PRWT10*...*PRWT14)					
		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]	8,192	17	1,717	4,648	12,779	1,323,156	8.06	5	1,708	4,620	12,943	1,234,015	7.56	4	1,692	4,589	12,922	1,237,337	7.40
Medium to Small MSA[1]	9,833	10	874	2,423	7,178	876,058	9.61	4	807	2,307	6,911	1,004,156	11.13	3	788	2,290	6,933	1,011,952	11.08
Non-MSA,[1] Urban	380	52	742	1,980	5,241	854,597	35.92	16	762	1,878	5,022	805,478	32.27	16	737	1,852	5,022	806,173	32.41
Non-MSA,[1] Rural	1,054	38	717	1,978	6,264	1,550,385	35.14	30	852	2,366	6,703	804,009	13.38	29	859	2,382	6,979	639,872	10.49
Group Quarters																			
Group	312	17	429	1,061	2,898	34,734	4.01	25	408	910	2,667	35,420	3.79	25	418	877	2,730	28,121	3.70
Non-Group	19,147	10	1,150	3,234	9,414	1,550,385	10.09	4	1,109	3,185	9,334	1,234,015	9.49	3	1,093	3,175	9,352	1,237,337	9.28
Pair Relationship Domain[4]																			
Parent-Child (12-14)	1,981	83	2,456	5,444	11,989	163,632	2.95	80	2,413	5,537	12,561	196,486	3.13	75	2,355	5,499	12,516	192,088	3.14
Parent-Child (12-17)	3,748	15	2,755	5,933	12,968	230,684	2.95	6	2,630	5,869	13,546	196,486	3.04	6	2,547	5,798	13,561	192,088	3.04
Parent-Child (12-20)	4,443	15	2,915	6,347	14,088	296,605	3.18	6	2,825	6,450	14,879	299,472	3.27	6	2,793	6,413	14,820	279,736	3.21
Sibling (12-14)-Sibling (15-17)	2,191	14	588	1,175	2,404	34,664	2.58	4	562	1,193	2,423	30,611	2.44	3	555	1,184	2,449	32,280	2.47
Sibling (12-17)-Sibling (18-25)	2,477	16	844	1,789	3,667	37,794	2.36	8	828	1,843	3,811	23,928	2.28	8	823	1,839	3,849	24,220	2.28
Spouse-Spouse/ Partner-Partner	3,664	30	1,097	4,481	16,693	1,550,385	10.77	16	1,108	4,420	17,392	1,234,015	9.77	14	1,065	4,375	17,434	1,237,337	9.62
Spouse-Spouse/ Partner-Partner with Children (Younger Than 18)	1,640	51	1,055	3,442	10,817	1,323,156	16.13	70	1,595	5,143	15,628	1,234,015	12.24	69	1,575	5,108	15,656	1,237,337	12.17

[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: Pair Analysis Manual Excerpt

Appendix M: Pair Analysis Manual Excerpt

This appendix provides background information on pair data analysis for the National Survey on Drug Use and Health (NSDUH). This excerpt is from Section 3.2 in *How To Prepare and Analyze Pair Data in the National Survey on Drug Use and Health* by Frechtel, Warren, and Porter (in press).

Inferential Population and Multiplicities

There are different perspectives through which pair data can be analyzed: (1) with pairs as the focus, or (2) with one member of the pair as the focus. When the focus is on the pair, the *PRANALWT* variable can be used to weight the data directly with no adjustments. However, when the focus is on one member of the pair, an adjustment often needs to be made to the weight to account for this. For example, the analysis outlined in Table 1 is focused on the child's behavior when the father talked to him or her about substance use. However, if the analyst was interested in the child's behavior when a parent talked to him or her about substance use, regardless of the gender of the parent, the weight would need to be adjusted to account for the fact that the child may be a member of more than one parent-child pair. For the analysis in Table 1, there is no multiple counting problem as long as an assumption is made that no child lives with more than one father.

Adjustments for the multiple counting problem are done using "multiplicities" (Chromy & Singh, 2001). These multiplicities have been computed for the analyst and can be accounted for by simply adjusting the weight variable (described in Section 4.4). Analysts should exercise care or seek assistance in computing multiplicities for any pair types not listed. The process by which the multiplicities are created is described in detail in Chapter 10 of Frechtel et al. (2013). Table 5 lists the pair domains for which multiplicities can be used to perform person-level analyses.

Table 5. Pair Domains and Multiplicities

Pair Domain			Multiplicity Variable	
IRPRREL **Levels**	**Description**	**Focus**[1]	**Name**	**Description**
1	Parent-child, child aged 12-14	Parent	*IRMPCP14*	Number of children aged 12-14 living with responding parent
		Child	*IRMPCC14*	Number of parents living with responding child aged 12-14
1,2	Parent-child, child aged 12-17	Parent	*IRMPCP17*	Number of children aged 12-17 living with responding parent
		Child	*IRMPCC17*	Number of parents living with responding child aged 12-17
1,2,3	Parent-child, child aged 12-20	Parent	*IRMPCP20*	Number of children aged 12-20 living with responding parent
		Child	*IRMPCC20*	Number of parents living with responding child aged 12-20
2	Parent-child, child aged 15-17	Parent	*IRMPCP57*	Number of children aged 15-17 living with responding parent
		Child	*IRMPCC57*	Number of parents living with responding child aged 15-17
5	Sibling-sibling, older sibling aged 15-17, younger sibling aged 12-14	Older sibling	*IRMS1417*	Number of siblings aged 12-14 living with responding sibling aged 15-17
		Younger sibling	*IRMS1714*	Number of siblings aged 15-17 living with responding sibling aged 12-14
6	Sibling-sibling, older sibling aged 18-25, younger sibling aged 12-17	Older sibling	*IRMS1725*	Number of siblings aged 12-17 living with responding sibling aged 18-25
		Younger sibling	*IRMS2517*	Number of siblings aged 18-25 living with responding sibling aged 12-17
8,9	Spouse-spouse and partner-partner	No multiplicity necessary: assume only one spouse per person		
8	Spouse-spouse and partner-partner, with children aged 0-17	No multiplicity necessary: assume only one spouse per person		

[1] No weight adjustment is needed when the inferential focus is on the pair.

To help clarify this concept, some bulleted examples are included below. A full list of multiplicities and definitions is included in Table 5 above. The examples are based on the analysis in Table 6, which is the same as the analysis in Table 1 except that the parent can be a father or a mother.

Table 6. Example of Pair-Level Analysis Requiring Multiplicities

Parent Reports Talking to Child Aged 12-17 about Substance Use	Child Used an Illicit Drug in the Past Year	
	Yes	No
Yes	(1,1)	(1,2)
No	(2,1)	(2,2)

- To populate Table 6 with estimates related to children (i.e., row percentages), the proper weight is *PRANALWT*/*IRMPCC17*. For example, this weight could be used to estimate the probability that a child aged 12 to 17 used an illicit drug in the past year, given that a parent talked to him or her about substance use in the past year (cell (1,1) in the table).
- If Table 6 instead showed estimates related to parents (i.e., column percentages), the proper weight would be *PRANALWT*/*IRMPCP17*. For example, this weight could be used to estimate the probability that a parent talked to his or her child aged 12 to 17 about substance use in the past year, given that his or her child used an illicit drug in the past year (again, cell (1,1) in the table).
- As stated above, if estimates related to children are desired, but interest is restricted to either fathers or mothers, standard practice is to assume the presence of no more than one father and no more than one mother in the dwelling unit. In these cases, no multiplicity is necessary, and *PRANALWT* can be used as the analysis weight. For example, *PRANALWT* should be used to estimate the probability that a child aged 12 to 17 used an illicit drug in the past year, given that his or her mother talked to him or her about substance use in the past year.
- If Table 6 showed estimates related to parent-child pairs and did not focus on either member of the pair (i.e., cell percentages), then no multiplicity would be necessary. The variable *PRANALWT* is designed for this. An analysis like this is of questionable value because of the multiple counting problem, however. The inferential population includes all parent-child pairs where the child is aged 12 to 17. Many persons are represented in more than one pair. A dwelling unit with two parents and three children aged 12 to 17 would represent six pairs in the population, with each parent being a member of three of the six and each child being a member of two of the six. A dwelling unit with one parent and one child aged 12 to 17 would represent only one pair in the population. Analyses that focus on only one member of the pair usually have a more natural interpretation.