2006 NATIONAL SURVEY ON DRUG USE AND HEALTH

SAMPLE DESIGN REPORT

Prepared for the 2006 Methodological Resource Book

RTI Project No. 0209009.230.004
Contract No. 283-2004-00022
Phase II, Deliverable 8

Authors:

Katherine B. Morton
James R. Chromy
Susan R. Hunter
Peilan C. Martin

Project Director: Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

January 31, 2007

2006 NATIONAL SURVEY ON DRUG USE AND HEALTH

SAMPLE DESIGN REPORT

Prepared for the 2006 Methodological Resource Book

Contract No. 283-2004-00022
RTI Project No. 0209009.230.004

Deliverable No. 8

Authors:

Katherine B. Morton
James R. Chromy
Susan R. Hunter
Peilan C. Martin

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

January 31, 2007

Acknowledgments

This publication was developed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Office of Applied Studies (OAS), by RTI International (a trade name of Research Triangle Institute), Research Triangle Park, North Carolina, under Contract No. 283-2004-00022. Significant contributors at RTI include Katherine B. Morton, James R. Chromy, Susan R. Hunter, Peilan C. Martin, Patrick Chen, David Chrest, Richard Straw, Joyce Clay-Brooks, and Thomas G. Virag (Project Director).

Table of Contents

Chapter **Page**

List of Tables v
List of Exhibits v

1. Overview 1
   - 1.1 Target Population 1
   - 1.2 Design Overview 1
   - 1.3 5-Year Design 2
   - 1.4 Stratification and First- and Second-Stage Sample Selections 2
   - 1.5 Dwelling Units and Persons 4

2. The Coordinated 5-Year Sample 7
   - 2.1 Formation of and Objectives for Using the Composite Size Measures 7
   - 2.2 Stratification 8
   - 2.3 First- and Second-Stage Sample Selection 9
   - 2.4 Survey Year and Quarter Assignment 11
   - 2.5 Creation of Variance Estimation Strata 11

3. General Sample Allocation Procedures for the Main Study 15
   - 3.1 Notation 15
   - 3.2 Determining Person Sample Sizes by State and Age Group 18
   - 3.3 Third-Stage Sample Allocation for Each Segment 20
     - 3.3.1 Dwelling Unit Frame Construction—Counting and Listing 20
     - 3.3.2 Determining Dwelling Unit Sample Size 25
   - 3.4 Determining Fourth-Stage Sample (Person) Selection Probabilities for Each Segment 25
   - 3.5 Sample Size Constraints: Guaranteeing Sufficient Sample for Additional Studies and Reducing Field Interviewer Burden 26
   - 3.6 Dwelling Unit Selection and Release Partitioning 27
   - 3.7 Half-Open Interval Rule and Procedure for Adding Dwelling Units 27
   - 3.8 Quarter-by-Quarter Deviations 27
   - 3.9 Impact of Hurricanes Katrina and Rita on the 2006 NSDUH Sample 33
   - 3.10 Sample Weighting Procedures 34

4. General Sample Allocation Procedures for the Reliability Study and Income Split Sample 41
   - 4.1 Respondent Universe and Sampling Methods for the Reliability Study 41
   - 4.2 Dwelling Unit and Person Sample Allocation Procedures for the Reliability Study 42
   - 4.3 Dwelling Unit Allocation Procedures for the Income Split Sample 43

Table of Contents (continued)

Chapter **Page**

References 45

Appendix

A 2005-2009 NSDUH State Sampling Regions 47

B 2006 NSDUH Procedure for Subsegmenting 101

C 2006 NSDUH Procedure for Adding Missed Dwelling Units 103

List of Tables

Table 2.1 Number of Census Tracts and Segments on Sampling Frame, by State 9

Table 2.2 Segment Identification Number Suffixes and Quarter Assignment 12

Table 3.1 Expected Relative Standard Errors, by Age Group 19

Table 3.2 Expected Main Study Sample Sizes, by State and Age Group 21

Table 3.3 Number of Map Pages, by State and Segment 23

Table 3.4 Segment and Dwelling Unit Summary 24

Table 3.5 Quarterly Sample Sizes and Percentages Released 28

Table 3.6 Definitions of Levels for Potential Variables for Dwelling Unit Nonresponse Adjustment 35

Table 3.7 Definitions of Levels for Potential Variables for Dwelling Unit Poststratification and Respondent Poststratification at the Person Level 36

Table 3.8 Definitions of Levels for Potential Variables for Selected Person Poststratification and Person-Level Nonresponse Adjustment 37

Table 3.9 Model Group Definitions 39

Table 4.1 Reliability Study Design Parameters 42

List of Exhibits

Exhibit 1. Flowchart of Sample Weighting Steps 39

Exhibit 2. Sample Weight Components 40

Chapter 1. Overview

1.1 Target Population

The respondent universe for the 2006 National Survey on Drug Use and Health[1] (NSDUH) was the civilian, noninstitutionalized population aged 12 years or older residing within the United States and the District of Columbia. Consistent with the NSDUH designs since 1991, the 2006 NSDUH universe included residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories, and group homes), residents of Alaska and Hawaii, and civilians residing on military bases. Coverage before the 1991 survey was limited to residents of the coterminous 48 States, and it excluded residents of group quarters and all persons (including civilians) living on military bases. Persons excluded from the 2006 universe included those with no fixed household address (e.g., homeless transients not in shelters) and residents of institutional group quarters, such as jails and hospitals.

1.2 Design Overview

Beginning in 1999 and continuing through subsequent years, the Substance Abuse and Mental Health Services Administration (SAMHSA) implemented major changes in the way NSDUH would be conducted. The surveys are conducted using computer-assisted interviewing (CAI) methods and provide improved State estimates based on minimum sample sizes per State. The total targeted sample size of 67,500 is equally allocated across three age groups: persons aged 12 to 17, persons aged 18 to 25, and persons aged 26 or older. This large sample size allows SAMHSA to continue reporting precise demographic subgroups at the national level without needing to oversample specially targeted demographics, as required in the past. This large sample is referred to as the "main sample" or the "CAI sample." The achieved sample for the 2006 CAI sample was 67,802 persons.

Beginning with the 2002 NSDUH and continuing through the 2006 NSDUH, survey respondents were given a $30 incentive payment for participation. As expected, the incentive had the effect of increasing response rates, thereby requiring fewer selected households than previous surveys.

An additional design change was made in 2002 and continued through 2006. A new pair-sampling strategy was implemented that increased the number of pairs selected in dwelling units (DUs) with older persons on the roster (Chromy & Penne, 2002). With the increase in the number of pairs came a moderate decrease in the response rate for older persons.

A Reliability Study was conducted during quarters 2 through 4 of the 2006 NSDUH. The Reliability Study sample was embedded within the main study sample. Only households in which one person was selected were eligible for the study. Reliability Study respondents were

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

administered the survey instrument on two occasions. Details on the Reliability Study sample are provided in Chapter 4.

Additionally, a new income module was tested in the 2006 NSDUH in which respondents received a reduced set of income questions. Details on the assignment of sample households to income module treatments are provided in Chapter 4.

Finally, Hurricanes Katrina and Rita hit the Gulf Coast in the fall of 2005. Large-scale devastation was experienced in the States of Louisiana, Mississippi, Alabama, and Texas. The sample was supplemented in the areas determined to be hardest hit by the hurricanes in quarter 1 of the 2006 NSDUH. As a result, a total of 7,207 segments were fielded in the 2006 survey. The sample supplement is discussed in greater detail in Section 3.9.

1.3 5-Year Design

The 2006 NSDUH is part of a coordinated 5-year sample design. Although there is no planned overlap with the 1999-2004 samples, a coordinated design for 2005 through 2009 facilitated 50 percent overlap in second-stage units (area segments) within each successive 2-year period from 2005 through 2009. This design was intended to increase the precision of estimates in year-to-year trend analyses, using the expected positive correlation resulting from the overlapping sample between successive NSDUH years.

The 2006 design provides for estimates by State in all 50 States plus the District of Columbia. States may therefore be viewed as the first level of stratification as well as a reporting variable. Eight States, referred to as the "large" States,[2] had samples designed to yield 3,600 respondents per State for the 2006 survey. This sample size was considered adequate to support direct State estimates. The remaining 43 States[3] had samples designed to yield 900 respondents per State in the 2006 survey. In these 43 States, adequate data were available to support reliable State estimates based on small area estimation (SAE) methodology.

1.4 Stratification and First- and Second-Stage Sample Selections

Within each State, State sampling (SS) regions were formed. Based on a composite size measure, States were geographically partitioned into roughly equal-sized regions according to population. In other words, regions were formed such that each area yielded, in expectation, roughly the same number of interviews during each data collection period. The smaller States were partitioned into 12 SS regions, whereas the eight large States were divided into 48 SS regions. Therefore, the partitioning of the United States resulted in the formation of a total of 900 SS regions. SS region maps can be found in Appendix A.

Unlike the 1999 through 2001 NHSDAs and the 2002 through 2004 NSDUHs, the first stage of selection for the 2005 through 2009 NSDUHs was census tracts. This stage was included

[2] The large States are California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas.

[3] For reporting and stratification purposes, the District of Columbia is treated the same as a State, and no distinction is made in the discussion.

to contain sample segments within a single census tract to the extent possible.[4] In prior years, segments that crossed census tract boundaries made merging to external data sources difficult.

The first stage of selection began with the construction of an area sample frame that contained one record for each census tract in the United States. If necessary, census tracts were aggregated within SS regions until each tract[5] had, at a minimum, 150 DUs[6] in urban areas and 100 DUs in rural areas.[7]

Before selecting census tracts, additional implicit stratification was achieved by sorting the first-stage sampling units by a CBSA/SES[8] (core-based statistical area/socioeconomic status) indicator[9] and by the percentage of the population that is non-Hispanic and white. From this well-ordered sample frame, 48 census tracts per SS region were selected with probabilities proportionate to a composite size measure and with minimum replacement (Chromy, 1979).

Because census tracts generally exceed the minimum DU requirement, one smaller geographic region was selected within each sampled census tract. For this second stage of sampling, each selected census tract was partitioned into compact clusters[10] of DUs by aggregating adjacent census blocks. Consistent with the terminology used in previous NSDUHs, these geographic clusters of blocks are referred to as "segments." A sample DU in NSDUH refers to either a housing unit or a group-quarters listing unit, such as a dormitory room or a shelter bed. Similar to census tracts, segments were formed to contain a minimum of 150 DUs in urban areas and 100 DUs in rural areas. This minimum DU requirement will support the overlapping sample design and any special supplemental samples or field tests that SAMHSA may wish to conduct.

One segment was selected within each sampled census tract with probability proportionate to size. The 48 selected segments then were randomly assigned to a survey year and quarter of data collection as described in Section 2.4.

---

[4] Some census tracts had to be aggregated in order to meet the minimum DU requirement.

[5] For the remainder of the discussion, first-stage sampling units are referred to as "census tracts" even though each first-stage sampling unit contains one or more census tracts.

[6] DU counts were obtained from the 2000 census data supplemented with revised population counts from Claritas.

[7] The basis for the differing minimum DU requirement in urban and rural areas is that it is more difficult to meet the requirement in rural areas, and 100 DUs is sufficient to support one field test and two main study samples.

[8] CBSAs include metropolitan and micropolitan statistical areas as defined by the Office of Management and Budget on June 6, 2003.

[9] Four categories are defined as (1) CBSA/low SES, (2) CBSA/high SES, (3) non-CBSA/low SES, and (4) non-CBSA/high SES. To define SES, block group-level median rents and property values were given a rank (1,…,5) based on State and CBSA quintiles. The rent and value ranks then were averaged, weighted by the percentages renter- and owner-occupied DUs, respectively. If the resulting score fell in the lower 25th percentile by State and CBSA, the area was considered "low SES"; otherwise, it was considered "high SES."

[10] Although the entire cluster is compact, the final sample of DUs represents a noncompact cluster. Noncompact clusters (selection from a list) differ from compact clusters in that not all units within the cluster are included in the sample. Although compact cluster designs are less costly and more stable, a noncompact cluster design was used because it provides for greater heterogeneity of dwellings within the sample. Also, social interaction (contagion) among neighboring dwellings is sometimes introduced with compact clusters (Kish, 1965).

1.5 Dwelling Units and Persons

After sample segments for the 2006 NSDUH were selected, specially trained field household listers visited the areas and obtained complete and accurate lists of all eligible DUs within the sample segment boundaries. These lists served as the frames for the third stage of sample selection.

The primary objective of the third stage of sample selection (listing units) was to determine the minimum number of DUs needed in each segment to meet the targeted sample sizes for all age groups. Thus, listing unit sample sizes for the segment were determined using the age group with the largest sampling rate, which we refer to as the "driving" age group. Using 2000 census data adjusted to more recent data from Claritas, State- and age-specific sampling rates were computed. These rates then were adjusted by the segment's probability of selection; the subsegmentation inflation factor,[11] if any; the probability of selecting a person in the age group (equal to the maximum, or 0.99, for the driving age group); and an adjustment for the "maximum of two" rule.[12] In addition to these factors, historical data from the 2004, 2005, and 2006 NSDUHs were used to compute predicted screening and interviewing response rate adjustments. The final adjusted sampling rate then was multiplied by the actual number of DUs found in the field during counting and listing activities. The product represents the segment's listing unit sample size.

Some constraints were put on the listing unit sample sizes. For example, to ensure adequate samples for supplemental studies, the listing unit sample size could not exceed 100 or half of the actual listing unit count. Similarly, if five unused listing units remained in the segment, a minimum of five listing units per segment was required for cost efficiency.

Using a random start point and interval-based (systematic) selection, the actual listing units were selected from the segment frame. After DU selections were made, an interviewer visited each selected DU to obtain a roster of all persons residing in the DU. As in previous years, during the data collection period, if an interviewer encountered any new DU in a segment or found a DU that was missed during the original counting and listing activities, the new or missed dwellings were selected into the 2006 NSDUH using the half-open interval selection technique.[13] This selection technique eliminates any frame bias that might be introduced because of errors and/or omissions in the counting and listing activities, and it also eliminates any bias that might be associated with using "old" segment listings.

---

[11] Segments found to be very large in the field are partitioned into *subsegments*. Then one subsegment is chosen at random with probability proportional to the size to be fielded. The subsegmentation inflation factor accounts for the narrowing down of the segment.

[12] Brewer's Selection Algorithm never allows for greater than two persons per household to be chosen. Thus, sampling rates are adjusted to satisfy this constraint.

[13] In summary, this technique states that, if a DU is selected for the 2006 study and an interviewer observes any new or missed DUs between the selected DU and the DU appearing immediately after the selection on the counting and listing form, all new or missed dwellings falling in this interval will be selected. If a large number of new or missed DUs are encountered (greater than 10), a sample of the new or missing DUs will be selected. For more information, please refer to Appendix C.

Using the roster information obtained from an eligible member of the selected DU, 0, 1, or 2 persons were selected for the survey. Sampling rates were preset by age group and State. Roster information was entered directly into the electronic screening instrument, which automatically implemented this fourth stage of selection based on the State and age group sampling parameters.

One exciting advantage of using an electronic screening instrument in NSDUH is the ability to impose a more complicated person-level selection algorithm on the fourth stage of the NSDUH design. Similar to the 1999 through 2005 designs, one feature that was included in the 2006 design was that any two survey-eligible persons within a DU had some chance of being selected (i.e., all survey-eligible pairs of persons had some nonzero chance of being selected). This design feature was of interest to NSDUH researchers because, for example, it allows analysts to examine how the drug use propensity of one individual in a family relates to the drug use propensity of another family member residing in the same DU (e.g., the relationship of drug use between a parent and his or her child).

Chapter 2. The Coordinated 5-Year Sample

As was mentioned previously, the sample design was developed simultaneously for each of the 2005 through 2009 National Surveys on Drug Use and Health (NSDUHs). Starting with a census block-level frame, first- and second-stage sampling units (census tracts and area segments, respectively) were formed. A sufficient number of segments then were selected within sampled census tracts to support the 5-year design as well as any supplemental studies the Substance Abuse and Mental Health Services Administration (SAMHSA) chose to field.

2.1 Formation of and Objectives for Using the Composite Size Measures

The composite size measure procedure is used to obtain self-weighting samples for multiple domains in multistage designs. The NSDUH sample design has employed the composite size measure methodology since 1988. Our goal was to specify size measures for sample areas (segments) and dwelling units (DUs) that would achieve the following objectives:

- Yield the targeted domain sample sizes in expectation (E_s) over repeated samples; that is, if m_{ds} is the domain d sample size achieved by sample s, then

$$E_s(m_{ds}) = m_d \text{ for } d = 1,...,D. \quad (1)$$

- Constrain the maximum number of selections per DU at a specified value; specifically, we limited the total number of within-DU selections across all age groups to a maximum of 2.

- Minimize the number of sample DUs that must be screened to achieve the targeted domain sample sizes.

- Eliminate all variation in the sample inclusion probabilities within a domain, except for the variation in the within-DU/within-domain probabilities of selection. The inverse probabilities of selection for each sample segment were used to determine the number of sample lines to select from within each segment. As a consequence, all DUs within a specific stratum were selected with approximately the same probability, and therefore, approximately equalized DU sampling weights. This feature minimizes the variance inflation that results from unnecessary variation in sampling weights.

- Equalize the expected number of sample persons per cluster to balance the interviewing workload and to facilitate the assignment of interviewers to regions and segments. This feature also minimizes adverse effects on precision resulting from extreme cluster size variations.

- Simplify the size measure data requirements so that census data (block-level counts) are adequate to implement the method.

Using the 2000 census data supplemented with revised population projections, a composite size measure was computed for each census block defined within the United States.

The composite size measure began by defining the rate $f_h(d)$ at which we wished to sample each age group domain d (d = 1,...,5 for 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 years or older) from State h.

Let $C_{hijk}(d)$ be the population count from domain d in census block k of segment j of State sampling (SS) region i within each State h. The composite size measure for block k was defined as

$$S_{hijk} = \sum_{d=1}^{5} f_h(d) \quad C_{hijk}(d). \tag{2}$$

The composite size measure for segment j was calculated as

$$S_{hij+} = \sum_{d=1}^{5} f_h(d) \quad \sum_{k=1}^{N_{hij}} C_{hijk}(d), \tag{3}$$

where N_{hij} equals the number of blocks within segment j of SS region i and State h.

2.2 Stratification

Because the 5-year NSDUH design provides for estimates by State in all 50 States plus the District of Columbia, States may be viewed as the first level of stratification. The objective of the next level of stratification was to distribute the number of interviews, in expectation, equally among SS regions. Within each State, census tracts were joined to form mutually exclusive and exhaustive SS regions of approximately equal sizes (aggregate composite size measures of roughly 100). Using desktop computer mapping software, the regions were formed, taking into account geographical boundaries, such as mountain ranges and rivers, to the extent possible. Therefore, the resulting regions facilitated ease of access and distributed the workload evenly among regions. Twelve SS regions were formed in each State, except in California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas, where 48 regions were formed.[14]

[14] The design called for 300 persons in each of three age groups (12 to 17, 18 to 25, and 26 or older) equally allocated to four quarters within each small sample State. Based on an analysis of the cost variance tradeoffs, an average cluster size of 3.125 persons in each of the three age groups (or an average of 9.375 persons over the three age groups combined) was considered near optimal. When applied to the small States, a quarterly sample of 75 persons per quarter per age group could be obtained from 24 clusters or area segments. For unbiased variance estimation purposes, at least two observations are required per stratum (Chromy, 1981); maximum geographic stratification was obtained by defining 12 strata (SS regions) with 2 area segments each per quarter. Two additional segments were selected for each of the other three quarters, yielding 8 area segments per stratum, or 96 area segments per small sample State. This approach supported a target sample size for the small States of 300 persons per age group, or a total of 900 for the year. In the large sample States, 4 times as large a sample was required. Optimum cluster size configuration and maximum stratification given the need for unbiased variance estimation were maintained by simply quadrupling the number of strata (SS regions) to 48 per large sample State, yielding a sample 300 persons per age group per quarter, 1,200 per age group over four quarters, and 3,600 per year over all three age groups.

2.3 First- and Second-Stage Sample Selection

Once the SS regions were formed, the first-stage sampling units were formed by collapsing adjacent census tracts within regions as needed. Although most census tracts contained 150 DUs in urban areas and 100 DUs in rural areas, some had to be collapsed in order to meet the minimum requirement. Once first-stage sampling units were formed, a probability proportional to the size sample was selected with minimum replacement within each SS region. The sampling frame was stratified implicitly by sorting the first-stage sampling units by a CBSA/SES (core-based statistical area/socioeconomic status) indicator and by the percentage of the population that is non-Hispanic and white. Table 2.1 summarizes the census tract sampling frame by State.

Table 2.1 Number of Census Tracts and Segments on Sampling Frame, by State

State	State Abbreviation	State FIPS Code	Number of Census Tracts on Sampling Frame	Total Number of Census Tracts/Segments Selected	Number of Segments on Sampling Frame	Number Selected for 5-Year Sample	Unique Segments in 5-Year Sample
Total U.S.			64,505	43,200	382,598		
Northeast							
Connecticut	CT	09	807	576	5,095	288	287
Maine	ME	23	343	576	3,533	288	287
Massachusetts	MA	25	1,355	576	6,163	288	288
New Hampshire	NH	33	272	576	3,076	288	286
New Jersey	NJ	34	1,914	576	5,657	288	288
New York	NY	36	4,738	2,304	19,057	1,152	1,149
Pennsylvania	PA	42	3,088	2,304	21,704	1,152	1,150
Rhode Island	RI	44	233	576	2,305	288	283
Vermont	VT	50	179	576	1,648	288	285
Midwest							
Illinois	IL	17	2,901	2,304	20,733	1,152	1,147
Indiana	IN	18	1,408	576	6,863	288	287
Iowa	IA	19	790	576	5,366	288	288
Kansas	KS	20	719	576	5,120	288	288
Michigan	MI	26	2,689	2,304	18,765	1,152	1,148
Minnesota	MN	27	1,293	576	5,955	288	288
Missouri	MO	29	1,303	576	7,193	288	287
Nebraska	NE	31	495	576	4,075	288	288
North Dakota	ND	38	215	576	1,618	288	279
Ohio	OH	39	2,902	2,304	20,342	1,152	1,149
South Dakota	SD	46	212	576	2,001	288	284
Wisconsin	WI	55	1,310	576	6,773	288	288

(continued)

Table 2.1 Number of Census Tracts and Segments on Sampling Frame, by State (continued)

State	State Abbreviation	State FIPS Code	Number of Census Tracts on Sampling Frame	Total Number of Census Tracts/Segments Selected	Number of Segments on Sampling Frame	Number Selected for 5-Year Sample	Unique Segments in 5-Year Sample
South							
Alabama	AL	01	1,079	576	6,958	288	288
Arkansas	AR	05	618	576	6,128	288	288
Delaware	DE	10	196	576	1,721	288	282
District of Columbia	DC	11	179	576	1,049	288	270
Florida	FL	12	3,140	2,304	25,374	1,152	1,150
Georgia	GA	13	1,609	576	7,682	288	288
Kentucky	KY	21	992	576	6,301	288	288
Louisiana	LA	22	1,099	576	5,841	288	288
Maryland	MD	24	1,204	576	5,477	288	288
Mississippi	MS	28	601	576	6,448	288	287
North Carolina	NC	37	1,550	576	8,708	288	287
Oklahoma	OK	40	977	576	5,654	288	286
South Carolina	SC	45	862	576	7,365	288	288
Tennessee	TN	47	1,246	576	7,534	288	288
Texas	TX	48	4,351	2,304	26,096	1,152	1,152
Virginia	VA	51	1,513	576	6,448	288	286
West Virginia	WV	54	466	576	4,319	288	287
West							
Alaska	AK	02	154	576	1,348	288	283
Arizona	AZ	04	1,089	576	6,759	288	287
California	CA	06	6,978	2,304	22,973	1,152	1,152
Colorado	CO	08	1,050	576	6,231	288	288
Hawaii	HI	15	274	576	1,784	288	285
Idaho	ID	16	277	576	3,224	288	285
Montana	MT	30	256	576	2,417	288	288
Nevada	NV	32	474	576	3,919	288	288
New Mexico	NM	35	431	576	3,839	288	288
Oregon	OR	41	752	576	6,219	288	288
Utah	UT	49	485	576	4,024	288	288
Washington	WA	53	1,312	576	6,425	288	288
Wyoming	WY	56	125	576	1,291	288	283

FIPS = Federal Information Processing Standards.

To form segments within sampled census tracts, adjacent census blocks were collapsed until the total number of DUs within the area was at least 150 in urban areas and 100 in rural areas. Latitude, longitude, and sorting within tracts were used to obtain geographic ordering of the blocks. One segment was selected per sampled census tract with probability proportionate to the size of the segment.

As Table 2.1 indicates, 48 census tracts/segments per SS region were chosen for a total of 576 segments in each State, except in the large States where a total of 2,304 segments were

chosen. Although only 24 segments per SS region were needed to support the 5-year study, an additional 24 segments were selected to serve as replacements when segment lines are depleted and/or to support any supplemental studies embedded within NSDUH.

2.4 Survey Year and Quarter Assignment

The 48 sampled segments per SS region were randomly assigned to survey years by drawing 6 equal probability subsamples of 4 segments. Prior to selecting the second subsample, the first subsample segments were removed from the pool of eligible segments. The second subsample then was selected from the remaining segments. This process was repeated 5 times until the 48 sampled segments were assigned to 6 subsamples of 4 and a "reserve" sample of 24 segments.

The first subsample of segments was assigned to the 2005 NSDUH and constituted the panel of segments to be used for that year only. The second subsample of segments was assigned to the 2005 NSDUH and was used again in the 2006 survey; the third was assigned to the 2006 and 2007 surveys; and so on. Within each subsample, segments were assigned to survey quarters 1 through 4 in the order that they were selected.

Using the survey year and quarter assignments, a sequential segment identification number (SEGID) then was assigned. Table 2.2 describes the relationship between segment identification numbers and quarter assignment. The last two digits in the SEGID are called the "segment suffix." The 5-year survey corresponds to segment suffixes 1 through 24.

2.5 Creation of Variance Estimation Strata

The nature of the stratified, clustered sampling design requires that the design structure be taken into consideration when computing variances of survey estimates. Key nesting variables were created to capture explicit stratification and to identify clustering. For the 2005-2009 NSDUHs, each SS region appears in a different variance estimation stratum every quarter.

To define the variance estimation strata for the 2005-2009 NSDUHs, the 900 SS regions were first placed in random order (States were randomly sorted, and regions were randomly sorted within States). This list, numbered 1 to 900, defined the quarter 1 regions in the strata. For the quarter 2 regions, SS region 151 was assigned to stratum 1, and the stratum assignments continued down the list through the 900 SS regions, wrapping to the beginning of the list once the end was reached (i.e., SS region number 900 was assigned to stratum 750, SS region number 1 was assigned to stratum 751, etc.). For the quarter 3 regions, another 150 were added to the previous starting point (i.e., the new starting point was SS region number 301), and regions were assigned to strata starting from there. As before, the assignments wrapped to the beginning of the list, once the end was reached (i.e., SS region number 900 was assigned to stratum 600, SS region number 1 was assigned to stratum 601, etc.). And finally for the fourth quarter regions, 150 were added to the third starting point (or 450 from the beginning of the list), and regions were assigned to strata from there. This method had the effect of assigning the regions to strata in a pseudo-random fashion while ensuring that each stratum consists of four SS regions from four different States.

Table 2.2 Segment Identification Number Suffixes and Quarter Assignment

Segment Suffix	2005 NSDUH	2006 NSDUH	2007 NSDUH	2008 NSDUH	2009 NSDUH	Variance Replicate
01	x (Q1)					1
02	x (Q2)					1
03	x (Q3)					1
04	x (Q4)					1
05	x (Q1)	x (Q1)				2
06	x (Q2)	x (Q2)				2
07	x (Q3)	x (Q3)				2
08	x (Q4)	x (Q4)				2
09		x (Q1)	x (Q1)			1
10		x (Q2)	x (Q2)			1
11		x (Q3)	x (Q3)			1
12		x (Q4)	x (Q4)			1
13			x (Q1)	x (Q1)		2
14			x (Q2)	x (Q2)		2
15			x (Q3)	x (Q3)		2
16			x (Q4)	x (Q4)		2
17				x (Q1)	x (Q1)	1
18				x (Q2)	x (Q2)	1
19				x (Q3)	x (Q3)	1
20				x (Q4)	x (Q4)	1
21					x (Q1)	2
22					x (Q2)	2
23					x (Q3)	2
24					x (Q4)	2

Note: The segment suffix is defined as the last two digits of the segment identification number.

The 2005-2009 definition of variance estimation strata has the effect of increasing the number of degrees of freedom for State-level estimates while preserving the number of degrees of freedom for national estimates (900). Each small sample State is in 48 different strata; therefore, there are 48 degrees of freedom available for State estimates. Similarly, each large sample State is in 192 strata and therefore has 192 degrees of freedom for estimation.

Two replicates per year were defined within each variance stratum. Each variance replicate consists of four segments, one for each quarter of data collection. The first replicate consists of those segments that are "phasing out" or will not be used in the next survey year. The second replicate consists of those segments that are "phasing in" or will be fielded again the following year, thus constituting the 50 percent overlap between survey years. Table 2.2 describes the assignment of segments to variance estimation replicates.

All weighted statistical analyses for which variance estimates are needed should use the stratum and replicate variables to identify nesting. Variance estimates can be computed using a clustered data analysis software package such as SUDAAN® (RTI International, 2004). The SUDAAN® software package computes variance estimates for nonlinear statistics using such procedures as a first-order Taylor series approximation of the deviations of estimates from their expected values. The approximation is unbiased for sufficiently large samples. SUDAAN® also recognizes positive covariance among estimates involving data from 2 or more years.

Chapter 3. General Sample Allocation Procedures for the Main Study

In this chapter, the computational details of the procedural steps used to determine both person and dwelling unit (DU) sample sizes are discussed. The within-DU age group–specific selection probabilities for the design of the main study of the 2006 National Survey on Drug Use and Health (NSDUH) also are addressed. This optimization procedure was designed specifically to address the Substance Abuse and Mental Health Services Administration's (SAMHSA's) multiple precision and design requirements while simultaneously minimizing the cost of data collection. Costs were minimized by determining the smallest number of interviews and selected DUs necessary to achieve the various design requirements. In summary, this three-step optimization procedure proceeded as follows:

1. In the first step, we determined the optimal number of interviews (i.e., responding persons) by domains of interest needed to satisfy the precision requirements for several drug use outcome measures. In other words, we initially sought to determine 255 unknown m_{ha} values for each State h (51) and age group a (5). A solution to this multiple constraint optimization was achieved using Chromy′s Algorithm (Chromy, 1987). This is described in further detail in Section 3.2.

2. Using the m_{ha} determined from Step 1, the next step was to determine the optimal number of selected dwelling (D_{hj}) units (i.e., third-stage sample) necessary. This step was achieved by applying parameter constraints (e.g., probabilities of selection and expected response rates) at the segment level j or the stage at which DUs would be selected, which was done on a quarterly basis using approximately 25 percent of the m_{ha} values. This step is described in further detail in Section 3.3.

3. The final step in this procedure entailed determining age group–specific probabilities of selection (S_{hja}) for each segment given the m_{ha} and D_{hj} from Steps 1 and 2. This was achieved using a modification of Brewer's Method of Selection (Cochran, 1977, pp. 261-263). The modification was designed to select 0, 1, or 2 persons from each DU.[15] A detailed discussion of the final step is given in Section 3.4. After calculating the required DUs and the selection probabilities, we applied sample size constraints to ensure adequate samples for supplemental studies and to reduce the field interviewer (FI) burden. Limits on the total number of expected interviews per segment also were applied. This process became iterative to reallocate the reduction in sample size to other segments not affected by such constraints. Details of this step in the optimization procedure are given in Section 3.5.

3.1 Notation

h = 50 States plus the District of Columbia.

[15] Direct application of Brewer's method would require a fixed sample size.

a = Age group a = 1,...,5 and represents the following groups: 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older.

j = Individual segment indicator (total of 7,200; 1,800 per quarter).

m_{ha} = Number of completed interviews (person respondents) desired in each State h and age group a. Computation of m_{ha} is discussed in Section 3.2. For quarterly computation of selected DU sample size, approximately 25 percent of the yearly estimate is used.

y_{ha} = Estimated number of persons in the target population in State h and age group a. The 2006 population is estimated using the 2000 census data adjusted to the 2007 Claritas population projections in the compound interest formula, $y = Ae^{Bx}$, where

y = population at time x,
A = initial population,
e = base of the system of natural logarithms,
B = growth rate per unit of time, and
x = period of time over which growth occurs.

First, B is computed as $[ln(y/A)]/x$, where y = the population in 2007, A = the population in 2000, and x = 7. Then the 2006 population (y^*_{ha}) is computed using the original formula and this time allowing x to be 6. Finally, the 2006 population is adjusted by the ratio of estimated eligible listed DUs to the Claritas DU counts (U_{hj}). This adjustment factor considers the number of added DUs expected to be obtained through the half-open interval rule (1.01) and the probability of a DU being eligible (ε_h), both determined via historic data. The coefficient adjustment of 1.01 is estimated using historical data and is the proportion of all screened DUs (includes added DUs) over the original total of selected DUs (excluding added DUs). So, $y_{ha} = \{[1.01 * \varepsilon_h * L_{hj} * (1/I_{hj}) / U_{hj}]\} * y^*_{ha}$, where ε_h, L_{hj}, and I_{hj} are defined further below. This adjustment is computed at the census block level and then aggregated to the State level.

f_{ha} = m_{ha} / y_{ha}. State-specific age group sampling fraction.

F_h = $Max[f_{ha} / (\phi_h * \lambda_{ha} * \delta_{ha}), a = 1\text{-}5]$.

P_{hj} = Inverse of the segment selection probability (includes the census tract selection probability). DU sample sizes are computed on a quarterly basis, and segments are selected on a yearly basis. Because each quarter contains only a fourth of the selected segments, these probabilities are adjusted by a factor of 4 so that weights will add to the yearly totals.

I_{hj} = Subsegmentation inflation factor. For segments too large to count and to list efficiently in both time and cost, field listing personnel are allowed to subsegment the segment into roughly equal-sized subdivisions. They perform a quick count (best guess: L^*_{hj}) of the entire segment and then subdivide (taking also a best guess estimate of the number of DUs

in each subsegment: B^*_{hj}). Using a selection algorithm provided by RTI International, one subsegment is selected for regular counting and listing. For the subsegment to represent the entire segment, the weights are adjusted up to reflect the unused portion of the segment.

$= (B^*_{hj} / L^*_{hj})$.

= 1, if no subsegmenting was done.

D_{hj} = Minimum number of DUs to select for screening in segment j to meet the targeted sample sizes for all age groups.

L_{hj} = Final segment count of DUs available for screening.

S_{hja} = State- and segment-specific probability of selecting a person in age group a. One implemented design constraint was that no single age group selection probability could exceed 1. The maximum allowable probability was then set to 0.99.

ε_h = State-specific DU eligibility rate. This rate was derived from 2004 NSDUH quarter 4 and 2005 NSDUH quarters 1 through 3 data by taking the average eligibility rate within each State.

ϕ_h = State-specific screening response rates. These rates were calculated using the same methodology as described for the DU eligibility rate (ε_h).

λ_{ha} = State- and age group–specific interview response rate. Using data from quarter 4 of the 2004 NSDUH and quarters 1 through 3 of the 2005 NSDUH, the additive effects of State and age group on interview response were determined by taking the average interview response rate within each State.

γ_{ha} = Expected number of persons within an age group per DU. This number was calculated using 2004 NSDUH quarter 4 and 2005 NSDUH quarters 1 through 3 data by dividing the weighted total number of rostered persons in an age group by the weighted total number of complete screened DUs by State.

δ_{ha} = State- and age group–specific maximum-of-two rule adjustment. The survey design restricts the number of interviews per DU to a total of two. This is achieved through a modified Brewer's Method of Selection, which results in a loss of potential interviews in DUs where selection probabilities sum greater than 2. The adjustment is designed to inflate the number of required DUs to compensate for this loss. Using data from all four quarters of the 2004 NSDUH, the adjustment was computed by taking the average maximum-of-two rule adjustment within each State.

3.2 Determining Person Sample Sizes by State and Age Group

The first step in the design of the fourth stage of selection was to determine the optimal number of respondents needed in each of the 255 domains to minimize the costs associated with data collection, subject to multiple precision requirements established by SAMHSA. In summary, the precision requirements were that the expected relative standard error (RSE) on a prevalence of 10 percent not exceed the following:

- 3.00 percent for total population statistics, and
- 5.00 percent for statistics in three age group domains: 12 to 17, 18 to 25, and 26 or older.

In preparation for the 2005-2009 NSDUHs, several optimization models and other related analyses were conducted. Using historical 2001 survey data, estimates and RSEs for each of nine outcome measures of interest were computed. Estimates then were standardized to a prevalence of 10 percent. The outcome measures of interest were included to address not only the NSDUH recency-of-use estimates, but also such related generic substance abuse measures as treatment received for alcohol and illicit drug use and dependence on alcohol and illicit drugs.

Specifically, the nine classes of NSDUH outcomes we considered were as follows:

Use of Legal (Licit) Substances

1. *Cigarette Use in the Past Month.* Smoked cigarettes at least once within the past month.
2. *Alcohol Use in the Past Month.* Had at least one drink of an alcoholic beverage (beer, wine, liquor, or a mixed alcohol drink) within the past month.

Use of Illicit Substances

3. *Illicit Drug Use in the Past Month.* Includes use of hallucinogens, heroin, marijuana, cocaine, inhalants, opiates, or nonmedical use of sedatives, tranquilizers, stimulants, or pain relievers.
4. *Illicit Drug Use Other Than Marijuana in the Past Month.* Past month use of illicit drugs excluding those whose only illicit drug use was marijuana.
5. *Cocaine Use in the Past Month.* Use within the past month of cocaine in any form, including crack.

Note that current use of illicit drugs provides a broad measure of illicit drug use; however, it is dominated by marijuana and cocaine use. Therefore, estimates of marijuana use and cocaine use are included because these two measures reflect different types of drug abuse.

Drug or Alcohol Dependence

6. *Dependent on Illicit Drugs in the Past Year.* Dependent on the same drugs listed in class 3, *Illicit Drug Use in the Past Month,* above. Those who are dependent on both alcohol and another illicit substance are included, but those who are dependent on alcohol only are not.

7. *Dependent on Alcohol and Not Illicit Drugs in the Past Year.* Dependent on alcohol and not dependent on illicit drugs.

Treatment for Drugs and Alcohol Problems

8. *Received Treatment for Illicit Drugs in the Past Year.* Received treatment in the past 12 months at any location (including hospitals, clinics, self-help groups, or doctors' offices) for illicit drug use.
9. *Received Treatment for Alcohol Use but Not Illicit Drug Use in the Past Year.* Received treatment in the past 12 months at any location (including hospitals, clinics, self-help groups, or doctors' offices) for drinking. These estimates exclude those who received treatment in the past 12 months for both drinking and illicit drug use.

These outcome measures, as well as the precision that is expected from this 2006 NSDUH design, are presented in Table 3.1. RSEs were based on an average prevalence rate of 10 percent for each measure.

Table 3.1 Expected Relative Standard Errors, by Age Group

	Total			12-17		
Outcome Measure	**Estimate**	**RSE**	**SRSE**	**Estimate**	**RSE**	**SRSE**
Past Month Cigarette Use	0.25	1.30	2.25	0.13	2.17	2.52
Past Month Alcohol Use	0.48	0.74	2.14	0.17	1.92	2.64
Past Month Use of Illicit Drugs	0.07	2.21	1.83	0.11	2.40	2.51
Past Month Use of Illicit Drugs Other Than Marijuana	0.03	3.52	1.89	0.05	3.46	2.35
Past Month Cocaine Use	0.01	7.76	2.01	0.00	13.86	2.79
Past Year, Dependent on Illicit Drugs	0.02	4.38	1.68	0.02	5.18	2.44
Past Year, Dependent on Alcohol	0.02	4.00	1.72	0.01	6.79	2.33
Past Year, Received Treatment for Illicit Drug Use	0.01	8.61	2.26	0.01	9.04	2.74
Past Year, Received Treatment for Alcohol Use	0.00	9.46	1.88	0.00	16.42	2.34
Average Relative Standard Error			1.96			2.52
Target Relative Standard Error			3.00			5.00
	18-25			**26+**		
Outcome Measure	**Estimate**	**RSE**	**SRSE**	**Estimate**	**RSE**	**SRSE**
Past Month Cigarette Use	0.39	1.20	2.89	0.24	1.68	2.85
Past Month Alcohol Use	0.59	0.84	3.02	0.51	0.91	2.76
Past Month Use of Illicit Drugs	0.19	2.17	3.13	0.05	4.04	2.64
Past Month Use of Illicit Drugs Other Than Marijuana	0.08	3.39	2.97	0.02	6.39	2.77
Past Month Cocaine Use	0.02	6.80	2.85	0.01	12.12	2.78
Past Year, Dependent on Illicit Drugs	0.05	4.71	3.15	0.01	8.87	2.63
Past Year, Dependent on Alcohol	0.05	4.07	2.66	0.02	5.95	2.34
Past Year, Received Treatment for Illicit Drug Use	0.01	7.73	2.64	0.01	12.82	3.08
Past Year, Received Treatment for Alcohol Use	0.01	10.82	2.61	0.00	12.25	2.41
Average Relative Standard Error			2.88			2.70
Target Relative Standard Error			5.00			5.00

RSE = relative standard error.
SRSE = standardized relative standard error.

Additionally, initial sample size requirements were implemented:

- Minimum sample size of 3,600 persons per State in the eight large States and 900 persons in the remaining 43 States.

- Equal allocation of the sample across the three age groups: 12 to 17, 18 to 25, and 26 or older within each State.

Similar to the 1999 through 2005 surveys, racial groups were not oversampled for the 2006 NSDUH. Consistent with previous surveys, the 2006 NSDUH was designed to oversample the younger age groups.

Among the 51 States, a required total sample size of 67,500 respondents was necessary to meet all precision and sample size requirements. Table 3.2 shows expected State by age group sample sizes. Because of the shorter calendar length of quarters 1 and 4 (due to interviewer training and the holidays, respectively), a decision was made to allocate the quarterly State by age group sample sizes (25 percent of the annual sample) to the four quarters in ratios of 96, 104, 104, and 96 percent, respectively. Only minor increases in unequal weighting resulted from not distributing the sample equally across quarters.

3.3 Third-Stage Sample Allocation for Each Segment

Given the desired respondent sample size for each State and age group (m_{ha}) needed to meet the design parameters established by SAMHSA, the next step was to determine the minimal number of DUs to select for each segment to meet the targeted sample sizes. In short, this step involved determining the sample size of the third stage of selection. This sample size determination was performed on a quarterly basis to take advantage of both segment differences and, if necessary, make adjustments to design parameters. Procedures described below were developed originally for initial implementation in quarter 1 of the survey. The description is specific to quarter 1. Any modifications or corrections were made in subsequent quarters and are explained in detail in Section 3.8.

3.3.1 Dwelling Unit Frame Construction—Counting and Listing

The process by which the DU frame is constructed is called counting and listing. In summary, a certified lister visits the selected area and lists a detailed and accurate address (or description, if no address is available) for each DU within the segment boundaries. The lister is given a series of maps on which to mark the locations of these DUs. The number of map pages per State and the average number of map pages per segment are summarized in Table 3.3. The resulting list of DUs is entered into a database and serves as the frame from which the third-stage sample is drawn.

In some situations, the number of DUs within the segment boundaries was much larger than the specified maximum. To obtain a reasonable number of DUs for the frame, the lister first counted the DUs in such an area. The sampling staff at RTI International then partitioned the segment into smaller pieces or subsegments and randomly selected one to be listed. The number of

segments that were subsegmented in the 2006 NSDUH sample is summarized in Table 3.4. For more information on the subsegmenting procedures, see Appendix B.

During counting and listing, the lister moves about the segment in a prescribed fashion called the "continuous path of travel." The lister attempts to move in a clockwise fashion, makes each possible right turn, makes U-turns at segment boundaries, and does not break street sections. Following these defined rules and always looking for DUs on the right-hand side of the street, the lister minimizes the chance of not listing a DU within the segment. Also, using a defined path of travel makes it easier for the field interviewer (FI) assigned to the segment to locate the sampled DUs. Finally, the continuous path of travel lays the groundwork for the half-open interval procedure for recovering missed DUs, as described in Section 3.7 of this report. A detailed description of the counting and listing procedures is provided in the *2006 NSDUH: Counting and Listing General Manual* (RTI International, 2005).

Table 3.2 Expected Main Study Sample Sizes, by State and Age Group

State	State FIPS	SS Regions	Total Segments	Total Respondents					
				12-17	18-25	26-34	35-49	50+	Total
Total Population		900	7,207	22,500	22,500	6,382	9,459	6,659	67,500
Northeast									
Connecticut	09	12	96	300	300	78	130	92	900
Maine	23	12	96	300	300	72	131	97	900
Massachusetts	25	12	96	300	300	85	126	89	900
New Hampshire	33	12	96	300	300	77	136	87	900
New Jersey	34	12	96	300	300	83	128	89	900
New York	36	48	384	1,200	1,200	349	498	353	3,600
Pennsylvania	42	48	384	1,200	1,200	306	500	394	3,600
Rhode Island	44	12	96	300	300	80	126	94	900
Vermont	09	12	96	300	300	72	134	94	900
Midwest									
Illinois	17	48	384	1,200	1,200	360	501	339	3,600
Indiana	18	12	96	300	300	85	126	89	900
Iowa	19	12	96	300	300	78	124	98	900
Kansas	20	12	96	300	300	83	127	90	900
Michigan	26	48	384	1,200	1,200	335	512	353	3,600
Minnesota	27	12	96	300	300	84	131	85	900
Missouri	29	12	96	300	300	81	126	93	900
Nebraska	31	12	96	300	300	83	126	91	900
North Dakota	38	12	96	300	300	78	127	95	900
Ohio	39	48	384	1,200	1,200	327	508	365	3,600
South Dakota	46	12	96	300	300	78	127	95	900
Wisconsin	55	12	96	300	300	81	129	90	900

(continued)

Table 3.2 Expected Main Study Sample Sizes, by State and Age Group (continued)

State	State FIPS	SS Regions	Total Segments	Total Respondents					
				12-17	18-25	26-34	35-49	50+	Total
South									
Alabama	01	12	96	300	300	84	123	93	900
Arkansas	05	12	96	300	300	82	121	97	900
Delaware	10	12	96	300	300	84	126	90	900
District of Columbia	11	12	96	300	300	103	114	83	900
Florida	12	48	384	1,200	1,200	312	477	411	3,600
Georgia	13	12	96	300	300	97	127	76	900
Kentucky	21	12	96	300	300	86	125	89	900
Louisiana	22	12	101	300	300	86	127	87	900
Maryland	24	12	96	300	300	84	132	84	900
Mississippi	28	12	98	300	300	87	123	90	900
North Carolina	37	12	96	300	300	91	123	86	900
Oklahoma	40	12	96	300	300	82	124	94	900
South Carolina	45	12	96	300	300	86	124	90	900
Tennessee	47	12	96	300	300	87	124	89	900
Texas	48	48	384	1,200	1,200	386	506	308	3,600
Virginia	51	12	96	300	300	88	129	83	900
West Virginia	54	12	96	300	300	76	122	102	900
West									
Alaska	02	12	96	300	300	88	144	68	900
Arizona	04	12	96	300	300	92	120	88	900
California	06	48	384	1,200	1,200	383	502	315	3,600
Colorado	08	12	96	300	300	93	131	76	900
Hawaii	15	12	96	300	300	85	125	90	900
Idaho	16	12	96	300	300	86	127	87	900
Montana	30	12	96	300	300	71	131	98	900
Nevada	32	12	96	300	300	92	123	85	900
New Mexico	35	12	96	300	300	83	128	89	900
Oregon	41	12	96	300	300	85	125	90	900
Utah	49	12	96	300	300	106	120	74	900
Washington	53	12	96	300	300	86	130	84	900
Wyoming	56	12	96	300	300	76	133	91	900

FIPS = Federal Information Processing Standards; SS = State sampling.

Table 3.3 Number of Map Pages, by State and Segment

State	Total Segments	Cumulative Number of Map Pages per State	Average Number of Map Pages per Segment
Total Population	7,207	43,451	6.0
Alabama	96	698	7.3
Alaska	96	635	6.6
Arizona	96	503	5.2
Arkansas	96	798	8.3
California	384	1,543	4.0
Colorado	96	559	5.8
Connecticut	96	415	4.3
Delaware	96	484	5.0
District of Columbia	96	334	3.5
Florida	384	1,767	4.6
Georgia	96	506	5.3
Hawaii	96	384	4.0
Idaho	96	830	8.6
Illinois	384	2,058	5.4
Indiana	96	653	6.8
Iowa	96	681	7.1
Kansas	96	820	8.5
Kentucky	96	586	6.1
Louisiana	101	573	5.7
Maine	96	675	7.0
Maryland	96	355	3.7
Massachusetts	96	502	5.2
Michigan	384	1,962	5.1
Minnesota	96	553	5.8
Mississippi	98	770	7.9
Missouri	96	648	6.8
Montana	96	1,077	11.2
Nebraska	96	894	9.3
Nevada	96	462	4.8
New Hampshire	96	553	5.8
New Jersey	96	454	4.7
New Mexico	96	892	9.3
New York	384	1,432	3.7
North Carolina	96	506	5.3
North Dakota	96	1,072	11.2
Ohio	384	2,008	5.2
Oklahoma	96	658	6.9
Oregon	96	556	5.8
Pennsylvania	384	2,339	6.1
Rhode Island	96	545	5.7
South Carolina	96	617	6.4
South Dakota	96	988	10.3
Tennessee	96	602	6.3
Texas	384	2,053	5.3

(continued)

Table 3.3 Number of Map Pages by State and Segment (continued)

State	Total Segments	Cumulative Number of Map Pages per State	Average Number of Map Pages per Segment
Utah	96	630	6.6
Vermont	96	642	6.7
Virginia	96	429	4.5
Washington	96	512	5.3
West Virginia	96	784	8.2
Wisconsin	96	588	6.1
Wyoming	96	1,866	19.4

Table 3.4 Segment and Dwelling Unit Summary

State	Total Segments	Subsegmented Segments	Listed Dwelling Units	Added Dwelling Units
Total Population	7,207	577	1,501,304	1,634
Alabama	96	6	19,488	20
Alaska	96	10	23,135	25
Arizona	96	9	20,092	19
Arkansas	96	4	18,126	16
California	384	28	82,460	47
Colorado	96	5	20,238	13
Connecticut	96	6	20,861	39
Delaware	96	11	20,932	26
District of Columbia	96	11	25,155	38
Florida	384	70	82,774	64
Georgia	96	14	23,286	10
Hawaii	96	18	22,582	19
Idaho	96	9	17,953	24
Illinois	384	20	79,475	80
Indiana	96	1	19,723	18
Iowa	96	2	17,908	16
Kansas	96	2	18,686	16
Kentucky	96	5	18,647	28
Louisiana	101	9	21,803	37
Maine	96	3	20,164	32
Maryland	96	14	22,301	30
Massachusetts	96	5	20,669	38
Michigan	384	20	75,224	64
Minnesota	96	4	19,778	23
Mississippi	98	3	18,440	22
Missouri	96	5	20,249	14
Montana	96	4	16,451	16
Nebraska	96	1	17,856	26
Nevada	96	13	17,522	9
New Hampshire	96	6	19,352	71
New Jersey	96	4	21,252	10
New Mexico	96	9	20,729	5

(continued)

Table 3.4 Segment and Dwelling Unit Summary (continued)

State	Total Segments	Subsegmented Segments	Listed Dwelling Units	Added Dwelling Units
New York	384	58	84,326	104
North Carolina	96	4	20,260	23
North Dakota	96	6	20,414	26
Ohio	384	25	82,684	42
Oklahoma	96	7	19,713	8
Oregon	96	7	19,151	49
Pennsylvania	384	12	78,091	82
Rhode Island	96	7	20,424	60
South Carolina	96	10	18,972	9
South Dakota	96	12	17,362	26
Tennessee	96	6	18,640	33
Texas	384	49	79,884	29
Utah	96	6	18,431	14
Vermont	96	4	19,408	60
Virginia	96	13	21,783	45
Washington	96	7	20,957	38
West Virginia	96	1	20,985	20
Wisconsin	96	4	18,219	27
Wyoming	96	8	18,289	24

3.3.2 Determining Dwelling Unit Sample Size

For the main study, the optimization formula is as follows:

$$f_{ha} = P_{hj} * I_{hj} * (\frac{D_{hj}}{L_{hj}}) * S_{hja} * \phi_h * \lambda_{ha} * \delta_{ha}. \quad (4)$$

At this point in the procedure, only two components in the formula are unknown: D_{hj} and S_{hja}. Selection probabilities are segment- and age group–specific, and to maximize the number of selected persons within a DU, the age group whose adjusted sampling fraction $[f_{ha} / (\phi_h * \lambda_{ha} * \delta_{ha})] = F_h$, known now as the driving age group, is set to the largest allowable selection probability (S_{hja}) of 0.99. D_{hj} then is computed as

$$D_{hj} = \frac{f_{ha}}{(P_{hj} * I_{hj} * S_{hja} * \phi_h * \lambda_{ha} * \delta_{ha})} * L_{hj}. \quad (5)$$

3.4 Determining Fourth-Stage Sample (Person) Selection Probabilities for Each Segment

$$S_{hja} = \frac{f_{ha}}{P_{hj} * I_{hj} * (\frac{D_{hj}}{L_{hj}}) * \phi_h * \lambda_{ha} * \delta_{ha}}. \quad (6)$$

Having solved for D_{hj}, the selection probabilities for the remaining age groups were solved. If L_{hj} equals 0, D_h and S_{hja} are set to 0.

3.5 Sample Size Constraints: Guaranteeing Sufficient Sample for Additional Studies and Reducing Field Interviewer Burden

A major area of interest for the survey is to ensure that an adequate sample of eligible DUs remain within each segment. This sample surplus is needed to allow SAMHSA to implement supplemental studies if desired.

In addition, concern was noted about guaranteeing that FIs would be able to complete the amount of work assigned to them within the quarterly timeframe. These concerns prompted adjustments to the D_{hj} sample size:

1. Number of selected DUs for screening: <100 or <½L_{hj}. Adjustments were made by adjusting the D_{hj} counts to equal the minimum of 100 or ½L_{hj}.

2. Number of selected DUs: > 5. For cost purposes, if at least five DUs remain in the segment, the minimum number of selected DUs was set to five.

3. Expected number of interviews: <40.

This expected number of interviews (m^*_{hja}) was computed as follows:

$$m^*_{hja} = D^*_{hj} * \varepsilon_h * \phi_h * \gamma_{ha} * S_{hja} * \lambda_{ha} * \delta_{ha}, \quad (7)$$

where D^*_{hj} has been adjusted for constraint 1. This value is the total number of interviews expected within each segment. The calculation of the first adjustment, the screening adjustment, is

$$5 / D^*_{hj} . \quad (8)$$

Similarly, the interview adjustment is computed as

$$40 / m^*_{hja} . \quad (9)$$

This second adjustment is applied to D_{hj} under the assumption of an equal number of screened DUs for each completed interview.

Both constraints 1 and 3 reduce the third-stage sample, which could in turn reduce the expected fourth-stage sample size. Therefore, the reduction in the third-stage sample is reallocated back to the segments by applying a marginal adjustment to the fourth-stage sample size (m_{ha}) at the State and age group level. As a result, segments that were not subject to these constraints could be affected. This adjustment to reallocate the DU sample is iterative until the expected person sample sizes are met.

3.6 Dwelling Unit Selection and Release Partitioning

After derivation of the required DU sample size (D_{hj}), the sample was selected from the frame of counted and listed DUs for each segment (L_{hj}). The frame was ordered in the same manner as described in Section 3.3.1, and selection was completed using systematic sampling with a random start value.

To compensate for quarterly variations in response rates and yields, a sample partitioning procedure was implemented in all quarters. The entire sample (D_{hj}) still would be selected, but only certain percentages of the total would be released into the field. An initial percentage would be released to all segments at the beginning of the quarter and, based on interquarter work projections, additional percentages would be released if field staff could handle the added workload. Each partitioning of the sample is a valid sample and helps control the amount of nonresponse without jeopardizing the validity of the study. Incidentally, a reserve sample of 20 percent also was selected, over and above the required D_{hj} sample, to allow for supplemental releases based on State experiences within each quarter. Thus, the 96 percent quarter 1 sample was increased to the 115.2 percent level. In quarter 1, the D_{hj} sample was allocated out to States in the following release percentages:

Release 1: 67 percent of entire sample (80/120, main sample + 20 percent reserve);
Release 2: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 3: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 4: 8 percent of entire sample (10/120, main sample + 20 percent reserve);
Release 5: 8 percent of entire sample (10/120, main sample + 20 percent reserve); and
Release 6: 8 percent of entire sample (10/120, main sample + 20 percent reserve).

A summary of the quarterly sample sizes and percentages released is provided in Table 3.5.

3.7 Half-Open Interval Rule and Procedure for Adding Dwelling Units

To guarantee that every DU had a chance of selection and to eliminate any bias associated with incomplete frames, NSDUH implemented a procedure called the half-open interval rule. This procedure required that the interviewer look both on the property of each selected DU and between that DU and the next listed DU for any unlisted units. When found in these specific locations, the unlisted units became part of the sample (added DUs). If the number of added DUs linked to any particular sample DU did not exceed 5, or if the number for the entire segment was less than or equal to 10, the FI was instructed to consider these DUs as part of his or her assignment. If either of these limits was exceeded, special subsampling procedures were implemented, as described in Appendix C. The number of added DUs in the 2006 NSDUH sample is summarized in Table 3.4.

3.8 Quarter-by-Quarter Deviations

This section describes corrections and/or modifications that were implemented in the process of design optimization. "Design" refers to deviations from the original proposed plan of design. "Procedural" refers to changes made in the calculation methodologies. Finally, "Dwelling Unit Selection" addresses changes that occurred after sample size derivations, specifically corrections implemented during fielding of the sample (i.e., sample partitioning as described in

Table 3.5 Quarterly Sample Sizes and Percentages Released

State	Quarter 1			Quarter 2		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
Total Population	50,603	43,279	86	54,820	48,335	88
Northeast						
Connecticut	678	567	84	658	626	95
Maine	833	832	100	871	800	92
Massachusetts	731	731	100	673	673	100
New Hampshire	776	644	83	948	708	75
New Jersey	716	596	83	717	686	96
New York	2,978	2,480	83	3,375	3,222	95
Pennsylvania	3,005	2,507	83	3,091	2,456	79
Rhode Island	708	593	84	765	672	88
Vermont	680	569	84	767	605	79
Midwest						
Illinois	2,649	2,215	84	2,806	2,686	96
Indiana	624	573	92	643	617	96
Iowa	623	521	84	696	580	83
Kansas	661	549	83	718	568	79
Michigan	2,582	2,151	83	2,718	2,269	83
Minnesota	550	457	83	600	600	100
Missouri	609	510	84	667	555	83
Nebraska	672	556	83	661	634	96
North Dakota	729	608	83	725	666	92
Ohio	2,662	2,217	83	2,957	2,715	92
South Dakota	604	507	84	624	624	100
Wisconsin	645	539	84	705	705	100

(continued)

Table 3.5 Quarterly Sample Sizes and Percentages Released (continued)

State	Quarter 1			Quarter 2		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
South						
Alabama	622	522	84	745	556	75
Arkansas	592	492	83	735	613	83
Delaware	679	566	83	732	636	87
District of Columbia	1,078	892	83	1,195	841	70
Florida	3,129	2,592	83	3,506	3,071	88
Georgia	725	663	91	774	612	79
Kentucky	660	548	83	729	638	88
Louisiana	585	491	84	674	674	100
Maryland	640	536	84	765	511	67
Mississippi	665	582	88	705	705	100
North Carolina	688	688	100	799	799	100
Oklahoma	751	625	83	809	809	100
South Carolina	701	701	100	807	807	100
Tennessee	656	543	83	665	530	80
Texas	2,268	1,879	83	2,460	2,260	92
Virginia	689	689	100	746	620	83
West Virginia	836	730	87	878	696	79
West						
Alaska	651	548	84	774	643	83
Arizona	594	523	88	622	543	87
California	2,096	1,926	92	2,373	2,274	96
Colorado	664	580	87	725	666	92
Hawaii	678	570	84	744	656	88
Idaho	625	549	88	656	541	82
Montana	779	649	83	769	645	84
Nevada	658	550	84	635	581	91
New Mexico	638	537	84	676	480	71
Oregon	745	618	83	800	633	79
Utah	452	394	87	461	408	89
Washington	607	558	92	673	616	92
Wyoming	737	616	84	803	604	75

(continued)

Table 3.5 Quarterly Sample Sizes and Percentages Released (continued)

State	Quarter 3			Quarter 4		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
Total Population	51,961	45,989	89	48,477	43,222	89
Northeast						
Connecticut	658	631	96	555	484	87
Maine	834	802	96	807	738	91
Massachusetts	670	645	96	660	518	78
New Hampshire	928	737	79	761	570	75
New Jersey	717	687	96	742	713	96
New York	3,160	2,901	92	2,954	2,705	92
Pennsylvania	2,875	2,518	88	2,493	2,383	96
Rhode Island	734	549	75	648	543	84
Vermont	723	543	75	661	606	92
Midwest						
Illinois	2,673	2,446	92	2,561	2,342	91
Indiana	689	575	83	604	554	92
Iowa	630	603	96	594	568	96
Kansas	678	592	87	569	477	84
Michigan	2,445	2,038	83	2,451	2,143	87
Minnesota	590	589	100	626	573	92
Missouri	717	627	87	633	559	88
Nebraska	620	597	96	606	554	91
North Dakota	685	655	96	645	617	96
Ohio	2,677	2,342	87	2,619	2,291	87
South Dakota	619	619	100	591	591	100
Wisconsin	693	664	96	567	494	87
South						
Alabama	707	676	96	631	472	75
Arkansas	735	701	95	634	556	88
Delaware	673	591	88	620	594	96
District of Columbia	1,195	1,045	87	1,088	988	91
Florida	3,226	2,418	75	2,868	2,393	83
Georgia	666	499	75	591	493	83
Kentucky	678	648	96	636	579	91
Louisiana	607	607	100	629	629	100
Maryland	778	681	88	618	568	92
Mississippi	662	607	92	580	475	82
North Carolina	751	716	95	778	778	100
Oklahoma	816	816	100	753	657	87
South Carolina	773	578	75	671	558	83
Tennessee	605	579	96	631	526	83

(continued)

Table 3.5 Quarterly Sample Sizes and Percentages Released (continued)

State	Quarter 3			Quarter 4		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
Texas	2,319	2,214	95	2,196	1,909	87
Virginia	762	731	96	656	545	83
West Virginia	800	596	75	737	648	88
West						
Alaska	576	459	80	624	573	92
Arizona	641	563	88	653	652	100
California	2,339	1,953	83	2,219	2,039	92
Colorado	684	654	96	650	595	92
Hawaii	680	488	72	691	666	96
Idaho	722	632	88	637	506	79
Montana	686	598	87	647	566	87
Nevada	635	584	92	608	556	91
New Mexico	651	651	100	561	535	95
Oregon	723	633	88	665	612	92
Utah	419	365	87	429	378	88
Washington	644	619	96	691	601	87
Wyoming	793	727	92	738	552	75

Section 3.6). Quarter 1 deviations are not included because the methods and procedures described above were all implemented in quarter 1. Subsequently, any changes would have been made after quarter 1.

Quarter 2

Design: An additional 20 percent reserve sample was added to the 104 percent quarterly sample to allow for supplemental releases where needed. Thus, the total quarter 2 sample was increased to the 124.8 percent level.

Procedural: To predict State response rates more accurately, the most current four quarters of data were used in the computation of State-specific yield and response rates. Thus, data from quarters 1 through 4 of the 2005 NSDUH were used to compute average yields, DU eligibility, screening response, and interviewer response rates.

Dwelling Unit Selection: The quarter 2 D_{hj} sample was partitioned into the following release percentages:

Release 1: 67 percent of entire sample (80/120, main sample + 20 percent reserve);
Release 2: 4 percent of entire sample (5/120, main sample + 20 percent reserve);

Release 3: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 4: 8 percent of entire sample (10/120, main sample + 20 percent reserve);
Release 5: 8 percent of entire sample (10/120, main sample + 20 percent reserve); and
Release 6: 8 percent of entire sample (10/120, main sample + 20 percent reserve).

Quarter 3

Design: Using the completed cases from quarter 1 and the projected number of completes from quarter 2, each State's midyear surplus/shortfall was computed. The quarter 3 104 percent sample then was adjusted by this amount. An additional 20 percent sample also was included, bringing the total quarter 3 adjusted sample to the 124.8 percent level.

Procedural: Data from quarters 2 through 4 of the 2005 NSDUH and quarter 1 of the 2006 NSDUH were used to compute State-specific average yields, DU eligibility, screening response, and interviewer response rates.

Dwelling Unit Selection: The quarter 3 D_{hj} sample was partitioned into the following release percentages:

Release 1: 67 percent of entire sample (80/120, main sample + 20 percent reserve);
Release 2: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 3: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 4: 8 percent of entire sample (10/120, main sample + 20 percent reserve);
Release 5: 8 percent of entire sample (10/120, main sample + 20 percent reserve); and
Release 6: 8 percent of entire sample (10/120, main sample + 20 percent reserve).

Quarter 4

Design: The State and age 96 percent quarterly sample sizes were adjusted to meet the yearly targets based on completed cases from quarters 1 and 2 and the projected number of completes from quarter 3. An additional 20 percent sample also was included, bringing the total quarter 4 adjusted sample to the 115.2 percent level.

Procedural: Data from quarters 3 and 4 of the 2005 NSDUH and quarters 1 and 2 of the 2006 NSDUH were used to compute State-specific average yields, DU eligibility, screening response, and interviewer response rates.

Dwelling Unit Selection: The quarter 4 D_{hj} sample was partitioned into the following release percentages:

Release 1: 67 percent of entire sample (80/120, main sample + 20 percent reserve);
Release 2: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 3: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 4: 8 percent of entire sample (10/120, main sample + 20 percent reserve);
Release 5: 8 percent of entire sample (10/120, main sample + 20 percent reserve); and
Release 6: 8 percent of entire sample (10/120, main sample + 20 percent reserve).

3.9 Impact of Hurricanes Katrina and Rita on the 2006 NSDUH Sample

At the end of August 2005, Hurricane Katrina caused large-scale damage and destruction in the coastal regions of Louisiana, Mississippi, and Alabama. Later, Hurricane Rita devastated portions of Texas and Louisiana. To minimize the impact of these hurricanes on the sample, several actions were taken.

First, we identified the areas that were most likely to be affected according to the paths of the hurricanes. The quarter 1 sample then was supplemented with the retired 2005 NSDUH quarter 1 segment in seven State sampling (SS) regions determined to be the hardest hit. In each of these regions, the sample size was not increased; the sample size was distributed among three segments rather than two.[16] Because entire segments could be ineligible due to hurricane damage, having the third segment in the SS region created an additional location from which to draw the sample.

The impact of the hurricanes on the sample was reexamined prior to selecting the quarter 2 sample, and it was determined that the sample supplement was no longer necessary. Therefore, the 2006 NSDUH sample consisted of 7,207 segments.

In addition to supplementing the quarter 1 sample, field staff were reminded to apply standard procedures to handle unusual situations. Specifically, field staff were instructed to apply

[16] Segment-level weights were adjusted for all quarter 1 segments in these SS regions.

the residency rule for eligibility[17] and to pick up displaced persons wherever they currently were residing. Additionally, temporary housing units were picked up by applying the half-open interval rule.

3.10 Sample Weighting Procedures

At the conclusion of data collection for the last quarter, design weights are constructed for each quarter of the State-level study, reflecting the various stages of sampling. The weights for the 2006 NSDUH have not been computed yet. However, the planned procedure is described in this section.

The calculation of the sampling weights will be based on the stratified, four-stage design of the study. Specifically, the person-level sampling weights will be the product of the four stagewise sampling weights, each equal to the inverse of the selection probability for that stage. In review, the stages are as follows:

Stage 1: Selection of census tract.

Stage 2: Selection of segment.

Stage 3: Selection of DU.
Three possible adjustments exist with this stage of selection:
(1) subsegmentation inflation: by-product of counting and listing;
(2) added DU: results from the half-open interval rule when subsampling is needed; and
(3) release adjustment.

Stage 4: Selection of person within a DU.

A total of seven weight adjustments will be necessary for the calculation of the final analysis sample weight. All weight adjustments will be implemented using a generalized exponential model technique. These adjustments are listed in the order in which they will be implemented:

1. *Nonresponse Adjustment at the Dwelling Unit Level.* This adjustment is to account for the failure to complete the within-DU roster. The potential list of variables for the 51-State main study DU nonresponse modeling is presented in Table 3.6.
2. *Dwelling Unit–Level Poststratification.* This adjustment involves using screener data of demographic information (e.g., age, race, gender). DU weights will be adjusted to the intercensal population estimates derived from the 2000 census for various demographic domains. In short, explanatory variables used during modeling will consist of counts of eligible persons within each DU that fall into the various demographic categories. Consequently, these counts, multiplied by the newly adjusted DU weight and summed

[17] The residency rule for eligibility requires that a person reside at a selected DU at least half of the quarter in order to be eligible for the survey.

Table 3.6 Definitions of Levels for Potential Variables for Dwelling Unit Nonresponse Adjustment

Group Quarters Indicator
1: College Dorm
2: Other Group Quarter
3: Nongroup Quarter
Percentage of Owner-Occupied Dwelling Units in Segment (% Owner)
1: 0 - <10%
2: 10% - <50%
3: 50% - 100%
Percentage of Blacks in Segment (% Black)
1: 0 - <10%
2: 10% - <50%
3: 50% - 100%
Percentage of Hispanics in Segment (% Hispanic)
1: 0 - <10%
2: 10% - <50%
3: 50% - 100%
Population Density
1: CBSA > 1,000,000
2: CBSA < 1,000,000
3: Non-CBSA Urban
4: Non-CBSA Rural
Quarter
1: Quarter 1
2: Quarter 2
3: Quarter 3
4: Quarter 4
Segment Combined Median Rent and Housing Value (Rent/Housing)
1: First Quintile
2: Second Quintile
3: Third Quintile
4: Fourth Quintile
5: Fifth Quintile
State

CBSA = core-based statistical area.
Note: Interactions among the main effect variables also are considered.

across all DUs for various domains, will add to the census population estimates. This adjustment is useful for providing more stable control totals for subsequent adjustments and pair weights. Potential explanatory variables are listed in Table 3.7.

3. *Extreme Weight Treatment at the Dwelling Unit Level.* If it is determined that design-based weights (stages 1 and 2) along with any of their respective adjustments result in an unsatisfactory unequal weighting effect (i.e., variance of the DU-level weights is too high, with high frequency of extreme weights), then extreme weights will be further adjusted. This adjustment will be implemented by doing another weight calibration. The control totals are the DU-level poststratified weights, and the same explanatory variables as in DU-level poststratification will be used so that the extreme weights are controlled and all the distributions in various demographic groups are preserved.

Table 3.7 Definitions of Levels for Potential Variables for Dwelling Unit Poststratification and Respondent Poststratification at the Person Level

Age
1: 12-17
2: 18-25
3: 26-34
4: 35-49
5: 50+[a]
Gender
1: Male
2: Female
Hispanicity
1: Hispanic
2: Non-Hispanic
Quarter
1: Quarter 1
2: Quarter 2
3: Quarter 3
4: Quarter 4
Race
1: White
2: Black
3: American Indian/Alaska Native
4: Asian
5: Two or More Races
State

Note: Interactions among the main effect variables also are considered.

[a] For person-level respondent poststratification adjustment, the age category of 50+ is further divided into 50-64 and 65+ categories.

4. *Selected Person Weight Adjustment for Poststratification to Roster Data.* This step utilizes control totals derived from the DU roster that are already poststratified to the census population estimates. This adjustment assists in bias reduction and improved precision by taking advantage of the properties of a two-phase design. Selected person sample weights (i.e., those that have been adjusted at the DU level and account for fourth-stage sampling) are adjusted to the DU weight sums of all eligible rostered persons. Any demographic information used in modeling is based solely on screener information because this is the only information available for all rostered persons. Potential explanatory variables for this adjustment are a combination of the variables presented in Table 3.8.
5. *Person-Level Nonresponse Adjustment.* This adjustment allows for the correction of weights resulting from the failure of selected sample persons to complete the interview. Respondent sample weights will be adjusted to the weight of all selected persons. Again, demographic information used in modeling is based solely on screener information. Potential explanatory variables for this adjustment are a combination of the variables presented in Table 3.8.

Table 3.8 Definitions of Levels for Potential Variables for Selected Person Poststratification and Person-Level Nonresponse Adjustment

Group Quarters Indicator
- 1: College Dorm
- 2: Other Group Quarter
- 3: Nongroup Quarter

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

Percentage of Blacks in Segment (% Black)
- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Percentage of Hispanics in Segment (% Hispanic)
- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Population Density
- 1: CBSA > 1,000,000
- 2: CBSA < 1,000,000
- 3: Non-CBSA Urban
- 4: Non-CBSA Rural

Quarter
- 1: Quarter 1
- 2: Quarter 2
- 3: Quarter 3
- 4: Quarter 4

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

State

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

Gender
- 1: Male
- 2: Female

(continued)

Table 3.8 Definitions of Levels for Potential Variables for Selected Person Poststratification and Nonresponse Adjustment (continued)

Hispanicity
1: Hispanic
2: Non-Hispanic
Race
1: White
2: Black
3: American Indian/Alaska Native
4: Asian
5: Two or More Races
Relation to Householder
1: Householder or Spouse
2: Child
3: Other Relative
4: Non-Relative

CBSA = core-based statistical area.
Note: Interactions among the main effect variables also are considered.

6. *Person-Level Poststratification.* This step is to adjust the final person sample weights to the census population estimates derived from the 2000 census. These are the same outside control totals used in the second adjustment. However, demographic variables for this adjustment are based on questionnaire data, not screener data as in adjustments 2, 4, and 5. Potential explanatory variables used in modeling are presented in Table 3.7.

7. *Extreme Weight Treatment at the Person Level.* This adjustment will be implemented in the same manner as described in adjustment 3, except that the weights reflect the fourth stage of selection.

All weight adjustments for the 2006 main study's final analysis weights will be derived from a generalized exponential model. To help reduce computational burden at all adjustment steps, separate models will be fit for clusters of States, based on census division definitions as shown in Table 3.9. Furthermore, model variable selection at each adjustment will be done using a combination method of forward and backward selection processes. The forward selection will be used for the model enlargement. Within each enlargement, backward selection will be used. The final adjusted weight, which is the product of weight components 1 through 15, is the analysis weight used in estimation. Exhibit 1 presents a flowchart of steps used in the weighting process, and Exhibit 2 displays all individual weight components.

Full details of the finalized modeling procedures, as well as final variables used in each adjustment step, will be described in the forthcoming report on the person-level sampling weight calibration for the 2006 NSDUH (Chen et al., 2007).

Table 3.9 Model Group Definitions

Model	Defined State
1	Connecticut, Maine, New Hampshire, Rhode Island, Vermont, Massachusetts
2	New Jersey, New York, Pennsylvania
3	Illinois, Indiana, Michigan, Wisconsin, Ohio
4	Iowa, Kansas, Minnesota, Missouri, Nebraska, South Dakota, North Dakota
5	Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia
6	Alabama, Kentucky, Mississippi, Tennessee
7	Arkansas, Louisiana, Oklahoma, Texas
8	Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Arizona
9	Alaska, Hawaii, Oregon, Washington, California

Exhibit 1. Flowchart of Sample Weighting Steps

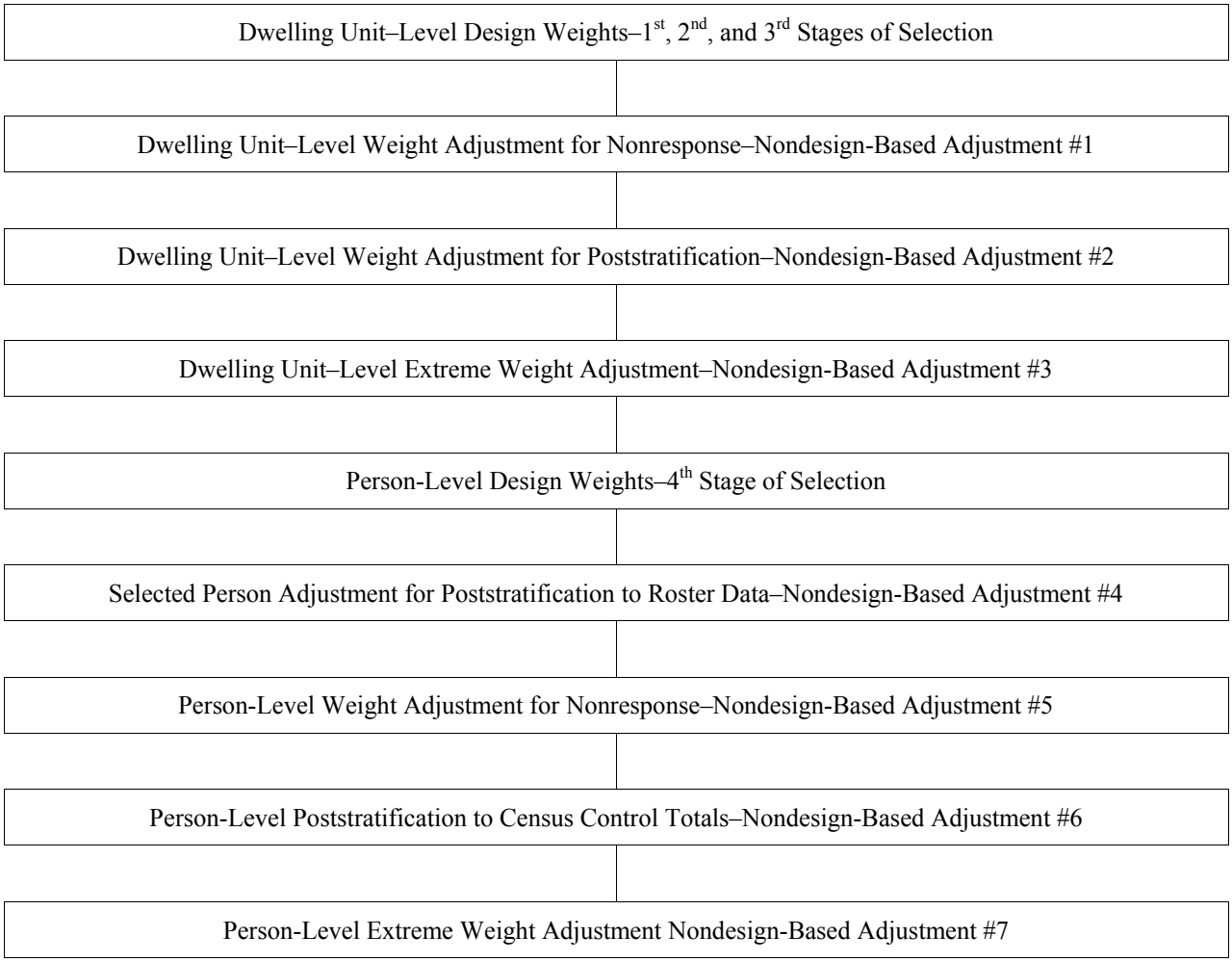

Exhibit 2. Sample Weight Components

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 Dwelling Unit
#6.	Inverse Probability of Added/Subsampled Dwelling Unit
#7.	Dwelling Unit Release Adjustment
#8.	Dwelling Unit Nonresponse Adjustment
#9.	Dwelling Unit Poststratification Adjustment
#10.	Dwelling Unit Extreme Weight Adjustment
Person-Level Design Weight Components	
#11.	Inverse Probability of Selecting a Person within a Dwelling Unit
#12.	Selected Person Poststratification to Roster Adjustment
#13.	Person-Level Nonresponse Adjustment
#14.	Person-Level Poststratification Adjustment
#15.	Person-Level Extreme Weight Adjustment

Chapter 4. General Sample Allocation Procedures for the Reliability Study and Income Split Sample

A Reliability Study and an income split sample were included in the 2006 National Survey on Drug Use and Health (NSDUH) to measure the reliability of NSDUH measures and the impact of a reduced set of income questions, respectively. The Reliability Study and income split sample originally were planned for all four quarters of the 2006 NSDUH. Due to a programming logic error, the Reliability Study was pulled from the quarter 1 sample. As a result, field staff were required to complete approximately 3,100 reinterviews in three quarters rather than four.

The income split sample was to have been embedded within the Reliability Study in all four quarters. Even though the income module treatments were assigned within the quarter 1 Reliability Study households, the treatments were maintained in the sample when the Reliability Study was pulled. Therefore, a larger number of new (reduced) income modules were completed than originally planned in quarter 1. In the other three quarters, when a household "screened out" of the Reliability Study, household respondents received the old (full) income module regardless of the initial assignment.

The following sections provide more detail on the sample allocation procedures used for the Reliability Study and the income split sample.

4.1 Respondent Universe and Sampling Methods for the Reliability Study

The Reliability Study was fielded during quarters 2 through 4 of the 2006 NSDUH. The Reliability Study sample was embedded within the main study sample. Similar to the main study, persons eligible for the Reliability Study were the civilian, noninstitutionalized population aged 12 years old or older. Unlike the main study, the respondent universe excluded residents of Alaska and Hawaii, residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories), and persons who do not speak English. Additionally, households in which two persons were selected were excluded from the study; that is, reinterviews were conducted in households that were designated for the Reliability Study where only one person was selected.

An equal probability sample of 400 State sampling (SS) regions was selected for the Reliability Study. Each quarter, a random sample of dwelling units (DUs) in the 800 Reliability Study segments was designated for the study.

The Reliability Study had two substudies:

- same interviewer substudy in which both the initial interview and the follow-up interview were conducted by the same interviewer; and

- paired interviewer substudy in which a follow-up interview was conducted by a different interviewer working in a nearby area.

The purpose of these substudies was to allow a comparison of reliability measures based on the same interviewer reinterview and a different interviewer reinterview. In quarter 2, one third of

the Reliability Study SS regions were designated for the same interviewer substudy, and the remaining two thirds were in the paired interviewer substudy. In quarter 3, a different one third of the regions was selected for the same interviewer substudy. Finally, in quarter 4, the remaining one third was assigned to the same interviewer substudy. Thus, each Reliability SS region was in the same interviewer substudy for one quarter and the paired interviewer substudy for the other two quarters.

4.2 Dwelling Unit and Person Sample Allocation Procedures for the Reliability Study

In preparation for the Reliability Study, we estimated that 26,098 selected DUs would be needed to yield 3,100 completed reinterviews for an effective sample size of approximately 1,800 (Table 4.1). This estimate assumed an 82 percent interview response rate (IRR) for initial interviews and a 92 percent IRR for reinterviews. Additionally, we assumed that 3 percent of respondents would be ineligible for reinterview due to having completed the first interview in Spanish. Finally, we assumed that 84 percent of the DUs would be eligible for the main study and 21 percent of those would result in a single person selection; thus, the expected eligibility rate for Reliability Study dwelling units was 18 percent.

Table 4.1 Reliability Study Design Parameters

Dwelling Unit Level: Total Sample	**Rate**	***N***						
State Sampling Regions		400						
Segments		3,200						
Selected Lines		26,098						
Expected Eligible Dwelling Units	0.18	4,698						
Expected Completed Screening Interviews	0.91	4,275						
	Overall		**Aged 12-17**		**Aged 18-25**		**Aged 26+**	
Person Level: Total Sample	**Rate**	***N***	**Rate**	***N***	**Rate**	***N***	**Rate**	***N***
Expected Selected Persons (1st Interview)	1.00	4,275	1.00	1,288	1.00	1,351	1.00	1,636
Expected Completed Interviews (1st Interview)	0.82	3,488	0.89	1,140	0.84	1,133	0.74	1,214
Expected Selected Persons (2nd Interview)	0.97	3,372	0.99	1,123	0.96	1,088	0.96	1,161
Expected Completed Interviews (2nd Interview)	0.92	3,100	0.92	1,033	0.95	1,033	0.89	1,033
Expected Sample Size Based on Assumed Design Effect	1.7	1,824	1.7	607.8	1.7	607.8	1.7	607.8

In the implementation, the 26,098 DUs were distributed among quarters 2 through 4 in proportions of 35, 35, and 30 percent, respectively. Each quarter, the Reliability Study sample was allocated to the 800 segments proportionally to the main study sample size (excluding sampled group quarters units).

To allow greater control over the Reliability Study sample, the DUs were drawn from specific main study partitions. In quarter 2, the Reliability Study sample was limited to release 1 (80/120 of main sample). In quarters 3 and 4, the Reliability Study sample was drawn from releases 1 (80/120 of main sample), 3 (5/120), and 6 (10/120). In all three quarters, partition 1 was

released. In quarters 3 and 4, consideration was given to the Reliability Study when releasing partitions 3 and 6; however, the main study sample was given priority. The Reliability Study sample resulted in a total of 3,136 reinterviews. For more information on the results of the Reliability Study, visit the Substance Abuse and Mental Health Services Administration (SAMHSA) website for the forthcoming 2006 NSDUH Questionnaire Reliability Study report (see http://oas.samhsa.gov/nsduh/methods.cfm).

4.3 Dwelling Unit Allocation Procedures for the Income Split Sample

A proportion of the 2006 NSDUH respondents received a new (reduced) set of income questions. In quarter 1, the income module was randomly assigned at the DU level. If two persons were selected in the household, both persons received the same income module treatment. A total of 2,050 quarter 1 interviews were completed with the new income module.

In quarters 2 through 4, the income split sample was embedded within the Reliability Study. Fifty percent of the households designated for the Reliability Study were assigned to the new income module. The remaining 50 percent received the old (full) set of income questions. Similar to the Reliability Study, only households with a single person selected were eligible to receive the new income module. If a household was flagged to receive the new income module and two persons were selected (i.e., the household "screened out" of the Reliability Study and new income module), the flag was set back to the default value of old (full) income module. A total of 1,797 new income module interviews were completed in quarters 2 through 4 for a total of 3,847 for the year.

References

Chen, P., Dai, L., Gordek, H., Laufenberg, J., Sathe, S. N., & Westlake, M. (2007). Person-level sampling weight calibration [2006]. In *2006 National Survey on Drug Use and Health: Methodological resource book* (Section 3, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009). Research Triangle Park, NC: RTI International. [To be available as a PDF at http://www.oas.samhsa.gov/nhsda/methods.cfm#2k6]

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

Chromy, J. R. (1981). Variance estimators for a sequential sample selection procedure. In D. Krewski, R. Platek, & J. N. K. Rao (Eds.), *Current topics in survey sampling: Proceedings of the symposium held May 7-9, 1980, sponsored by the Ottawa Chapter and the Survey Research Methods Section of the American Statistical Association* (pp. 329-347). New York: Academic Press.

Chromy, J. R. (1987). Design optimization with multiple objectives. In *Proceedings of the 1987 American Statistical Association, Survey Research Methods Section, San Francisco, CA* (pp. 194-199). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/SRMS/proceedings/]

Chromy, J. R., & Penne, M. (2002). Pair sampling in household surveys. In *Proceedings of the 2002 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, New York, NY [CD-ROM]* (pp. 552-554). Alexandria, VA: American Statistical Association.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley & Sons.

Kish, L. (1965). *Survey sampling*. New York: John Wiley & Sons.

Office of Management and Budget. (2003, June 6). *OMB Bulletin No. 03-04: Revised definitions of metropolitan statistical areas, new definitions of micropolitan statistical areas and combined statistical areas, and guidance on uses of the statistical definitions of these areas*. Retrieved June 29, 2005, from http://www.whitehouse.gov/omb/bulletins/b03-04.html

RTI International. (2004). *SUDAAN®, Release 9.0* [computer software]. Research Triangle Park, NC: Author.

RTI International. (2005). *2006 National Survey on Drug Use and Health: Counting and listing general manual* (RTI/0209009, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022). Research Triangle Park, NC: Author.