

2002 National Survey on Drug Use and Health

Sample Design Report

Contract No. 283-98-9008
RTI Project No. 7190
Phase IV, Deliverable No. 10

Authors:

Dawn M. Odom
Katherine R. Bowman
James R. Chromy
Peilan C. Martin

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:
RTI International

January 2004

2002 National Survey on Drug Use and Health

Sample Design Report

Authors:

Dawn M. Odom
Katherine R. Bowman
James R. Chromy
Peilan C. Martin

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Contract No. 283-98-9008
Deliverable No. 10

December 2003

Acknowledgments

This publication was developed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Office of Applied Studies (OAS), by RTI International, Research Triangle Park, North Carolina, under Contract No. 283-98-9008. Significant contributors at RTI include Katherine R. Bowman, James R. Chromy, Dawn M. Odom, Peilan C. Martin, R. K. Scott Chestnut, and Thomas G. Virag (Project Director).

Table of Contents

| Chapter | Page |
|---|------------|
| List of Tables..... | iv |
| Chapter 1: Overview | 1 |
| 1.1..... Target Population | 1 |
| 1.2..... Design Overview..... | 1 |
| 1.3..... 5-Year Design | 2 |
| 1.4..... Stratification and First-Stage Sample Selection..... | 2 |
| 1.5..... Dwelling Units and Persons | 3 |
| Chapter 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-Stage Sample Selection | 10 |
| 2.4..... Survey Year and Quarter Assignment..... | 11 |
| 2.5..... Creation of Variance Estimation Strata..... | 12 |
| Chapter 3: General Sample Allocation Procedures for the Main Study | 15 |
| 3.1..... Notation..... | 16 |
| 3.2..... Determining Person Sample Sizes by State and Age Group..... | 18 |
| 3.3..... Second-Stage Sample Allocation for Each Segment | 21 |
| 3.3.1 Dwelling Unit Frame Construction—Counting and Listing..... | 21 |
| 3.3.2 Determining Dwelling Unit Sample Size..... | 26 |
| 3.4..... Determining Third-Stage Sample (Person) Selection Probabilities for Each Segment..... | 26 |
| 3.5..... Sample Size Constraints: Guaranteeing Sufficient Sample for Additional Studies and Reducing Field Interviewer Burden | 27 |
| 3.6..... Dwelling Unit Selection and Release Partitioning..... | 28 |
| 3.7..... Half-Open Interval Rule and Procedure for Adding Dwelling Units..... | 28 |
| 3.8..... Quarter-by-Quarter Deviations | 29 |
| 3.9..... Sample Weighting Procedures | 34 |
| References | 43 |
| Appendix A 1999-2003 NHSDA/NSDUH Field Interview Regions..... | A-1 |
| Appendix B 2002 NSDUH Procedure for Subsegmenting..... | B-1 |
| Appendix C 2002 NSDUH Procedure for Adding Missed Dwelling Units..... | C-1 |

List of Tables

| | | |
|------------|---|----|
| Table 2.1 | Number of Segments on Sampling Frame, by State | 9 |
| Table 2.2 | Survey Year and Quarter Assignment Order for 96 Segments within Each FI Region..... | 11 |
| Table 2.3 | Segment Identification Number Suffixes for the 1999-2003 NSDUHs..... | 13 |
| Table 3.1 | Expected Relative Standard Errors by Race/Ethnicity and Age Group: Main Sample..... | 20 |
| Table 3.2 | Expected Main Study Sample Sizes by State and Age Group | 22 |
| Table 3.3 | Number of Map Pages by State and Segment | 24 |
| Table 3.4 | Segment and Dwelling Unit Summary..... | 25 |
| Table 3.5 | Quarterly Sample Sizes and Percentages Released..... | 29 |
| Table 3.6 | Definitions of Levels for Potential Variables for Dwelling Unit Nonresponse Adjustment | 36 |
| Table 3.7 | Definitions of Levels for Potential Variables for Dwelling Unit Poststratification and Respondent Poststratification at the Person Level | 37 |
| Table 3.8 | Definitions of Levels for Potential Variables for Selected Person Poststratification and Nonresponse Adjustment | 38 |
| Table 3.9 | Model Group Definitions | 39 |
| Table 3.10 | Flowchart of Sample Weighting Steps..... | 40 |
| Table 3.11 | Sample Weight Components..... | 41 |

Chapter 1: Overview

1.1 Target Population

The respondent universe for the 2002 National Survey on Drug Use and Health¹ (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 2002 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 2002 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

The Substance Abuse and Mental Health Services Administration (SAMHSA) implemented major changes in the way the NSDUH would be conducted, beginning in 1999 and continuing through subsequent years. 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: 12 to 17 year olds, 18 to 25 year olds, and 26 year olds 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 2002 CAI sample was 68,126 persons.

Included in the 2001 survey was an experimental study conducted to evaluate the effectiveness of respondent incentives on improving response rates and to examine the results of incentives on data quality, survey costs, and substance use estimates. The study compared the effectiveness of \$40, \$20, and \$0 incentive payments. The results of the experiment showed that the \$20 and \$40 treatments produced significantly better interview response rates than the control group. Based on the results of this experiment, SAMHSA chose to institute a \$30 incentive payment beginning with the 2002 NSDUH. As expected, response rates improved significantly in the 2002 survey. Due to these higher response rates, fewer selected households were required in 2002 than in previous surveys.

An additional change occurred in the 2002 NSDUH. Due to concerns regarding the sample size of pair-level data, more pairs were selected in 2002. The new pair sampling strategy increased the number of pairs selected in dwelling units with older persons rostered. Due to the high person selection probabilities for younger age persons, the previous algorithm already selected pairs in almost all dwelling units with more than one person aged 25 or younger. Since

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

older persons, particularly those 50 or older, have low selection probabilities the previous algorithm often selected only one person in dwelling units with mostly older persons. The adjustment in the sample selection algorithm achieved the same person selection probabilities in such older person dwelling units by more frequently selecting either none or two persons per dwelling unit rather than selecting only one from two or more eligible persons. The overall impact was fewer dwelling units with at least one person selected, and an increase in the number of dwelling units with two persons selected. Chromy and Penne (2002) reported on the impact of the new algorithm on response rates. Based on simulation with prior years' data, it was projected that a moderate decrease in the response rate for older persons would occur. An intermediate version of the algorithm was adopted to achieve a balance between increasing the number of pairs and decreasing response rates for older persons.

1.3 5-Year Design

A coordinated 5-year sample design was developed. The 2002 main sample is a subsample of the 5-year sample. Although there is no planned overlap with the 1998 sample, a coordinated design for 1999-2003 facilitated 50% overlap in first-stage units (area segments) within each successive 2-year period from 1999 through 2003. 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 1999-2003 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 2002 survey. This sample size was considered adequate to support direct State estimates. The remaining 43 States³ had samples designed to yield 900 respondents per State in the 2002 survey. In these 43 States, adequate data were available to support reliable State estimates based on small area estimation methodology.

1.4 Stratification and First-Stage Sample Selection

Within each State, field interviewer (FI) regions were formed. Based on a composite size measure, States were geographically partitioned into roughly equally size 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, thus distributing the workload equally among NSDUH interviewers. The smaller States were partitioned into 12 FI regions, whereas the eight "large" States were divided into 48 regions. Therefore, the partitioning of the United States resulted in the formation of a total of 900 FI regions. FI region maps can be found in Appendix A.

² For the 1999-2003 NSDUHs, 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.

For the first stage of sampling, each of the FI regions was partitioned into noncompact clusters⁴ of dwelling units 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 dwelling unit in the NSDUH refers to either a housing unit or a group-quarters listing unit, such as a dormitory room or a shelter bed. To support the overlapping sample design and any special supplemental samples or field tests that SAMHSA may wish to conduct, segments were formed to contain a minimum of 175 dwelling units⁵ on average. In prior years, this average minimum segment dwelling unit size was only 90.

Before selecting sample segments, additional implicit stratification was achieved by sorting the first-stage sampling units by an MSA/SES (metropolitan statistical area/socioeconomic status) indicator⁶ and by the percentage of the population that is non-Hispanic and white. From this well-ordered sample frame, 96 segments⁷ per FI region were selected with probabilities proportionate to a composite size measure and with minimum replacement (Chromy, 1979). The selected segments were then randomly assigned to a survey year and quarter of data collections, as described in Section 2.4. Twenty-four of these segments were designated for the coordinated 5-year sample, while the other 72 were designated as "reserve" segments.

1.5 Dwelling Units and Persons

After sample segments for the 2002 NSDUH were selected, specially trained field household listers visited the areas and obtained complete and accurate lists of all eligible dwelling units within the sample segment boundaries. These lists served as the frames for the second stage of sample selection.

The primary objective of the second stage of sample selection (listing units) was to determine the minimum number of dwelling units 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 1990 Census data adjusted to more recent data from Claritas, State- and age-specific sampling rates were computed. These rates were then adjusted by the segment's probability of

⁴ Noncompact clusters (selection from a list) differ from compact clusters in that not all units within the cluster are included in the sample. While 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).

⁵ Dwelling unit counts were obtained from the 1990 Decennial Census data supplemented with revised population counts from Claritas, a market research firm headquartered in San Diego, California (<http://cluster1.claritas.com/claritas/Default.jsp>).

⁶ Four categories are defined as (1) MSA/low SES, (2) MSA/high SES, (3) Non-MSA/low SES, and (4) Non-MSA/high SES. In order to define SES, block group-level median rents and property values were given a rank (1...5) based on State and MSA quintiles. The rent and value ranks were then averaged, weighting by the percentages renter- and owner-occupied dwelling units, respectively. If the resulting score fell in the lower 25th percentile by State and MSA, the area was considered "low SES"; otherwise, it was considered "high SES."

⁷ The 1999-2003 sample was planned so that 48 segments per FI region would be selected. In the implementation, however, an additional 48 segments were added to support any supplemental or field test samples.

selection; the subsegmentation inflation factor,⁸ 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.⁹ In addition to these factors, historical data from the 2000, 2001, and 2002 NSDUHs were used to compute predicted screening and interviewing response rate adjustments. The final adjusted sampling rate was then multiplied by the actual number of dwelling units 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 the overlapping design and/or 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 dwelling unit selections were made, an interviewer visited each selected dwelling unit to obtain a roster of all persons residing in the dwelling unit. As in previous years, during the data collection period, if an interviewer encountered any new dwelling unit in a segment or found a dwelling unit that was missed during the original counting and listing activities, then the new or missed dwellings were selected into the 2002 NSDUH using the half-open interval selection technique.¹⁰ The selection technique eliminates any frame bias that might be introduced because of errors and/or omissions in the counting and listing activities, and also eliminates any bias that might be associated with using "old" segment listings.

Using the roster information obtained from an eligible member of the selected dwelling unit, 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 third stage of selection based on the State and age group sampling parameters.

One exciting consequence of using an electronic screening instrument in the NSDUH is the ability to impose a more complicated person-level selection algorithm on the third stage of the NSDUH design. In 1999 and continuing through 2002, one feature that was included in the design was that any two survey-eligible people within a dwelling unit had some chance of being selected (i.e., all survey-eligible pairs of people had some nonzero chance of being selected). This design feature was of interest to NSDUH researchers because, for example, it allows

⁸ Segments found to be very large in the field are partitioned into *subsegments*. Then, one subsegment is chosen at random with probability proportional to size to be fielded. The subsegmentation inflation factor accounts for the narrowing down of the segment.

⁹ 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.

¹⁰ In summary, this technique States that, if a dwelling unit is selected for the 2002 study and an interviewer observes any new or missed dwelling units between the selected dwelling unit and the dwelling unit appearing immediately after the selection on the counting and listing form, then all new or missed dwellings falling in this interval will be selected. If a large number of new or missed dwelling units are encountered (generally greater than ten), then a sample of the new or missing dwelling units will be selected. For more information, please refer to Appendix C.

analysts to examine how the drug use propensity of one individual in a family relates to the drug use propensity of other family members residing in the same dwelling unit (e.g., the relationship of drug use between a parent and his or her child).

Chapter 2: The Coordinated 5-Year Sample

As was previously mentioned, the sample design was simultaneously developed for the 1999-2003 NSDUHs. Starting with a Census block-level frame, first-stage sampling units or area segments were formed. A sufficient number of segments was then selected to support the 5-year design as well as any supplemental studies SAMHSA may choose 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 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

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

- Constrain the maximum number of selections per dwelling unit at a specified value; specifically, we limited the total number of within-dwelling unit selections across all age groups to a maximum of two.
- Minimize the number of sample dwelling units 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-dwelling unit/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 dwelling units within a specific stratum were selected with approximately the same probability, and therefore, approximately equalized dwelling unit sampling weights. This feature minimizes 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 decennial Census data (block-level counts) are adequate to implement the method.

Using the 1990 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 ($d=1, \dots, 5$ 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 FI region i within each State h . The composite size measure for block k was defined as

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

The composite size measure for segment j was calculated as

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

where N_{hij} equals the number of blocks within segment j of FI 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 FIs. Within each State, Census tracts were joined to form mutually exclusive and exhaustive FI 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 NSDUH interviewers. Twelve FI regions were formed in each State, except in California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas, where 48 regions were formed.¹¹

¹¹ 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 with 2 area segments each, per quarter. Two additional segments were selected for each of the other 3 quarters, yielding 8 area segments per stratum, or 96 area segments per small sample State. This stratum configuration also corresponded with a reasonable average workload for a single FI, leading us to designate the geographic strata within States as FI regions. 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, four 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 FI 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.

To form segments within FI regions, adjacent Census blocks were collapsed until the total number of dwelling units within the area was at least 175 and the size measure was at least 9.38 times the maximum of F_1 , F_2 , F_3 , F_4 , and F_5 , where F_i is the person-level sampling rate for age group i in the State. The desired number of responding persons in each segment is 9.38. Latitude and longitude and sorting within block groups, tracts, and counties were used to obtain geographic ordering of the blocks. Segments were required to be entirely within FI region and county boundaries; however, they could span Census tracts and block groups. This crossing-over was avoided as much as possible. Table 2.1 summarizes the segment sampling frame by State.

Table 2.1 Number of Segments on Sampling Frame, by State

| State | State Abbreviation | State FIPS Code | Number of Segments on Sampling Frame | Total Number of Segments Selected | Number Selected for 5-Year Sample | Unique Segments in 5-Year Sample |
|-------------------|--------------------|-----------------|--------------------------------------|-----------------------------------|-----------------------------------|----------------------------------|
| Total U.S. | | | 499,287 | 86,400 | | |
| Northeast | | | | | | |
| Connecticut | CT | 09 | 5,978 | 1,152 | 288 | 288 |
| Maine | ME | 23 | 2,573 | 1,152 | 288 | 288 |
| Massachusetts | MA | 25 | 11,413 | 1,152 | 288 | 288 |
| New Hampshire | NH | 33 | 2,246 | 1,152 | 288 | 286 |
| New Jersey | NJ | 34 | 14,343 | 1,152 | 288 | 288 |
| New York | NY | 36 | 30,600 | 4,608 | 1,152 | 1,151 |
| Pennsylvania | PA | 42 | 24,256 | 4,608 | 1,152 | 1,151 |
| Rhode Island | RI | 44 | 1,912 | 1,152 | 288 | 282 |
| Vermont | VT | 50 | 1,248 | 1,152 | 288 | 284 |
| Midwest | | | | | | |
| Illinois | IL | 17 | 22,549 | 4,608 | 1,152 | 1,151 |
| Indiana | IN | 18 | 11,987 | 1,152 | 288 | 288 |
| Iowa | IA | 19 | 6,210 | 1,152 | 288 | 288 |
| Kansas | KS | 20 | 5,430 | 1,152 | 288 | 288 |
| Michigan | MI | 26 | 18,477 | 4,608 | 1,152 | 1,152 |
| Minnesota | MN | 27 | 9,364 | 1,152 | 288 | 288 |
| Missouri | MO | 29 | 10,871 | 1,152 | 288 | 288 |
| Nebraska | NE | 31 | 3,567 | 1,152 | 288 | 288 |
| North Dakota | ND | 38 | 1,330 | 1,152 | 288 | 286 |
| Ohio | OH | 39 | 21,500 | 4,608 | 1,152 | 1,151 |
| South Dakota | SD | 46 | 1,603 | 1,152 | 288 | 285 |
| Wisconsin | WI | 55 | 10,704 | 1,152 | 288 | 288 |

(continued)

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

| State | State Abbreviation | State FIPS Code | Number of Segments on Sampling Frame | Total Number of Segments Selected | Number Selected for 5-Year Sample | Unique Segments in 5-Year Sample |
|------------------|--------------------|-----------------|--------------------------------------|-----------------------------------|-----------------------------------|----------------------------------|
| South | | | | | | |
| Alabama | AL | 01 | 8,702 | 1,152 | 288 | 288 |
| Arkansas | AR | 05 | 5,411 | 1,152 | 288 | 288 |
| Delaware | DE | 10 | 1,346 | 1,152 | 288 | 281 |
| Washington, D.C. | DC | 11 | 943 | 1,152 | 288 | 273 |
| Florida | FL | 12 | 26,545 | 4,608 | 1,152 | 1,152 |
| Georgia | GA | 13 | 13,398 | 1,152 | 288 | 288 |
| Kentucky | KY | 21 | 7,718 | 1,152 | 288 | 287 |
| Louisiana | LA | 22 | 8,216 | 1,152 | 288 | 288 |
| Maryland | MD | 24 | 8,340 | 1,152 | 288 | 288 |
| Mississippi | MS | 28 | 5,473 | 1,152 | 288 | 288 |
| North Carolina | NC | 37 | 14,955 | 1,152 | 288 | 288 |
| Oklahoma | OK | 40 | 6,941 | 1,152 | 288 | 288 |
| South Carolina | SC | 45 | 7,437 | 1,152 | 288 | 287 |
| Tennessee | TN | 47 | 10,764 | 1,152 | 288 | 288 |
| Texas | TX | 48 | 34,367 | 4,608 | 1,152 | 1,151 |
| Virginia | VA | 51 | 11,666 | 1,152 | 288 | 288 |
| West Virginia | WV | 54 | 3,757 | 1,152 | 288 | 288 |
| West | | | | | | |
| Alaska | AK | 02 | 1,139 | 1,152 | 288 | 273 |
| Arizona | AZ | 04 | 8,212 | 1,152 | 288 | 288 |
| California | CA | 06 | 53,064 | 4,608 | 1,152 | 1,152 |
| Colorado | CO | 08 | 7,977 | 1,152 | 288 | 287 |
| Hawaii | HI | 15 | 1,658 | 1,152 | 288 | 276 |
| Idaho | ID | 16 | 2,611 | 1,152 | 288 | 288 |
| Montana | MT | 30 | 2,028 | 1,152 | 288 | 286 |
| Nevada | NV | 32 | 2,625 | 1,152 | 288 | 276 |
| New Mexico | NM | 35 | 3,369 | 1,152 | 288 | 288 |
| Oregon | OR | 41 | 6,835 | 1,152 | 288 | 288 |
| Utah | UT | 49 | 3,475 | 1,152 | 288 | 288 |
| Washington | WA | 53 | 11,086 | 1,152 | 288 | 287 |
| Wyoming | WY | 56 | 1,068 | 1,152 | 288 | 285 |

FIPS = Federal Information Processing Standards.

2.3 First-Stage Sample Selection

Once the segments were formed, a probability proportional to size sample of segments was selected with minimum replacement within each FI region. The sampling frame was implicitly stratified by sorting the first-stage sampling units by an MSA/SES indicator and by the percentage of the population that is non-Hispanic and white. As Table 2.1 indicates, 96 segments per FI region were chosen for a total of 1,152 segments in each State, except in the large States where a total of 4,608 segments were chosen. Although only 24 segments per FI region were needed to support the 5-year study, an additional 72 segments were selected to serve as

replacements when segment lines are depleted and/or to support any supplemental studies embedded within the NSDUH.

2.4 Survey Year and Quarter Assignment

Within each FI region, the 96 selected segments were assigned to a survey year and quarter in a random, systematic fashion. Because segments can be selected multiple times, the goal was to avoid putting the same segment in consecutive survey years. Therefore, survey years and quarters were assigned using a random starting point and the order defined in Table 2.2. The notation in the table is as follows:

- 99A = Segment for the 1999 NHSDA,
- 99B = Segment for the 1999 NHSDA and used again in the 2000 NHSDA,
- 00 = Segment for the 2000 NHSDA and used again in the 2001 NHSDA,
- 01 = Segment for the 2001 NHSDA and used again in the 2002 NSDUH,
- 02 = Segment for the 2002 NSDUH and used again in the 2003 NSDUH, and
- 03 = Segment for the 2003 NSDUH.

Table 2.2 Survey Year and Quarter Assignment Order for 96 Segments within Each FI Region

| Order | Survey Year | Quarter | Panel | Variance Replicate | Order | Survey Year | Quarter | Panel | Variance Replicate |
|-------|-------------|---------|-------|--------------------|-------|-------------|---------|-------|--------------------|
| 1 | 99A | 1 | 1 | 1 | 25 | 99A | 2 | 1 | 1 |
| 2 | Y00 | 1 | 15 | 1 | 26 | Y00 | 2 | 15 | 1 |
| 3 | X99B | 1 | 8 | 2 | 27 | X99B | 2 | 8 | 2 |
| 4 | Z01 | 1 | 22 | 2 | 28 | Z01 | 2 | 22 | 2 |
| 5 | 02 | 1 | 5 | 1 | 29 | 02 | 2 | 5 | 1 |
| 6 | Y99A | 1 | 13 | 1 | 30 | Y99A | 2 | 13 | 1 |
| 7 | X03 | 1 | 12 | 2 | 31 | X03 | 2 | 12 | 2 |
| 8 | Z99B | 1 | 20 | 2 | 32 | Z99B | 2 | 20 | 2 |
| 9 | 00 | 1 | 3 | 1 | 33 | 00 | 2 | 3 | 1 |
| 10 | Y02 | 1 | 17 | 1 | 34 | Y02 | 2 | 17 | 1 |
| 11 | X01 | 1 | 10 | 2 | 35 | X01 | 2 | 10 | 2 |
| 12 | Z03 | 1 | 24 | 2 | 36 | Z03 | 2 | 24 | 2 |
| 13 | 01 | 1 | 4 | 2 | 37 | 01 | 2 | 4 | 2 |
| 14 | Y03 | 1 | 18 | 2 | 38 | Y03 | 2 | 18 | 2 |
| 15 | X02 | 1 | 11 | 1 | 39 | X02 | 2 | 11 | 1 |
| 16 | Z99A | 1 | 19 | 1 | 40 | Z99A | 2 | 19 | 1 |
| 17 | 99B | 1 | 2 | 2 | 41 | 99B | 2 | 2 | 2 |
| 18 | Y01 | 1 | 16 | 2 | 42 | Y01 | 2 | 16 | 2 |
| 19 | X00 | 1 | 9 | 1 | 43 | X00 | 2 | 9 | 1 |
| 20 | Z02 | 1 | 23 | 1 | 44 | Z02 | 2 | 23 | 1 |
| 21 | 03 | 1 | 6 | 2 | 45 | 03 | 2 | 6 | 2 |
| 22 | Y99B | 1 | 14 | 2 | 46 | Y99B | 2 | 14 | 2 |
| 23 | X99A | 1 | 7 | 1 | 47 | X99A | 2 | 7 | 1 |
| 24 | Z00 | 1 | 21 | 1 | 48 | Z00 | 2 | 21 | 1 |

(continued)

Table 2.2 Survey Year and Quarter Assignment Order for 96 Segments within Each FI Region (continued)

| Order | Survey Year | Quarter | Panel | Variance Replicate | Order | Survey Year | Quarter | Panel | Variance Replicate |
|-------|-------------|---------|-------|--------------------|-------|-------------|---------|-------|--------------------|
| 49 | 99A | 3 | 1 | 1 | 73 | 99A | 4 | 1 | 1 |
| 50 | Y00 | 3 | 15 | 1 | 74 | Y00 | 4 | 15 | 1 |
| 51 | X99B | 3 | 8 | 2 | 75 | X99B | 4 | 8 | 2 |
| 52 | Z01 | 3 | 22 | 2 | 76 | Z01 | 4 | 22 | 2 |
| 53 | 02 | 3 | 5 | 1 | 77 | 02 | 4 | 5 | 1 |
| 54 | Y99A | 3 | 13 | 1 | 78 | Y99A | 4 | 13 | 1 |
| 55 | X03 | 3 | 12 | 2 | 79 | X03 | 4 | 12 | 2 |
| 56 | Z99B | 3 | 20 | 2 | 80 | Z99B | 4 | 20 | 2 |
| 57 | 00 | 3 | 3 | 1 | 81 | 00 | 4 | 3 | 1 |
| 58 | Y02 | 3 | 17 | 1 | 82 | Y02 | 4 | 17 | 1 |
| 59 | X01 | 3 | 10 | 2 | 83 | X01 | 4 | 10 | 2 |
| 60 | Z03 | 3 | 24 | 2 | 84 | Z03 | 4 | 24 | 2 |
| 61 | 01 | 3 | 4 | 2 | 85 | 01 | 4 | 4 | 2 |
| 62 | Y03 | 3 | 18 | 2 | 86 | Y03 | 4 | 18 | 2 |
| 63 | X02 | 3 | 11 | 1 | 87 | X02 | 4 | 11 | 1 |
| 64 | Z99A | 3 | 19 | 1 | 88 | Z99A | 4 | 19 | 1 |
| 65 | 99B | 3 | 2 | 2 | 89 | 99B | 4 | 2 | 2 |
| 66 | Y01 | 3 | 16 | 2 | 90 | Y01 | 4 | 16 | 2 |
| 67 | X00 | 3 | 9 | 1 | 91 | X00 | 4 | 9 | 1 |
| 68 | Z02 | 3 | 23 | 1 | 92 | Z02 | 4 | 23 | 1 |
| 69 | 03 | 3 | 6 | 2 | 93 | 03 | 4 | 6 | 2 |
| 70 | Y99B | 3 | 14 | 2 | 94 | Y99B | 4 | 14 | 2 |
| 71 | X99A | 3 | 7 | 1 | 95 | X99A | 4 | 7 | 1 |
| 72 | Z00 | 3 | 21 | 1 | 96 | Z00 | 4 | 21 | 1 |

X, Y, and Z denote extra segments for the corresponding NSDUH survey year. The 24 segments assigned to survey years not beginning with X, Y, or Z would then be used to field the 5-year study. Using the survey year and quarter assignments, a sequential segment identification number (SEGID) was then assigned. Table 2.3 describes the relationship between segment identification numbers and quarter assignment. The last two digits in the SEGID are called the "segment suffix" in Table 2.3. In Table 2.2, "panel" refers to a group of four segments (one per quarter) in an FI region that are either dropped or carried over to the following survey year. The 5-year survey consists of panels 1 through 6, which correspond 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 1999-2003 NSDUHs, each FI region comprised its own stratum.

Two replicates per year were defined within each variance stratum. The first replicate consists of those segments that are "phasing out" or will not be used in the next survey year. The second replicate is made up of those segments that are "phasing in" or will be fielded again the following year, thus constituting the 50 percent overlap between survey years. Each variance replicate consists of four segments, one for each quarter of data collection. 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 by using clustered data analysis software packages such as SUDAAN (RTI, 2001). The SUDAAN software package computes variance estimates for nonlinear statistics using procedures such 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 two or more years.

Table 2.3 Segment Identification Number Suffixes for the 1999-2003 NSDUHs

| Segment Suffix | 1999 NHSDA | 2000 NHSDA | 2001 NHSDA | 2002 NSDUH | 2003 NSDUH |
|-----------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 01 | x (Q1) | | | | |
| 02 | x (Q1) | x (Q1) | | | |
| 03 | x (Q2) | | | | |
| 04 | x (Q2) | x (Q2) | | | |
| 05 | x (Q3) | | | | |
| 06 | x (Q3) | x (Q3) | | | |
| 07 | x (Q4) | | | | |
| 08 | x (Q4) | x (Q4) | | | |
| 09 | | x (Q1) | x (Q1) | | |
| 10 | | x (Q2) | x (Q2) | | |
| 11 | | x (Q3) | x (Q3) | | |
| 12 | | x (Q4) | x (Q4) | | |
| 13 | | | x (Q1) | x (Q1) | |
| 14 | | | x (Q2) | x (Q2) | |
| 15 | | | x (Q3) | x (Q3) | |
| 16 | | | x (Q4) | x (Q4) | |
| 17 | | | | x (Q1) | x (Q1) |
| 18 | | | | x (Q2) | x (Q2) |
| 19 | | | | x (Q3) | x (Q3) |
| 20 | | | | x (Q4) | x (Q4) |
| 21 | | | | | x (Q1) |
| 22 | | | | | x (Q2) |
| 23 | | | | | x (Q3) |
| 24 | | | | | x (Q4) |

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

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 sample sizes will be discussed. The within-dwelling unit age group-specific selection probabilities for the 2002 NSDUH main study design are also addressed. This optimization procedure was specifically designed to address 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 dwelling units 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 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 utilizing 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., second-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 dwelling units would be selected. This 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 entails 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 dwelling unit.¹² A detailed discussion of the final step is given in Section 3.4. After calculating the required dwelling units and the selection probabilities, we applied sample size constraints¹³ to ensure adequate samples for overlapping designs and/or supplemental studies and to reduce the field interviewer burden. Limits on the total number of expected interviews per segment were also 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.

¹² Direct application of Brewer's method would require a fixed sample size.

¹³ Because of the overlap of the split sample, constraints were applied to the required dwelling unit sample sizes. Specifically, some segments would be revisited in the 2003 survey.

3.1 Notation

h = 50 States plus the District of Columbia.

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 quarter computation of selected dwelling unit 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 2002 population is estimated using the 1990 Census data adjusted to the 2001 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 2001, A = the population in 1990 and $x = 11$. Then, the 2002 population (y_{ha}^*) is computed using the original formula and this time allowing x to be 12. Finally, the 2002 population is adjusted by the ratio of estimated eligible listed dwelling units to the Claritas dwelling unit counts (U_{hj}). This adjustment factor considers the number of added dwelling units expected to be obtained through the half-open interval rule (1.01) and the probability of a dwelling unit being eligible (ϵ_s), both determined via historic data. The coefficient adjustment of 1.01 is estimated using historical data and is the proportion of all screened dwelling units (includes added) over the original total of selected dwelling units (excluding added dwelling units). So, $y_{ha} = \{[1.01 * \epsilon_s * L_{hj} * (1/I_{hj}) / U_{hj}]\} * y_{ha}^*$, where ϵ_s , L_{hj} , and I_{hj} are defined further below. This adjustment is computed at the Census block level, then aggregated to the State level.

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

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

P_{hj} = Inverse of the segment selection probability. Dwelling unit sample sizes are computed on a quarterly basis and segments are selected on a yearly basis. Since each quarter only contains 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 list efficiently in both time and cost, field listing personnel are allowed to subsegment the segment into roughly equal size 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 dwelling units in each subsegment: B_{hj}^*). Using a selection algorithm provided by RTI, 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 dwelling units to select for screening in segment j to meet the targeted sample sizes for all age groups.

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

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

ε_h = State-specific dwelling unit eligibility rate. Derived from 2000 NHSDA Quarter 4 and 2001 NHSDA Quarters 1 through 3 data by taking the average eligibility rate within each State.

ϕ_h = State-specific screening response rates. Calculated using the same methodology as described for the dwelling unit eligibility rate (ε_h).

λ_{ha} = State- and age group-specific interview response rate. Using data from Quarter 4 of the 2000 NHSDA and Quarters 1 through 3 of the 2001 NHSDA, the additive effects of State and age group on interview response were determined by taking the average interview response rate within each State. In addition, two adjustments were applied to the interview response rates to account for (1) the decreased rates in the older age groups due to the selection of additional pairs, and (2) the increased rates for all age groups due to the implementation of respondent incentives.

γ_{ha} = Expected number of persons within an age group per dwelling unit. Calculated using 2000 NHSDA Quarter 1 and 2001 NHSDA 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 dwelling units by State.

δ_{ha} = State- and age group-specific maximum-of-two rule adjustment. The survey design restricts the number of interviews per dwelling unit to a total of two. This is achieved through a modified Brewer's Method of Selection. This results in a loss of potential interviews in dwelling units where selection probabilities sum greater than two. The

adjustment is designed to inflate the number of required dwelling units to compensate for this loss. Using data from Quarter 4 of the 2000 NHSDA and Quarters 1 through 3 of the 2001 NHSDA, 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 third stage of selection was to determine the optimal number of respondents for each of the 255 domains that would be needed to minimize costs associated with data collection, subject to multiple precision requirements established by SAMHSA. In summary, the precision requirements on the relative standard error (RSE) of an estimate of 10 percent for SAMHSA's 17 subpopulations of interest are:

- RSE = 3.00 percent for the total national population.
- RSE = 5.00 percent for the national population in each of the four age groups: 12 to 17, 18 to 25, 26 to 34, 35 or older.
- RSE = 5.00 percent for the population within each of the four age groups for whites (i.e., nonblack, non-Hispanic).
- RSE = 11.00 percent for the population within each of the four age groups for blacks (i.e., black, non-Hispanic).
- RSE = 11.00 percent for the population within each of the four age groups for Hispanics.

One stratification feature we used in previous NSDUH designs that was worth including in the design of the current NSDUH is the expansion of the age group domain to 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older age groups. This age group stratification parallels SAMHSA's NSDUH subpopulation of interest, as implied by the precision constraints, except for the age group 35 or older. As we have done with the survey designs since 1992, we have chosen to further stratify this important age group by 35 to 49 and 50 or older to decrease the total number of 35 or older respondents needed to meet precision requirements. Since substance abuse is more prevalent among the 35 to 49 year olds compared to the 50 or older age group, oversampling this younger age group will increase the precision of the estimates generated for the 35 or older age group, while minimizing the total number of respondents aged 35 years or older needed in the sample.

To form precision constraints that reflect the above standard error requirements, we have set up a preliminary Step-1 Optimization using (1) design effects estimated from the 1994-1996 NHSDA data, (2) population counts obtained from Claritas, Inc., and (3) various outcome measures that were estimated for each block group in the United States from our 1991-1993 NHSDA small area estimation (SAE) project. Appropriate variance constraints were defined for nine outcome measures of interest. These 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 dependency on alcohol and illicit drug use.

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

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. *Any Illicit Drug Use in the Past Month.* Includes hallucinogens, heroin, marijuana, cocaine, inhalants, opiates, or nonmedical use of sedatives, tranquilizers, stimulants, or analgesics.
4. *Any Illicit Drug Use Other than Marijuana in the Past Month.* Past month use of any illicit drug 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 any illicit drug provides a broad measure of illicit drug use; however, it is dominated by marijuana and cocaine use. Therefore, estimates of marijuana and cocaine are included since 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, *Any 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 any illicit drug.

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 any illicit drugs.
9. *Received Treatment for Alcohol Use but Not 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 drinking. These estimates exclude those who received treatment in the past 12 months for both drinking and illicit drugs.

These outcome measures considered, as well as the precision that is expected from this 2002 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 Race/Ethnicity and Age Group: Main Sample

| Outcome Measure | Total Respondents | | | | | Hispanic Respondents | | | | |
|---|-------------------|-------|-------|-------|-------|----------------------|-------|-------|-------|-------|
| | 12-17 | 18-25 | 26-34 | 35+ | Total | 12-17 | 18-25 | 26-34 | 35+ | Total |
| Expected Relative Standard Error for Classes of Outcome Measures | | | | | | | | | | |
| Past Year, Dependence on Alcohol (not Illicit Drugs) | 2.62 | 2.70 | 5.15 | 3.23 | 2.31 | 6.49 | 7.54 | 12.86 | 10.56 | 6.15 |
| Past Month Alcohol Use | 2.71 | 2.71 | 5.08 | 3.25 | 2.52 | 6.77 | 7.47 | 12.74 | 10.33 | 6.54 |
| Past Month Cigarette Use | 2.43 | 2.62 | 4.96 | 2.99 | 2.26 | 7.29 | 7.11 | 12.37 | 10.87 | 7.03 |
| Past Month Cocaine Use | 2.41 | 2.50 | 4.28 | 2.08 | 1.58 | 6.66 | 7.42 | 12.25 | 9.02 | 5.28 |
| Past Year Received Treatment for Illicit Drug Use | 2.57 | 2.57 | 4.30 | 2.69 | 1.90 | 6.88 | 7.17 | 12.53 | 9.72 | 5.75 |
| Past Year Received Treatment for Alcohol Use | 2.56 | 2.51 | 4.22 | 2.76 | 2.06 | 6.82 | 7.24 | 12.05 | 9.67 | 5.93 |
| Past Month Use of Any Illicit Drug but Marijuana | 2.43 | 2.49 | 4.32 | 2.75 | 1.85 | 6.78 | 7.57 | 12.48 | 10.04 | 5.23 |
| Dependence on Illicit Drugs | 2.56 | 2.63 | 4.33 | 2.66 | 1.80 | 6.84 | 7.42 | 12.51 | 9.62 | 5.02 |
| Past Month Illicit Drug Use | 2.57 | 2.57 | 4.32 | 2.86 | 1.83 | 6.84 | 7.13 | 12.37 | 9.92 | 5.29 |
| Average Relative Standard Error | 2.54 | 2.59 | 4.55 | 2.81 | 2.01 | 6.82 | 7.34 | 12.46 | 9.97 | 5.80 |
| Target Relative Standard Error | 5.00 | 5.00 | 5.00 | 5.00 | 3.00 | 11.00 | 11.00 | 11.00 | 11.00 | n/a |
| | | | | | | 0 | | | | |
| Outcome Measure | Black Respondents | | | | | White Respondents | | | | |
| | 12-17 | 18-25 | 26-34 | 35+ | Total | 12-17 | 18-25 | 26-34 | 35+ | Total |
| Expected Relative Standard Error for Classes of Outcome Measures | | | | | | | | | | |
| Past Year, Dependence on Alcohol (not Illicit Drugs) | 6.75 | 7.14 | 12.15 | 9.19 | 6.40 | 2.94 | 3.10 | 5.20 | 3.35 | 2.56 |
| Past Month Alcohol Use | 7.01 | 7.19 | 12.03 | 9.32 | 6.34 | 3.04 | 3.11 | 5.20 | 3.38 | 2.83 |
| Past Month Cigarette Use | 6.63 | 7.31 | 12.20 | 9.16 | 6.54 | 2.85 | 3.02 | 5.40 | 3.27 | 2.53 |
| Past Month Cocaine Use | 6.70 | 6.48 | 11.07 | 8.04 | 5.65 | 2.90 | 2.85 | 4.98 | 2.37 | 1.67 |
| Past Year Received Treatment for Illicit Drug Use | 6.41 | 6.98 | 12.27 | 8.29 | 5.88 | 2.97 | 3.07 | 4.97 | 2.92 | 2.09 |
| Past Year Received Treatment for Alcohol Use | 6.42 | 6.52 | 12.21 | 8.55 | 6.22 | 2.94 | 3.00 | 4.90 | 2.91 | 2.30 |
| Past Month Use of Any Illicit Drug but Marijuana | 6.67 | 6.84 | 11.95 | 8.44 | 5.34 | 2.82 | 2.87 | 5.02 | 3.04 | 2.00 |
| Dependence on Illicit Drugs | 6.45 | 7.01 | 12.17 | 8.50 | 5.89 | 2.93 | 3.15 | 5.09 | 2.85 | 2.00 |
| Past Month Illicit Drug Use | 6.43 | 6.85 | 12.18 | 8.67 | 5.36 | 2.93 | 3.07 | 5.04 | 3.04 | 2.04 |
| Average Relative Standard Error | 6.61 | 6.92 | 12.03 | 8.68 | 5.96 | 2.92 | 3.03 | 5.09 | 3.01 | 2.22 |
| Target Relative Standard Error | 11.00 | 11.00 | 11.00 | 11.00 | n/a | 5.00 | 5.00 | 5.00 | 5.00 | n/a |

Note: Relative Standard Errors are based on a prevalence rate of 10%.
n/a = not applicable.

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.

Furthermore, race/ethnicity groups were not oversampled for the 2002 main study. However, consistent with previous NHSDAs, the 2002 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 percent, 104 percent, 104 percent, and 96 percent. Only minor increases in unequal weighting resulted from not distributing the sample equally across quarters.

3.3 Second-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 dwelling units to select for each segment to meet the targeted sample sizes. In short, this step involved determining the sample size of the second-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 originally developed for initial implementation in Quarter 1 of the survey. The description below is specific to Quarter 1. Any modifications or corrections were made in subsequent quarters and are explained in detail in Section 3.7.

3.3.1 Dwelling Unit Frame Construction—Counting and Listing

The process by which the dwelling unit 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 dwelling unit within the segment boundaries. The lister is given a series of maps on which to mark the locations of these dwelling units. 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 dwelling units is entered into a database and serves as the frame from which the second-stage sample is drawn.

In some situations, the number of dwelling units within the segment boundaries was much larger than the specified maximum. To obtain a reasonable number of dwelling units for the frame, the lister first counted the dwelling units in such an area. The sampling staff at RTI then partitioned the segment into smaller pieces or subsegments and randomly selected one to be listed. The number of segments which were subsegmented in the 2002 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." In short, the lister attempts to move in a clockwise fashion, makes each possible right turn, makes U-turns at segment boundaries, and doesn't break street sections. Following these defined rules and always looking for dwelling units on the right-hand side of the street, the lister minimizes the chance of not listing a dwelling unit within the segment. Also, using a defined path of travel makes it easier for the FI assigned to the segment to locate the sampled dwelling units. Finally, the continuous path of travel lays the groundwork for the half-open interval procedure for recovering missed dwelling units, as described in Section 3.7 of this report. A detailed description of the counting and listing procedures is provided in the *2002 NSDUH: Counting and listing general manual* (RTI, 2001).

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

| State | State FIPS | FI Regions | Total Segments | Total Respondents | | | | | |
|-------------------------|------------|------------|----------------|-------------------|--------|-------|--------|-------|--------|
| | | | | 12-17 | 18-25 | 26-34 | 35-49 | 50+ | Total |
| Total Population | | 900 | 7,200 | 22,500 | 22,500 | 6,500 | 10,000 | 6,000 | 67,500 |
| Northeast | | | | | | | | | |
| Connecticut | 09 | 12 | 96 | 300 | 300 | 85 | 134 | 81 | 900 |
| Maine | 23 | 12 | 96 | 300 | 300 | 79 | 138 | 82 | 900 |
| Massachusetts | 25 | 12 | 96 | 300 | 300 | 93 | 131 | 77 | 900 |
| New Hampshire | 33 | 12 | 96 | 300 | 300 | 87 | 142 | 71 | 900 |
| New Jersey | 34 | 12 | 96 | 300 | 300 | 85 | 135 | 80 | 900 |
| New York | 36 | 48 | 384 | 1,200 | 1,200 | 356 | 524 | 320 | 3,600 |
| Pennsylvania | 42 | 48 | 384 | 1,200 | 1,200 | 331 | 519 | 350 | 3,600 |
| Rhode Island | 44 | 12 | 96 | 300 | 300 | 91 | 129 | 80 | 900 |
| Vermont | 09 | 12 | 96 | 300 | 300 | 86 | 139 | 75 | 900 |
| Midwest | | | | | | | | | |
| Illinois | 17 | 48 | 384 | 1,200 | 1,200 | 358 | 535 | 307 | 3,600 |
| Indiana | 18 | 12 | 96 | 300 | 300 | 89 | 133 | 79 | 900 |
| Iowa | 19 | 12 | 96 | 300 | 300 | 83 | 130 | 87 | 900 |
| Kansas | 20 | 12 | 96 | 300 | 300 | 85 | 134 | 81 | 900 |
| Michigan | 26 | 48 | 384 | 1,200 | 1,200 | 351 | 538 | 311 | 3,600 |
| Minnesota | 27 | 12 | 96 | 300 | 300 | 86 | 139 | 75 | 900 |
| Missouri | 29 | 12 | 96 | 300 | 300 | 84 | 133 | 83 | 900 |
| Nebraska | 31 | 12 | 96 | 300 | 300 | 83 | 134 | 83 | 900 |
| North Dakota | 38 | 12 | 96 | 300 | 300 | 85 | 130 | 86 | 900 |
| Ohio | 39 | 48 | 384 | 1,200 | 1,200 | 344 | 530 | 326 | 3,600 |
| South Dakota | 46 | 12 | 96 | 300 | 300 | 82 | 134 | 85 | 900 |
| Wisconsin | 55 | 12 | 96 | 300 | 300 | 86 | 135 | 80 | 900 |

(continued)

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

| State | State FIPS | FI Regions | Total Segments | Total Respondents | | | | | |
|----------------------|------------|------------|----------------|-------------------|-------|-------|-------|-----|-------|
| | | | | 12-17 | 18-25 | 26-34 | 35-49 | 50+ | Total |
| South | | | | | | | | | |
| Alabama | 01 | 12 | 96 | 300 | 300 | 87 | 129 | 83 | 900 |
| Arkansas | 05 | 12 | 96 | 300 | 300 | 83 | 127 | 90 | 900 |
| Delaware | 10 | 12 | 96 | 300 | 300 | 90 | 133 | 77 | 900 |
| District of Columbia | 11 | 12 | 96 | 300 | 300 | 95 | 127 | 78 | 900 |
| Florida | 12 | 48 | 384 | 1,200 | 1,200 | 307 | 501 | 392 | 3,600 |
| Georgia | 13 | 12 | 96 | 300 | 300 | 93 | 137 | 69 | 900 |
| Kentucky | 21 | 12 | 96 | 300 | 300 | 86 | 132 | 82 | 900 |
| Louisiana | 22 | 12 | 96 | 300 | 300 | 88 | 132 | 80 | 900 |
| Maryland | 24 | 12 | 96 | 300 | 300 | 88 | 140 | 72 | 900 |
| Mississippi | 28 | 12 | 96 | 300 | 300 | 91 | 128 | 81 | 900 |
| North Carolina | 37 | 12 | 96 | 300 | 300 | 89 | 131 | 80 | 900 |
| Oklahoma | 40 | 12 | 96 | 300 | 300 | 82 | 130 | 88 | 900 |
| South Carolina | 45 | 12 | 96 | 300 | 300 | 87 | 132 | 81 | 900 |
| Tennessee | 47 | 12 | 96 | 300 | 300 | 87 | 133 | 80 | 900 |
| Texas | 48 | 48 | 384 | 1,200 | 1,200 | 366 | 544 | 290 | 3,600 |
| Virginia | 51 | 12 | 96 | 300 | 300 | 90 | 136 | 74 | 900 |
| West Virginia | 54 | 12 | 96 | 300 | 300 | 80 | 127 | 94 | 900 |
| West | | | | | | | | | |
| Alaska | 02 | 12 | 96 | 300 | 300 | 88 | 154 | 58 | 900 |
| Arizona | 04 | 12 | 96 | 300 | 300 | 86 | 130 | 84 | 900 |
| California | 06 | 48 | 384 | 1,200 | 1,200 | 385 | 539 | 276 | 3,600 |
| Colorado | 08 | 12 | 96 | 300 | 300 | 85 | 142 | 73 | 900 |
| Hawaii | 15 | 12 | 96 | 300 | 300 | 83 | 135 | 82 | 900 |
| Idaho | 16 | 12 | 96 | 300 | 300 | 85 | 133 | 81 | 900 |
| Montana | 30 | 12 | 96 | 300 | 300 | 76 | 136 | 88 | 900 |
| Nevada | 32 | 12 | 96 | 300 | 300 | 83 | 137 | 80 | 900 |
| New Mexico | 35 | 12 | 96 | 300 | 300 | 85 | 137 | 78 | 900 |
| Oregon | 41 | 12 | 96 | 300 | 300 | 80 | 135 | 85 | 900 |
| Utah | 49 | 12 | 96 | 300 | 300 | 102 | 128 | 70 | 900 |
| Washington | 53 | 12 | 96 | 300 | 300 | 85 | 140 | 75 | 900 |
| Wyoming | 56 | 12 | 96 | 300 | 300 | 79 | 140 | 81 | 900 |

FIPS = Federal Information Processing Standards.

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 | 7,200 | 40,757 | 5.7 |
| Alabama | 96 | 680 | 7.1 |
| Alaska | 96 | 534 | 5.6 |
| Arizona | 96 | 472 | 4.9 |
| Arkansas | 96 | 701 | 7.3 |
| California | 384 | 1,531 | 4.0 |
| Colorado | 96 | 491 | 5.1 |
| Connecticut | 96 | 358 | 3.7 |
| Delaware | 96 | 469 | 4.9 |
| District of Columbia | 96 | 256 | 2.7 |
| Florida | 384 | 2,096 | 5.5 |
| Georgia | 96 | 511 | 5.3 |
| Hawaii | 96 | 302 | 3.1 |
| Idaho | 96 | 668 | 7.0 |
| Illinois | 384 | 1,862 | 4.8 |
| Indiana | 96 | 616 | 6.4 |
| Iowa | 96 | 732 | 7.6 |
| Kansas | 96 | 709 | 7.4 |
| Kentucky | 96 | 540 | 5.6 |
| Louisiana | 96 | 629 | 6.6 |
| Maine | 96 | 640 | 6.7 |
| Maryland | 96 | 370 | 3.9 |
| Massachusetts | 96 | 443 | 4.6 |
| Michigan | 384 | 2,340 | 6.1 |
| Minnesota | 96 | 616 | 6.4 |
| Mississippi | 96 | 657 | 6.8 |
| Missouri | 96 | 465 | 4.8 |
| Montana | 96 | 868 | 9.0 |
| Nebraska | 96 | 819 | 8.5 |
| Nevada | 96 | 414 | 4.3 |
| New Hampshire | 96 | 525 | 5.5 |
| New Jersey | 96 | 395 | 4.1 |
| New Mexico | 96 | 704 | 7.3 |
| New York | 384 | 1,520 | 4.0 |
| North Carolina | 96 | 590 | 6.1 |
| North Dakota | 96 | 1,363 | 14.2 |
| Ohio | 384 | 1,604 | 4.2 |
| Oklahoma | 96 | 581 | 6.1 |
| Oregon | 96 | 531 | 5.5 |
| Pennsylvania | 384 | 2,438 | 6.3 |
| Rhode Island | 96 | 543 | 5.7 |
| South Carolina | 96 | 673 | 7.0 |
| South Dakota | 96 | 1,062 | 11.1 |
| Tennessee | 96 | 536 | 5.6 |
| Texas | 384 | 1,913 | 5.0 |

(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 | 458 | 4.8 |
| Vermont | 96 | 541 | 5.6 |
| Virginia | 96 | 460 | 4.8 |
| Washington | 96 | 381 | 4.0 |
| West Virginia | 96 | 618 | 6.4 |
| Wisconsin | 96 | 572 | 6.0 |
| Wyoming | 96 | 960 | 10.0 |

Table 3.4 Segment and Dwelling Unit Summary

| State | Total Segments | Subsegmented Segments | Listed Dwelling Units | Added Dwelling Units |
|-------------------------|-----------------------|----------------------------------|----------------------------------|---------------------------------|
| Total Population | 7,200 | 810 | 1,606,893 | 1,766 |
| Alabama | 96 | 8 | 21,148 | 16 |
| Alaska | 96 | 15 | 21,234 | 28 |
| Arizona | 96 | 13 | 19,305 | 15 |
| Arkansas | 96 | 5 | 20,936 | 4 |
| California | 384 | 28 | 85,574 | 52 |
| Colorado | 96 | 12 | 20,672 | 15 |
| Connecticut | 96 | 6 | 22,308 | 47 |
| Delaware | 96 | 14 | 22,682 | 112 |
| District of Columbia | 96 | 25 | 25,980 | 13 |
| Florida | 384 | 83 | 89,724 | 108 |
| Georgia | 96 | 17 | 21,404 | 17 |
| Hawaii | 96 | 20 | 22,070 | 40 |
| Idaho | 96 | 5 | 19,190 | 14 |
| Illinois | 384 | 39 | 82,339 | 52 |
| Indiana | 96 | 6 | 20,921 | 27 |
| Iowa | 96 | 5 | 20,218 | 12 |
| Kansas | 96 | 13 | 21,029 | 12 |
| Kentucky | 96 | 9 | 22,385 | 29 |
| Louisiana | 96 | 5 | 23,814 | 9 |
| Maine | 96 | 6 | 23,206 | 40 |
| Maryland | 96 | 24 | 21,559 | 20 |
| Massachusetts | 96 | 7 | 21,671 | 42 |
| Michigan | 384 | 31 | 88,045 | 94 |
| Minnesota | 96 | 4 | 21,009 | 23 |
| Mississippi | 96 | 5 | 20,544 | 6 |
| Missouri | 96 | 10 | 22,838 | 37 |
| Montana | 96 | 13 | 18,722 | 27 |
| Nebraska | 96 | 11 | 21,073 | 8 |
| Nevada | 96 | 27 | 19,995 | 11 |
| New Hampshire | 96 | 8 | 23,228 | 134 |
| New Jersey | 96 | 10 | 20,888 | 22 |
| New Mexico | 96 | 19 | 21,551 | 10 |

(continued)

Table 3.4 Segment and Dwelling Unit Summary (continued)

| State | Total Segments | Subsegmented Segments | Listed Dwelling Units | Added Dwelling Units |
|----------------|----------------|-----------------------|-----------------------|----------------------|
| New York | 384 | 50 | 87,989 | 109 |
| North Carolina | 96 | 12 | 22,706 | 17 |
| North Dakota | 96 | 9 | 20,842 | 21 |
| Ohio | 384 | 21 | 83,845 | 52 |
| Oklahoma | 96 | 5 | 21,815 | 31 |
| Oregon | 96 | 11 | 20,478 | 6 |
| Pennsylvania | 384 | 29 | 84,295 | 70 |
| Rhode Island | 96 | 7 | 22,393 | 69 |
| South Carolina | 96 | 12 | 21,882 | 5 |
| South Dakota | 96 | 9 | 18,635 | 33 |
| Tennessee | 96 | 10 | 21,867 | 10 |
| Texas | 384 | 69 | 87,513 | 57 |
| Utah | 96 | 11 | 20,339 | 8 |
| Vermont | 96 | 3 | 21,552 | 89 |
| Virginia | 96 | 18 | 21,988 | 23 |
| Washington | 96 | 16 | 20,114 | 24 |
| West Virginia | 96 | 5 | 21,843 | 16 |
| Wisconsin | 96 | 7 | 19,285 | 15 |
| Wyoming | 96 | 3 | 20,250 | 15 |

3.3.2 Determining Dwelling Unit Sample Size

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

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

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 dwelling unit, 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} is then 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 Third-Stage Sample (Person) Selection Probabilities for Each Segment

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

Having solved for D_{hj} , solve the selection probabilities for the remaining age groups. If L_{hj} equals 0, then 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 dwelling units remain within each segment. This sample surplus is needed to provide for the yearly 50 percent overlap across segments and to allow SAMHSA to implement supplemental studies. An adequate remaining sample has two advantages: (1) for the 50 percent overlap design, this will provide better precision in year-to-year trend estimates because of the expected positive correlation between successive NSDUH years; and (2) it will reduce the amount of counting and listing costs.

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

1. Number of selected dwelling units for screening: < 100 or $< \frac{1}{2}L_{hj}$. Adjustments were made by adjusting the D_{hj} counts to equal the minimum of 100 or $\frac{1}{2}L_{hj}$.
2. Number of selected dwelling units: > 5 . For cost purposes, if at least five dwelling units remain in the segment, the minimum number of selected dwelling units was set to five.
3. Expected number of interviews: < 40 .

This expected number of interviews ($m^*_{hja(main)}$) was computed for the main study as follows:

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

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

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

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

Both constraints 1 and 3 reduce the second-stage sample. This in turn, could potentially reduce the expected third-stage sample size. Therefore, the reduction in the second-stage sample is reallocated back to the segments by applying a marginal adjustment to the third-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 dwelling unit sample is iterative until the expected person sample sizes are met.

3.6 Dwelling Unit Selection and Release Partitioning

After derivation of the required dwelling unit sample size (D_{hj}), the sample was selected from the frame of counted and listed dwelling units 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.

In order to compensate for quarterly variations in response rates and yields, a sample partitioning procedure was implemented in all quarters. The entire sample (D_{hj}) would still 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 10 percent was also 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 105.6 percent level. In Quarter 1, the D_{hj} sample was allocated out to FI regions in the following release percentages:

Release 1: 100 percent of main sample (96 percent of quarterly sample);

Release 2: 100 percent of reserve sample (10 percent of main sample).

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 dwelling unit had a chance of selection and to eliminate any bias associated with incomplete frames, the NSDUH implemented a procedure called the half-open interval rule. This procedure required that the interviewer look both on the property of each selected dwelling unit and between that dwelling unit and the next listed dwelling unit for any unlisted units. When found in these specific locations, the unlisted units became part of the sample (added dwelling units). If the number of added dwelling units linked to any particular sample dwelling unit did not exceed five, or if the number for the entire segment was less than or equal to ten, the FI was instructed to consider these dwelling units as part of their assignment. If either of these limits was exceeded, special subsampling procedures were implemented, as described in Appendix C. The number of added dwelling units in the 2002 NSDUH sample is summarized in Table 3.4.

3.8 Quarter-by-Quarter Deviations

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

Table 3.5 Quarterly Sample Sizes and Percentages Released

| State | Quarter 1 | | | Quarter 2 | | |
|-------------------------|------------|------------|------------|------------|------------|------------|
| | # Selected | # Released | Percentage | # Selected | # Released | Percentage |
| Total Population | 47,507 | 43,128 | 91% | 51,738 | 45,268 | 87% |
| Northeast | | | | | | |
| Connecticut | 743 | 677 | 91% | 806 | 625 | 78% |
| Maine | 731 | 664 | 91% | 832 | 721 | 87% |
| Massachusetts | 697 | 631 | 91% | 734 | 629 | 86% |
| New Hampshire | 754 | 685 | 91% | 814 | 626 | 77% |
| New Jersey | 666 | 607 | 91% | 712 | 581 | 82% |
| New York | 2,819 | 2,555 | 91% | 2,969 | 2,835 | 95% |
| Pennsylvania | 2,650 | 2,369 | 89% | 2,961 | 2,563 | 87% |
| Rhode Island | 633 | 578 | 91% | 705 | 546 | 77% |
| Vermont | 654 | 594 | 91% | 722 | 524 | 73% |
| Midwest | | | | | | |
| Illinois | 2,574 | 2,337 | 91% | 2,813 | 2,421 | 86% |
| Indiana | 619 | 562 | 91% | 733 | 600 | 82% |
| Iowa | 542 | 493 | 91% | 591 | 591 | 100% |
| Kansas | 544 | 494 | 91% | 610 | 446 | 73% |
| Michigan | 2,712 | 2,466 | 91% | 2,960 | 2,289 | 77% |
| Minnesota | 548 | 501 | 91% | 633 | 542 | 86% |
| Missouri | 705 | 639 | 91% | 772 | 736 | 95% |
| Nebraska | 487 | 443 | 91% | 539 | 539 | 100% |
| North Dakota | 648 | 589 | 91% | 703 | 575 | 82% |
| Ohio | 2,355 | 2,141 | 91% | 2,509 | 2,391 | 95% |
| South Dakota | 540 | 490 | 91% | 572 | 572 | 100% |
| Wisconsin | 634 | 574 | 91% | 708 | 515 | 73% |

(continued)

Table 3.5 Quarterly Sample Sizes and Percentages Released (continued)

| State | Quarter 1 | | | Quarter 2 | | |
|-------------------------|------------|------------|------------|------------|------------|------------|
| | # Selected | # Released | Percentage | # Selected | # Released | Percentage |
| Total Population | 47,507 | 43,128 | 91% | 51,738 | 45,268 | 87% |
| South | | | | | | |
| Alabama | 588 | 537 | 91% | 623 | 623 | 100% |
| Arkansas | 686 | 627 | 91% | 693 | 693 | 100% |
| Delaware | 594 | 540 | 91% | 665 | 665 | 100% |
| District of Columbia | 1,092 | 988 | 90% | 1,248 | 903 | 72% |
| Florida | 2,783 | 2,525 | 91% | 3,076 | 2,795 | 91% |
| Georgia | 602 | 550 | 91% | 633 | 573 | 91% |
| Kentucky | 648 | 588 | 91% | 722 | 722 | 100% |
| Louisiana | 540 | 493 | 91% | 607 | 607 | 100% |
| Maryland | 523 | 473 | 90% | 537 | 508 | 95% |
| Mississippi | 627 | 569 | 91% | 668 | 487 | 73% |
| North Carolina | 652 | 592 | 91% | 707 | 545 | 77% |
| Oklahoma | 614 | 558 | 91% | 692 | 630 | 91% |
| South Carolina | 652 | 592 | 91% | 750 | 547 | 73% |
| Tennessee | 760 | 693 | 91% | 785 | 679 | 86% |
| Texas | 1,996 | 1,820 | 91% | 2,206 | 2,114 | 96% |
| Virginia | 725 | 658 | 91% | 760 | 553 | 73% |
| West Virginia | 761 | 693 | 91% | 818 | 631 | 77% |
| West | | | | | | |
| Alaska | 647 | 587 | 91% | 736 | 736 | 100% |
| Arizona | 651 | 592 | 91% | 681 | 524 | 77% |
| California | 2,286 | 2,071 | 91% | 2,482 | 2,370 | 95% |
| Colorado | 585 | 530 | 91% | 609 | 443 | 73% |
| Hawaii | 609 | 554 | 91% | 654 | 593 | 91% |
| Idaho | 512 | 465 | 91% | 542 | 542 | 100% |
| Montana | 690 | 628 | 91% | 734 | 703 | 96% |
| Nevada | 636 | 578 | 91% | 656 | 656 | 100% |
| New Mexico | 593 | 543 | 92% | 612 | 445 | 73% |
| Oregon | 649 | 591 | 91% | 724 | 588 | 81% |
| Utah | 356 | 318 | 89% | 426 | 406 | 95% |
| Washington | 580 | 528 | 91% | 656 | 656 | 100% |
| Wyoming | 615 | 558 | 91% | 638 | 464 | 73% |

(continued)

Table 3.5 Quarterly Sample Sizes and Percentages Released (continued)

| State | Quarter 3 | | | Quarter 4 | | |
|-------------------------|------------|------------|------------|------------|------------|------------|
| | # Selected | # Released | Percentage | # Selected | # Released | Percentage |
| Total Population | 50,410 | 47,468 | 94% | 45,433 | 40,335 | 89% |
| Northeast | | | | | | |
| Connecticut | 732 | 667 | 91% | 734 | 702 | 96% |
| Maine | 845 | 769 | 91% | 732 | 634 | 87% |
| Massachusetts | 723 | 657 | 91% | 608 | 608 | 100% |
| New Hampshire | 745 | 678 | 91% | 617 | 474 | 77% |
| New Jersey | 729 | 729 | 100% | 615 | 615 | 100% |
| New York | 3,024 | 2,750 | 91% | 2,885 | 2,231 | 77% |
| Pennsylvania | 2,818 | 2,818 | 100% | 2,465 | 2,237 | 91% |
| Rhode Island | 708 | 706 | 100% | 643 | 559 | 87% |
| Vermont | 682 | 619 | 91% | 642 | 584 | 91% |
| Midwest | | | | | | |
| Illinois | 2,677 | 2,305 | 86% | 2,479 | 2,148 | 87% |
| Indiana | 723 | 620 | 86% | 585 | 452 | 77% |
| Iowa | 638 | 638 | 100% | 605 | 518 | 86% |
| Kansas | 632 | 517 | 82% | 488 | 464 | 95% |
| Michigan | 2,785 | 2,658 | 95% | 2,425 | 2,313 | 95% |
| Minnesota | 624 | 538 | 86% | 569 | 569 | 100% |
| Missouri | 744 | 711 | 96% | 665 | 602 | 91% |
| Nebraska | 581 | 581 | 100% | 526 | 383 | 73% |
| North Dakota | 698 | 667 | 96% | 587 | 455 | 78% |
| Ohio | 2,506 | 2,506 | 100% | 2,201 | 2,104 | 96% |
| South Dakota | 604 | 547 | 91% | 504 | 411 | 82% |
| Wisconsin | 629 | 544 | 86% | 553 | 504 | 91% |
| South | | | | | | |
| Alabama | 675 | 674 | 100% | 610 | 553 | 91% |
| Arkansas | 703 | 703 | 100% | 594 | 513 | 86% |
| Delaware | 680 | 618 | 91% | 679 | 650 | 96% |
| District of Columbia | 1,113 | 953 | 86% | 845 | 844 | 100% |
| Florida | 2,968 | 2,693 | 91% | 3,040 | 2,621 | 86% |
| Georgia | 574 | 518 | 90% | 548 | 548 | 100% |
| Kentucky | 764 | 691 | 90% | 670 | 611 | 91% |
| Louisiana | 683 | 654 | 96% | 550 | 426 | 77% |
| Maryland | 524 | 524 | 100% | 480 | 459 | 96% |
| Mississippi | 665 | 665 | 100% | 534 | 534 | 100% |
| North Carolina | 685 | 656 | 96% | 588 | 479 | 81% |
| Oklahoma | 655 | 596 | 91% | 627 | 485 | 77% |
| South Carolina | 732 | 698 | 95% | 592 | 490 | 83% |
| Tennessee | 783 | 783 | 100% | 567 | 567 | 100% |

(continued)

Table 3.5 Quarterly Sample Sizes and Percentages Released (continued)

| State | Quarter 3 | | | Quarter 4 | | |
|-------------------------|------------|------------|------------|------------|------------|------------|
| | # Selected | # Released | Percentage | # Selected | # Released | Percentage |
| Total Population | 50,410 | 47,468 | 94% | 45,433 | 40,335 | 89% |
| Texas | 2,162 | 2,160 | 100% | 1,933 | 1,578 | 82% |
| Virginia | 651 | 651 | 100% | 596 | 541 | 91% |
| West Virginia | 765 | 731 | 96% | 727 | 692 | 95% |
| West | | | | | | |
| Alaska | 644 | 559 | 87% | 607 | 498 | 82% |
| Arizona | 699 | 699 | 100% | 634 | 516 | 81% |
| California | 2,398 | 2,288 | 95% | 2,009 | 1,644 | 82% |
| Colorado | 552 | 552 | 100% | 559 | 559 | 100% |
| Hawaii | 660 | 540 | 82% | 575 | 549 | 95% |
| Idaho | 537 | 537 | 100% | 524 | 475 | 91% |
| Montana | 745 | 745 | 100% | 669 | 669 | 100% |
| Nevada | 655 | 655 | 100% | 697 | 634 | 91% |
| New Mexico | 592 | 488 | 82% | 542 | 464 | 86% |
| Oregon | 662 | 662 | 100% | 749 | 609 | 81% |
| Utah | 399 | 363 | 91% | 392 | 392 | 100% |
| Washington | 640 | 640 | 100% | 606 | 606 | 100% |
| Wyoming | 598 | 547 | 91% | 562 | 562 | 100% |

Quarter 2

Design: An additional 10 percent 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 114.4 percent level.

Procedural: In order 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 2001 NHSDA were used to compute average yields, dwelling unit eligibility, screening response, and interviewer response rates.

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

Release 1: 73 percent of entire sample (80/110, main sample + 10 percent reserve);

Release 2: 5 percent of entire sample (5/110, main sample + 10 percent reserve);

Release 3: 5 percent of entire sample (5/110, main sample + 10 percent reserve);

Release 4: 9 percent of entire sample (10/110, main sample + 10 percent reserve);

Release 5: 9 percent of entire sample (10/110, main sample + 10 percent reserve).

Quarter 3

Design: Using the completed cases from Quarter 1 and the projected number of completes from Quarter 2, each State's mid-year surplus/shortfall was computed. The Quarter 3 104 percent sample was then adjusted by this amount. An additional 10 percent sample was also included, bringing the total Quarter 3 adjusted sample to the 114.4 percent level.

Procedural: Data from Quarters 2 through 4 of the 2001 NHSDA and Quarter 1 of the 2002 NSDUH¹⁴ were used to compute State-specific average yields, dwelling unit 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: 73 percent of entire sample (80/110, main sample + 10 percent reserve);

Release 2: 5 percent of entire sample (5/110, main sample + 10 percent reserve);

Release 3: 5 percent of entire sample (5/110, main sample + 10 percent reserve);

Release 4: 9 percent of entire sample (10/110, main sample + 10 percent reserve);

Release 5: 9 percent of entire sample (10/110, main sample + 10 percent reserve).

Quarter 4

Design: The State and age 96 percent quarterly sample sizes were adjusted in order to meet the yearly targets based on completed cases from Quarter 1 and 2 and the projected number of completes from Quarter 3. An additional 10 percent sample was also included, bringing the total Quarter 4 adjusted sample to the 105.6 percent level.

¹⁴ No adjustment was made to the interview response rate when using the 2002 data since the incentives and additional pair sampling were already in place.

| | |
|--------------------------|--|
| Procedural: | Data from Quarters 3 and 4 of the 2001 NHSDA and Quarters 1 and 2 of the 2002 NSDUH ¹⁵ were used to compute State-specific average yields, dwelling unit eligibility, screening response, and interviewer response rates. |
| Dwelling Unit Selection: | The Quarter 4 D_{hj} sample was partitioned into the following release percentages: |
| Selection: | <p><i>Release 1:</i> 73 percent of entire sample (80/110, main sample + 10 percent reserve);</p> <p><i>Release 2:</i> 5 percent of entire sample (5/110, main sample + 10 percent reserve);</p> <p><i>Release 3:</i> 5 percent of entire sample (5/110, main sample + 10 percent reserve);</p> <p><i>Release 4:</i> 9 percent of entire sample (10/110, main sample + 10 percent reserve);</p> <p><i>Release 5:</i> 9 percent of entire sample (10/110, main sample + 10 percent reserve).</p> |

3.9 Sample Weighting Procedures

At the conclusion of data collection for the last quarter, design weights were constructed for each quarter of the State-level study, reflecting the various stages of sampling. The calculation of the sampling weights was based on the stratified, three-stage design of the study. Specifically, the person-level sampling weights were the product of the three 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 segment. |
| Stage 2: | <p>Selection of dwelling unit.</p> <p>Three possible adjustments exist with this stage of selection:</p> <ol style="list-style-type: none"> (1) Subsegmentation inflation – by-product of counting and listing, (2) Added dwelling unit – results from the half-open interval rule when subsampling is needed, and (3) Release adjustment. |
| Stage 3: | Selection of person within a dwelling unit. |

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

¹⁵ Again, no adjustment was made to the interview response rate when using the 2002 data since the incentives and additional pair sampling were already in place.

1. *Nonresponse Adjustment at the Dwelling Unit Level.* This was to account for the failure to complete the within-dwelling unit roster. The potential list of variables for the 51-State main study dwelling unit nonresponse modeling is presented in Table 3.6.
2. *Dwelling Unit–Level Poststratification.* This involved using screener data of demographic information (e.g., age, race, gender, etc.). Dwelling unit weights were adjusted to the intercensal population estimates derived from the 2000 U.S. Census for various demographic domains. In short, explanatory variables used during modeling consisted of counts of eligible persons within each dwelling unit that fell into the various demographic categories. Consequently, these counts, multiplied by the newly adjusted dwelling unit weight and summed across all dwelling units for various domains, 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 was determined that design-based weights (stages 1 and 2) along with any of their respective adjustments resulted in an unsatisfactory unequal weighting effect (i.e., variance of the dwelling unit–level weights was too high, with high frequency of extreme weights), then extreme weights were further adjusted. This was implemented by doing another weight calibration. The control totals are the dwelling unit–level poststratified weights, and the same explanatory variables as in dwelling unit–level poststratification were used so that the extreme weights were controlled and all the distributions in various demographic groups were preserved.
4. *Selected Person Weight Adjustment for Poststratification to Roster Data.* This step utilized control totals derived from the dwelling unit roster that were already poststratified to the Census population estimates. This assisted 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 dwelling unit level and account for third-stage sampling) were adjusted to the dwelling unit weight sums of all eligible rostered persons. Any demographic information used in modeling is based solely on screener information, since 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 allowed for the correction of weights resulting from the failure of selected sample persons to complete the interview. Respondent sample weights were 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.
6. *Person-Level Poststratification.* This step was to adjust the final person sample weights to the Census population estimates derived from the 2000 U.S. Census. These were the same outside control totals used in the second adjustment. However, demographic variables for this adjustment are based on questionnaire data, not

screeners 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 was implemented in the same manner as described above in adjustment 3, except the weights reflect the third stage of selection.

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

Group Quarter 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 Black in Segment (% Black)

- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Percentage of Hispanic in Segment (% Hispanic)

- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Population Density

- 1: MSA > 1,000,000
- 2: MSA < 1,000,000
- 3: Non-MSA urban
- 4: Non-MSA 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

Interactions among the main effect variables are also considered.

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+¹⁶

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: Indian/Native American
- 4: Asian
- 5: Multiple Race

State

Interactions among the main effect variables are also considered.

¹⁶ For person-level respondents poststratification adjustment, the age category of 50+ is further divided into 50-64 and 65+ categories.

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

Group Quarter 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 Black in Segment (% Black)

- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Percentage of Hispanic in Segment (% Hispanic)

- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Population Density

- 1: MSA > 1,000,000
- 2: MSA < 1,000,000
- 3: Non-MSA urban
- 4: Non-MSA 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: Indian/Native American |
| 4: Asian |
| 5: Multiple Race |
| Relation to Householder |
| 1: Householder or Spouse |
| 2: Child |
| 3: Other Relative |
| 4: Non-Relative |

Interactions among the main effect variables are also considered.

All weight adjustments for the 2002 main study final analysis weights were derived from a generalized exponential model. To help reduce computational burden at all adjustment steps, separate models were fit for clusters of States, based on Census Division definitions as shown in Table 3.9. Furthermore, model variable selection at each adjustment was done using a combination method of forward and backward selection processes. The forward selection is in the sense of model enlargement. Within each enlargement, backward selection was used. The final adjusted weight, which is the product of weight components 1 through 14, is the analysis weight used in estimation. Table 3.10 presents a flowchart of steps used in the weighting process, and Table 3.11 displays all individual weight components.

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 |

Table 3.10 Flowchart of Sample Weighting Steps

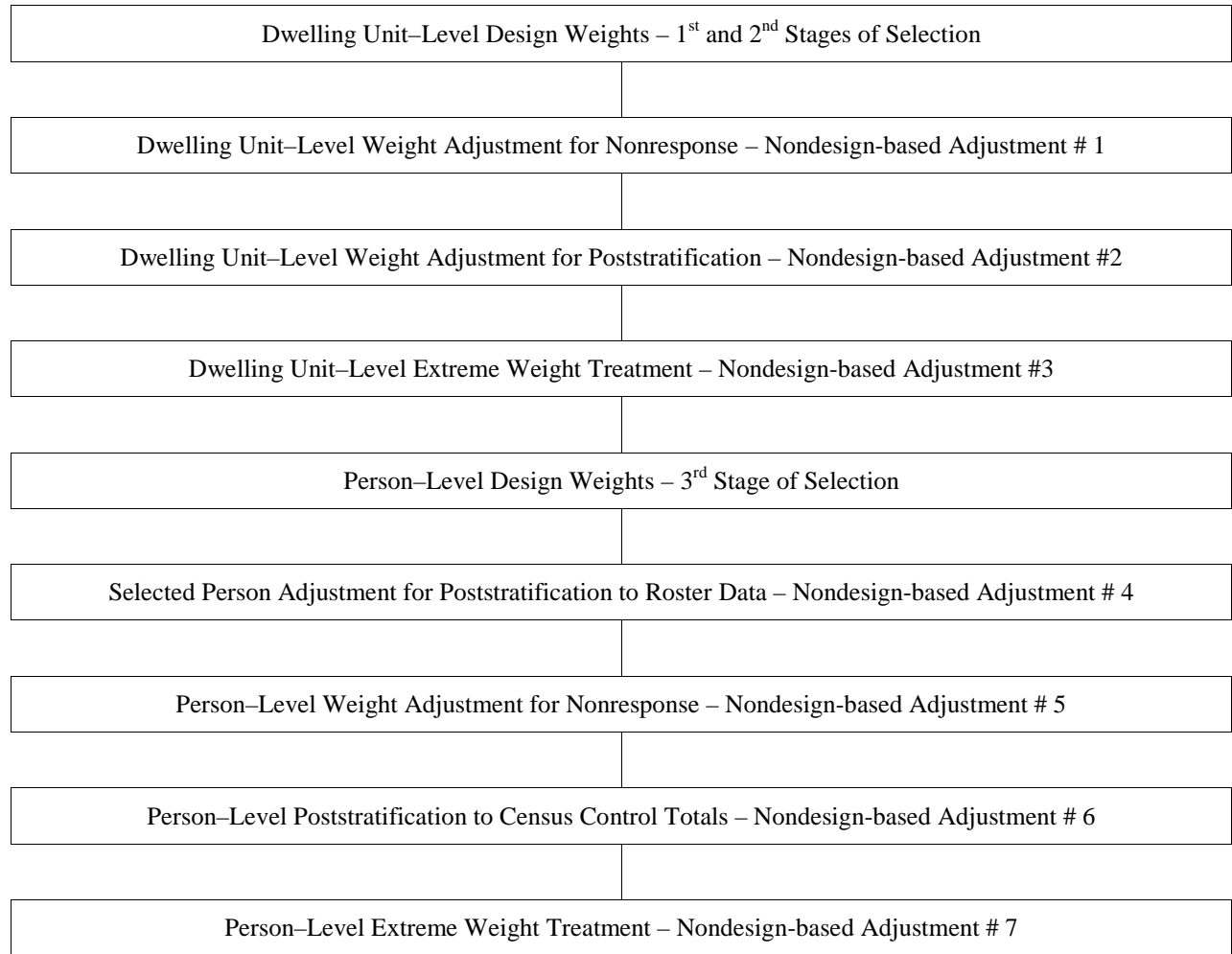


Table 3.11 Sample Weight Components

| Dwelling Unit–Level Design Weight Components | |
|---|--|
| #1. | Inverse Probability of Selecting Segment |
| #2. | Quarter Segment Weight Adjustment |
| #3. | Subsegmentation Inflation Adjustment |
| #4. | Inverse Probability of Selecting Dwelling Unit |
| #5. | Inverse Probability of Added/Subsampled Dwelling Unit |
| #6. | Dwelling Unit Release Adjustment |
| | |
| #7. | Dwelling Unit Nonresponse Adjustment |
| | |
| #8. | Dwelling Unit Poststratification Adjustment |
| | |
| #9. | Dwelling Unit Extreme Weight Adjustment |
| | |
| Person–Level Design Weight Components | |
| #10. | Inverse Probability of Selecting a Person Within a Dwelling Unit |
| | |
| #11. | Selected Person Poststratification to Roster Adjustment |
| | |
| #12. | Person-Level Nonresponse Adjustment |
| | |
| #13. | Person-Level Poststratification Adjustment |
| | |
| #14. | Person-Level Extreme Weight Adjustment |

Full details of the finalized modeling procedures, as well as final variables used in each adjustment step, can be found in *Person-level Sampling Weight Calibration for the 2002 NSDUH* (Chen et al., 2003).

References

- Chen, P., Gordek, H., Dai, L., Singh, A., Shi, W., & Westlake, M. (2003). Person-level sampling weight calibration for the 2002 NSDUH. In *2002 National Survey on Drug Use and Health: Methodological resource book* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-98-9008). Research Triangle Park, NC: Research Triangle Institute.
- Chromy, J. R. (1979). Sequential sample selection methods. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 401-406). Alexandria, VA: American Statistical Association.
- Chromy, J. R. (1981). Variance estimates 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 American Statistical Association, Survey Research Methods Sections* (pp. 194-199). Washington, DC: American Statistical Association.
- Chromy, J. R., & Penne, M. A. (August 2002). Pair sampling in household surveys. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 552-554) New York: American Statistical Association.
- Cochran, W. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley & Sons.
- Kish, L. (1965). *Survey sampling*. New York: John Wiley & Sons.
- RTI. (2001). *2002 National Survey on Drug Use and Health: Counting and listing general manual* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-98-9008). Research Triangle Park, NC: Author.
- RTI. (2001). *SUDAAN user's manual, release 8.0*. Research Triangle Park, NC: RTI.

