2001 National Household Survey on Drug Abuse

Sample Design Report

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

Authors:

Katherine R. Bowman
James R. Chromy
Dawn M. Odom
Michael A. Penne

Project Director: Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:
Research Triangle Institute

April 2003

2001 National Household Survey on Drug Abuse

Sample Design Report

Authors:
Katherine R. Bowman
James R. Chromy
Dawn M. Odom
Michael A. Penne

Project Director:
Thomas G. Virag

Prepared for:

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

April 2003

Acknowledgments

This publication was developed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Office of Applied Studies (OAS), by Research Triangle Institute (RTI), 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, Michael A. Penne, R. David Belton, Brenda K. Porter, and Thomas G. Virag (Project Director).

Table of Contents

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

2. The Coordinated 5-Year Sample 5

2.1 Formation of and Objectives for Using the Composite Size Measures 5
2.2 Stratification 6
2.3 First-Stage Sample Selection 8
2.4 Survey Year and Quarter Assignment 8
2.5 Creation of Variance Estimation Strata 10

3. General Sample Allocation Procedures for the Main Study 13

3.1 Notation 13
3.2 Determining Person Sample Sizes by State and Age Group 15
3.3 Second-Stage Sample Allocation for Each Segment 18
3.3.1 Dwelling Unit Frame Construction—Counting and Listing 18
3.3.2 Determining Dwelling Unit Sample Size 21
3.4 Determining Third-Stage Sample (Person) Selection Probabilities for Each Segment 22
3.5 Sample Size Constraints: Guaranteeing Sufficient Sample for Additional Studies and Reducing Field Interviewer Burden 22
3.6 Dwelling Unit Selection and Release Partitioning 23
3.7 Half-Open Interval Rule and Procedure for Adding Dwelling Units 23
3.8 Quarter-by-Quarter Deviations 23
3.9 Sample Weighting Procedures 28

4. References 37

Appendix A 1999-2003 NHSDA Field Interviewer Regions

Appendix B 2001 NHSDA Procedure for Subsegmenting

Appendix C 2001 NHSDA Procedure for Adding Missed Dwelling Units

List of Tables

2.1 Number of Segments on Sampling Frame by State 7
2.2 Survey Year and Quarter Assignment Order for 96 Segments within Each FI Region 9
2.3 Segment Identification Number Suffixes for the 1999-2003 NHSDAs 11

3.1 Expected Relative Standard Errors by Race/Ethnicity and Age Group: Main Sample 17
3.2 Expected Main Study Sample Sizes by State and Age Group 19
3.3 Segment and Dwelling Unit Summary 20
3.4 Quarterly Sample Sizes and Percent Released 24
3.5 Definitions of Levels for Potential Variables for Dwelling Unit Non-Response Adjustment 30
3.6 Definitions of Levels for Potential Variables for Selected Person Post-Stratification and Non-Response Adjustment 31
3.7 Definitions of Levels for Potential Variables for Dwelling Unit Post-Stratification and Respondent Post-Stratification at the Person Level 32
3.8 Model Group Definitions 34
3.9 Flowchart of Sample Weighting Steps 34
3.10 Sample Weight Components 35

Chapter 1: Overview

1.1 Target Population

The respondent universe for the 2001 National Household Survey on Drug Abuse (NHSDA) was the civilian, noninstitutionalized population aged 12 years or older residing within the United States and the District of Columbia. Consistent with the NHSDA designs since 1991, the 2001 NHSDA 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. Survey coverage before the 1991 NHSDA 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 2001 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 Administration (SAMHSA) implemented major changes in the way the NHSDA would be conducted beginning in 1999 and continuing through subsequent years. The 1999 survey was the first conducted using computer-assisted interviewing (CAI) methods. This survey also marked the first year in a transition to improved state estimates based on minimum sample sizes per state. In addition, it was also the first year in which cigarette brand information was obtained for the Centers for Disease Control and Prevention (CDC). To obtain the required precision at the state level and to improve the precision of cigarette brand data for youth at the national level, the total sample size of 67,500 was increased by 2,500 youths aged 12 to 17 to a total of 70,000 for the 1999 and 2000 surveys. Since no youth supplement was included in the 2001 NHSDA, the total sample size was targeted at the original 67,500. This large sample size allowed 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 2001 CAI sample was 68,929 persons.

Beginning in 2000, SAMHSA, RTI, and the University of Delaware conducted the Validity of Self-Reported Drug Use in Population Surveys (VSRDU) to evaluate and establish baseline information on the validity of survey research methods in assessing recent drug use among the general household population. Previous validity research had been conducted on only specific subgroups, resulting in the need to examine the validity on the overall population. The VSRDU was conducted with a national probability sample of all NHSDA-eligible persons aged 12 to 25. In order to obtain the required precision, the sample size consisted of 1,000 per age group (12 to 17 and 18 to 25) per year (2000 and 2001). Although different surveys, the VSRDU was conducted alongside the NHSDA and used the same interviewer staff. The design of the VSRDU will not be covered in this report, but full documentation will be published elsewhere.

Also included in the 2001 survey was an experimental study to evaluate the effectiveness of respondent incentives in improving response rates. In the first two quarters of 2001, a randomized, split-sample experimental design was included with the main study data collection of the NHSDA to compare the impact of $20 and $40 incentive treatments with a $0 control group on measures of respondent cooperation, data quality, survey costs, and population substance use estimates. While the design of the incentive experiment will not be covered in this report, full documentation can be found in the *2001 NHSDA Incentive Experiment: Combined Quarter 1 and Quarter 2 Analysis* (Eyerman and Bowman, 2002).

Finally, a supplemental sample was added to the New York City area following the September 11 attacks. Although the attacks fell in Quarter 3 of the survey, the majority of the quarter's cases had been completed by mid-September. Thus, the sample was not supplemented until Quarter 4. The targeted sample was increased by 600, 150, and 150 in New York, New Jersey, and Connecticut, respectively. The increased sample in Quarter 4 would allow SAMHSA to measure the impact of the attacks on drug use prevalence and mental health service utilization with greater precision.

1.3 5-Year Design

A coordinated 5-year sample design was developed. The 2001 main sample is a subsample of the 5-year sample. Although there is no overlap with the 1998 sample, a coordinated design for 1999-2003 facilitated 50% overlap in first-stage units (area segments) between each two successive years from 1999 through 2003. This design was intended to increase the precision of estimates in year-to-year trend analyses because of the expected positive correlation resulting from the overlapping sample between successive NHSDA 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,[1] had a sample designed to yield 3,600 respondents per state for the 2001 survey. This sample size was considered adequate to support direct state estimates. The remaining 43 states[2] had a sample designed to yield 900 respondents per state in the 2001 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 equal 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 NHSDA 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.

For the first stage of sampling, each of the FI regions was partitioned into noncompact clusters[3] of dwelling units by aggregating adjacent Census blocks. Consistent with the terminology used in previous NHSDAs, these geographic clusters of blocks are referred to as *segments*. A sample *dwelling unit* in the NHSDA 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

---

[1]For the 1999-2003 NHSDAs, the "large" states are California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas.

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

[3]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).

SAMHSA may wish to conduct, segments were formed to contain a minimum of 175 dwelling units[4] 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) indicator[5] and by the percent of the population that is non-Hispanic and white. From this well-ordered sample frame, 96[6] segments 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 collection as will be 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 2001 NHSDA 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 selection, the subsegmentation inflation factor,[7] 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.[8] In addition to these factors, historical data from the 1999, 2000, and 2001 NHSDAs 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

---

[4]Dwelling unit counts were obtained from the 1990 Decennial Census data supplemented with revised population counts from Claritas.

[5]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 percent 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.”

[6]The 1999-2003 sample was planned such 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.

[7]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.

[8]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.

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 2001 NHSDA using the half-open interval selection technique.[9] 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 NHSDA is the ability to impose a more complicated person-level selection algorithm on the third stage of the NHSDA design. In 1999 and continuing through 2001, 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 NHSDA researchers because, for example, it allows analysts to examine how the drug use propensity of one individual in a family relates to the drug use propensity of other family members residing in the same dwelling unit (e.g., the relationship of drug use between a parent and his or her child).

[9]In summary, this technique states that, if a dwelling unit is selected for the 2001 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.

Chapter 2: The Coordinated 5-Year Sample

As was previously mentioned, the sample design was simultaneously developed for the 1999-2003 NHSDAs. 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 NHSDA 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 achieve the following objectives:

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

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

- Constrain the maximum number of selections per dwelling unit at a specified value; specifically, we limit 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,...,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

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

The composite size measure for segment j was calculated as

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

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 NHSDA 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 as well as distributing the workload evenly among NHSDA 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.[10]

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 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.

---

[10]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 reasonable average workload for a single FI, leading us to designate the geographic strata within state 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.

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 Five-Year Sample	Unique Segments in Five-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
North Central						
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
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

(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	Number of Segments Selected	Number Selected for Five-Year Sample	Unique Segments in Five-Year Sample
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[11] and by the percent 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 NHSDA.

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 NHSDA,
02 = Segment for the 2002 NHSDA and used again in the 2003 NHSDA, and
03 = Segment for the 2003 NHSDA.

[11]Four categories are defined as: (1) MSA/low SES, (2) MSA/high SES, (3) Non-MSA/low SES, and (4) Non-MSA/high SES.

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

Order	Survey Year	Quarter	Panel	Variance Replicate
1	99A	1	1	1
2	Y00	1	15	1
3	X99B	1	8	2
4	Z01	1	22	2
5	02	1	5	1
6	Y99A	1	13	1
7	X03	1	12	2
8	Z99B	1	20	2
9	00	1	3	1
10	Y02	1	17	1
11	X01	1	10	2
12	Z03	1	24	2
13	01	1	4	2
14	Y03	1	18	2
15	X02	1	11	1
16	Z99A	1	19	1
17	99B	1	2	2
18	Y01	1	16	2
19	X00	1	9	1
20	Z02	1	23	1
21	03	1	6	2
22	Y99B	1	14	2
23	X99A	1	7	1
24	Z00	1	21	1
25	99A	2	1	1
26	Y00	2	15	1
27	X99B	2	8	2
28	Z01	2	22	2
29	02	2	5	1
30	Y99A	2	13	1
31	X03	2	12	2
32	Z99B	2	20	2
33	00	2	3	1
34	Y02	2	17	1
35	X01	2	10	2
36	Z03	2	24	2
37	01	2	4	2
38	Y03	2	18	2
39	X02	2	11	1
40	Z99A	2	19	1
41	99B	2	2	2
42	Y01	2	16	2
43	X00	2	9	1
44	Z02	2	23	1
45	03	2	6	2
46	Y99B	2	14	2
47	X99A	2	7	1
48	Z00	2	21	1
49	99A	3	1	1
50	Y00	3	15	1
51	X99B	3	8	2
52	Z01	3	22	2
53	02	3	5	1
54	Y99A	3	13	1
55	X03	3	12	2
56	Z99B	3	20	2
57	00	3	3	1
58	Y02	3	17	1
59	X01	3	10	2
60	Z03	3	24	2
61	01	3	4	2
62	Y03	3	18	2
63	X02	3	11	1
64	Z99A	3	19	1
65	99B	3	2	2
73	99A	4	1	1
74	Y00	4	15	1
75	X99B	4	8	2
76	Z01	4	22	2
77	02	4	5	1
78	Y99A	4	13	1
79	X03	4	12	2
80	Z99B	4	20	2
81	00	4	3	1
82	Y02	4	17	1
83	X01	4	10	2
84	Z03	4	24	2
85	01	4	4	2
86	Y03	4	18	2
87	X02	4	11	1
88	Z99A	4	19	1
89	99B	4	2	2

(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
66	Y01	3	16	2
67	X00	3	9	1
68	Z02	3	23	1
69	03	3	6	2
70	Y99B	3	14	2
71	X99A	3	7	1
72	Z00	3	21	1

Order	Survey Year	Quarter	Panel	Variance Replicate
90	Y01	4	16	2
91	X00	4	9	1
92	Z02	4	23	1
93	03	4	6	2
94	Y99B	4	14	2
95	X99A	4	7	1
96	Z00	4	21	1

X, Y, and Z denote extra segments for the corresponding NHSDA survey year. The 24 segments assigned to survey years not beginning with X, Y, and 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 five-year survey consists of panels 1 through 6 which correspond to segment suffices 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 NHSDAs, 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% 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 NHSDAs

Segment Suffix	1999 NHSDA	2000 NHSDA	2001 NHSDA	2002 NHSDA	2003 NHSDA
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.

This page intentionally left blank

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 2001 NHSDA 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} 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% of the m_{ha}'s. 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 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 calculation of the required dwelling units and the selection probabilities, sample size constraints[13] were applied to ensure adequate sample for overlapping designs and/or supplemental studies and to reduce 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.

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).

[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 2002 survey.

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% of the yearly estimate is used.

y_{ha} = Estimated number of persons in the target population in state h and age group a. The 2001 population is estimated using the 1990 Census data adjusted to the 2001 Claritas Population Projections.

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

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

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, 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 .99.

ε_h = State-specific, dwelling unit eligibility rate. Derived from 1999 NHSDA Quarter 4 and 2000 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 1999 NHSDA and quarters 1 through 3 of the 2000 NHSDA, the additive effects of state and age group on interviewer response were determined by taking the average interview response rate within each state.

γ_a = Expected number of persons within an age group per dwelling unit. Calculated using 1999 NHSDA Quarter 4 and 2000 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.

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 1999 NHSDA and quarters 1 through 3 of the 2000 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% for SAMHSA's 17 subpopulations of interest are:

- RSE = 3.40% for the total, national population.
- RSE = 5.00% 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% for the population within each of the four age groups for white (i.e., nonblack, non-Hispanic).
- RSE = 11.00% for the population within each of the four age groups for blacks (i.e., black, non-Hispanic).
- RSE = 11.00% for the population within each of the four age groups for Hispanics.

Note, one stratification feature that we used in previous NHSDA designs and was worth including in the design of the current NHSDA 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 NHSDA subpopulation of interest, as implied by the precision constraints, except for the age group 35 or older. As we have done with the NHSDA 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 NHSDA 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 NHSDA 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 *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, doctors) 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 for drinking (including hospitals, clinics, self-help groups, doctors). 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 2001 NHSDA design, are presented in Table 3.1. RSEs were based on an average prevalence rate of 10% for each measure.

Table 3.1 Expected Relative Standard Errors By Race/Ethnicity and Age Group: Main Sample

	Total Respondents					Hispanic Respondents				
Outcome Measure	**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	3.40	11.00	11.00	11.00	11.00	n/a

	Black Respondents					White Respondents				
Outcome Measure	**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 are not oversampled for the 2001 main study. However, consistent with previous NHSDAs, the 2001 NHSDA is designed to over-sample the younger age groups.

Among the 51 states, a required total sample size of 67,500 respondents is 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% of the annual sample) to the four quarters in ratios of 96%, 104%, 104%, and 96%. Only minor increases in unequal weighting result 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 is to determine the minimal number of dwelling units to select for each segment to meet the targeted sample sizes. In short, this step involves determining the sample size of the second-stage of selection. This sample size determination is 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/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 location of these dwelling units. 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 2001 NHSDA sample is summarized in Table 3.3. 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 is described

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
North Central									
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
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

(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
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 Segment and Dwelling Unit Summary

State	Total Segments	Subsegmented Segments	Listed Dwelling Units	Added Dwelling Units
Total Population	7,200	728	1,611,057	1,439
Alabama	96	10	22,180	15
Alaska	96	12	21,724	61
Arizona	96	12	21,100	6
Arkansas	96	4	21,296	11
California	384	33	84,866	51
Colorado	96	11	21,694	15
Connecticut	96	8	22,344	27
Delaware	96	21	22,583	21
District of Columbia	96	21	25,579	11
Florida	384	81	86,707	31
Georgia	96	14	23,420	15
Hawaii	96	18	21,831	46
Idaho	96	4	20,506	10
Illinois	384	28	83,969	36
Indiana	96	6	21,315	30
Iowa	96	7	19,749	11
Kansas	96	5	20,345	10
Kentucky	96	10	22,910	23
Louisiana	96	4	20,961	16
Maine	96	9	23,227	37
Maryland	96	23	22,707	16
Massachusetts	96	7	20,244	27
Michigan	384	29	87,413	74
Minnesota	96	4	20,998	27
Mississippi	96	3	20,336	15
Missouri	96	10	23,974	13
Montana	96	3	19,481	7

Table 3.3 Segment and Dwelling Unit Summary (continued)

State	Total Segments	Subsegmented Segments	Listed Dwelling Units	Added Dwelling Units
Nebraska	96	8	20,452	9
Nevada	96	28	21,884	6
New Hampshire	96	5	23,061	112
New Jersey	96	9	21,646	12
New Mexico	96	12	20,968	14
New York	384	53	88,261	105
North Carolina	96	10	21,890	4
North Dakota	96	3	22,605	38
Ohio	384	27	86,309	58
Oklahoma	96	5	21,568	28
Oregon	96	5	18,685	16
Pennsylvania	384	22	83,780	58
Rhode Island	96	3	21,646	59
South Carolina	96	11	22,281	7
South Dakota	96	4	19,876	10
Tennessee	96	6	20,962	21
Texas	384	63	87,000	28
Utah	96	8	19,404	12
Vermont	96	5	22,143	75
Virginia	96	16	21,191	44
Washington	96	15	20,224	14
West Virginia	96	5	22,199	12
Wisconsin	96	5	18,928	18
Wyoming	96	3	20,635	17

in Section 3.7 of this report. A detailed description of the counting and listing procedures is provided in the 2001 NHSDA: *Counting and Listing General Manual* (RTI, 2000).

3.3.2 Determining Dwelling Unit Sample Size

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

$$f_{ha} = P_{hj} * I_{hj} * \left(\frac{D_{hj}}{L_{hj}}\right) * S_{hja} * \varphi_h * \lambda_{ha} * \delta_{ha}. \quad (8)$$

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 ($f_{ha} / (\varphi_h * \lambda_{ha} * \delta_{ha})) = F_h$, known now as the driving age group, is set to the largest allowable selection probability (S_{hja}) of .99. D_{hj} is then computed as:

$$D_{hj} = \frac{f_{ha}}{(P_{hj} * I_{hj} * S_{hja} * \varphi_h * \lambda_{ha} * \delta_{ha})} * L_{hj} . \quad (9)$$

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

$$S_{hja} = \frac{f_{ha}}{P_{hj} * I_{hj} * (\frac{D_{hj}}{L_{hj}}) * \varphi_h * \lambda_{ha} * \delta_{ha}} \quad (10)$$

Having solved for D_{hj}, solve the selection probabilities for the remaining age groups. If L_{hj} equals 0 and subsequently D_{hj} equals 0, then all S_{hja} equals 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% overlap across segments and to allow SAMHSA to implement supplemental studies. An adequate remaining sample has two advantages: (1) for the 50% overlap design, this will provide better precision in year-to-year trend estimates because of the expected positive correlation between successive NHSDA 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 < ½*L_{hj}. Adjustments were made by adjusting the D_{hj} counts to equal the minimum of 100 or ½*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 ($m^*_{hja(\text{main})}$) was computed for the main study as follows:

$$m^*_{hja(\text{main})} = D^*_{hj} * \varepsilon_h * \varphi_h * \gamma_{ha} * S_{hja} * \lambda_{ha} * \delta_{\text{ha}}, \quad (11)$$

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:

$$5 / D^*_{hj} . \quad (12)$$

Similarly, the interview adjustment is computed as:

$$40 / m^*_{hja(\text{main})} . \quad (13)$$

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 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 is selected from the frame of counted and listed dwelling units for each segment (L_{hj}). The frame is ordered in the same manner as described in Section 3.3.1 and selection is 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 to control the amount of nonresponse without jeopardizing the validity of the study. Incidentally, a reserve sample of 10% 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% Quarter 1 sample was increased to the 105.6% level. In Quarter 1, the D_{hj} sample was allocated out to FI regions in the following release percentages:

Release 1: 100% of main sample (96% of quarterly sample);
Release 2: 100% of reserve sample (10% of main sample).

A summary of the quarterly sample sizes and percents released is provided in Table 3.4.

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

To guarantee that every dwelling unit has a chance of selection and to eliminate any bias associated with incomplete frames, the NHSDA implements a procedure called the half-open interval rule. This procedure requires 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 become part of the sample (added dwelling units). If the number of added dwelling units linked to any particular sample dwelling unit did not exceed six 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 2001 NHSDA sample is summarized in Table 3.3.

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* will address 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 3.4 Quarterly Sample Sizes and Percent Released

State	Quarter 1			Quarter 2		
	# Selected	# Released	Percent	# Selected	# Released	Percent
Total Population	52,686	47,870	91%	54,452	51,282	94%
Northeast						
Connecticut	786	711	90%	923	923	100%
Maine	882	803	91%	909	826	91%
Massachusetts	858	781	91%	852	775	91%
New Hampshire	816	742	91%	905	822	91%
New Jersey	759	687	91%	795	698	88%
New York	3,143	2,861	91%	3,239	3,236	100%
Pennsylvania	2,908	2,640	91%	2,980	2,700	91%
Rhode Island	754	687	91%	760	689	91%
Vermont	917	834	91%	807	735	91%
North Central						
Illinois	2,874	2,600	90%	2,984	2,718	91%
Indiana	728	661	91%	784	784	100%
Iowa	739	670	91%	703	637	91%
Kansas	611	555	91%	616	504	82%
Michigan	2,922	2,651	91%	2,945	2,674	91%
Minnesota	635	578	91%	647	589	91%
Missouri	839	757	90%	841	765	91%
Nebraska	605	551	91%	635	574	90%
North Dakota	664	605	91%	725	658	91%
Ohio	2,799	2,542	91%	2,804	2,804	100%
South Dakota	671	613	91%	637	637	100%
Wisconsin	702	641	91%	736	736	100%
South						
Alabama	743	671	90%	802	802	100%
Arkansas	764	697	91%	798	798	100%
Delaware	676	615	91%	647	590	91%
District of Columbia	1,346	1,214	90%	1,353	1,190	88%
Florida	3,008	2,731	91%	3,066	2,786	91%
Georgia	666	605	91%	687	687	100%
Kentucky	780	710	91%	705	641	91%
Louisiana	657	601	91%	706	641	91%
Maryland	584	533	91%	608	552	91%
Mississippi	600	546	91%	611	611	100%
North Carolina	651	590	91%	746	746	100%
Oklahoma	658	598	91%	687	687	100%
South Carolina	684	624	91%	726	726	100%
Tennessee	685	625	91%	703	703	100%
Texas	2,075	1,888	91%	2,233	2,233	100%
Virginia	698	636	91%	719	719	100%
West Virginia	761	691	91%	825	825	100%

(continued)

Table 3.4 Quarterly Sample Sizes and Percent Released (continued)

State	Quarter 1			Quarter 2		
	# Selected	# Released	Percent	# Selected	# Released	Percent
West						
Alaska	706	641	91%	712	712	100%
Arizona	715	650	91%	750	614	82%
California	2,625	2,382	91%	2,696	2,448	91%
Colorado	653	595	91%	720	658	91%
Hawaii	606	551	91%	664	664	100%
Idaho	611	554	91%	597	597	100%
Montana	711	647	91%	756	617	82%
Nevada	593	541	91%	614	556	91%
New Mexico	616	558	91%	667	667	100%
Oregon	628	568	90%	707	641	91%
Utah	370	342	92%	393	360	92%
Washington	679	619	91%	679	679	100%
Wyoming	525	477	91%	648	648	100%

State	Quarter 3			Quarter 4		
	# Selected	# Released	Percent	# Selected	# Released	Percent
Total Population	54,981	51,362	93%	54,153	51,507	95%
Northeast						
Connecticut	1,017	926	91%	1,119*	927*	83%*
Maine	839	764	91%	833	757	91%
Massachusetts	781	710	91%	648	648	100%
New Hampshire	934	850	91%	801	658	82%
New Jersey	772	702	91%	1,157*	1069*	92%*
New York	3,283	2,988	91%	5,387*	4679*	87%*
Pennsylvania	2,919	2,653	91%	3,000	2,998	100%
Rhode Island	761	691	91%	707	707	100%
Vermont	798	722	90%	702	640	91%
North Central						
Illinois	3,100	3,099	100%	2,663	2,647	99%
Indiana	710	710	100%	765	765	100%
Iowa	784	714	91%	582	479	82%
Kansas	589	537	91%	583	583	100%
Michigan	3,102	3,102	100%	3,156	3,156	100%
Minnesota	667	605	91%	530	436	82%
Missouri	786	717	91%	783	712	91%
Nebraska	643	585	91%	549	446	81%
North Dakota	700	635	91%	749	679	91%
Ohio	2,751	2,502	91%	2,449	2,449	100%
South Dakota	666	605	91%	538	495	92%
Wisconsin	758	689	91%	643	584	91%

*Sample sizes include the NYC supplement.

(continued)

Table 3.4 Quarterly Sample Sizes and Percent Released (continued)

State	Quarter 3			Quarter 4		
	# Selected	# Released	Percent	# Selected	# Released	Percent
South						
Alabama	726	659	91%	603	549	91%
Arkansas	844	844	100%	675	675	100%
Delaware	677	616	91%	561	561	100%
District of Columbia	1,350	1,226	91%	1,183	1,181	100%
Florida	3,231	2,936	91%	2,760	2,760	100%
Georgia	808	733	91%	621	565	91%
Kentucky	714	648	91%	659	659	100%
Louisiana	637	580	91%	544	496	91%
Maryland	603	603	100%	508	507	100%
Mississippi	678	678	100%	760	760	100%
North Carolina	818	818	100%	690	690	100%
Oklahoma	691	629	91%	646	583	90%
South Carolina	772	772	100%	793	793	100%
Tennessee	725	658	91%	914	827	90%
Texas	2,198	1,995	91%	2,247	2,247	100%
Virginia	739	739	100%	844	689	82%
West Virginia	889	889	100%	783	783	100%
West						
Alaska	824	747	91%	693	693	100%
Arizona	659	659	100%	765	699	91%
California	2,756	2,506	91%	2,586	2,358	91%
Colorado	664	603	91%	621	620	100%
Hawaii	686	686	100%	572	572	100%
Idaho	638	621	97%	630	575	91%
Montana	689	689	100%	739	739	100%
Nevada	610	550	90%	680	680	100%
New Mexico	567	511	90%	532	532	100%
Oregon	626	626	100%	666	666	100%
Utah	383	350	91%	326	326	100%
Washington	722	653	90%	589	589	100%
Wyoming	697	632	91%	619	619	100%

Quarter 2

Design: An additional 10% sample was added to the 104% quarterly sample to allow for supplemental releases where needed. Thus, the total Quarter 2 sample was increased to the 114.4% level.

Procedural: In order to predict state response rates more accurately, the most current four quarters of NHSDA data were used in the computation of state-specific yield and response rates. Thus, data from quarters 1 through 4 of the 2000 NHSDA were used to compute average yields, dwelling unit eligibility, screening response, and interviewer response rates.

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

Release 1: 82% of entire sample (90/110, main sample + 10% reserve);
Release 2: 9% of entire sample (10/110, main sample + 10% reserve);
Release 3: 9% of entire sample (10/110, main sample + 10% 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% sample was then adjusted by this amount. An additional 10% sample was also included, therefore bringing the total Quarter 3 adjusted sample to the 114.4% level.

Procedural: Data from quarters 2 through 4 of the 2000 NHSDA and Quarter 1 of the 2001 NHSDA 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: 82% of entire sample (90/110, main sample + 10% reserve);
Release 2: 9% of entire sample (10/110, main sample + 10% reserve);
Release 3: 9% of entire sample (10/110, main sample + 10% reserve).

Quarter 4

Design: The state and age 96% quarterly sample sizes were adjusted in order to meet the yearly targets. An additional 10% sample was also included, therefore bringing the total Quarter 4 adjusted sample to the 105.6% level.

Procedural: Data from quarters 3 and 4 of the 2000 NHSDA and quarters 1 and 2 of the 2001 NHSDA 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:
Release 1: 82% of entire sample (90/110, main sample + 10% reserve);
Release 2: 9% of entire sample (10/110, main sample + 10% reserve);
Release 3: 9% of entire sample (10/110, main sample + 10% reserve).

Quarter 4 New York City Supplement

Design: Following the September 11 attacks, a supplemental sample was added to the area surrounding New York City. Specifically, the FI regions that were supplemented were NY18 through NY48, NJ03 through NJ12, and CT01 through CT08. Dwelling unit sample sizes were computed for New York, New Jersey, and Connecticut based on state targets of 600, 150, and 150, respectively.

Procedural: Similar to the Quarter 4 sample allocation, data from quarters 3 and 4 of the 2000 NHSDA and quarters 1 and 2 of the 2001 NHSDA were used to compute state-specific average yields, dwelling unit eligibility, screening response, and interviewer response rates.

Dwelling Unit Selection: The full supplement was released, thus no sample partitions were needed.

3.9 Sample Weighting Procedures

At the conclusion of data collection for the last quarter, sample weights were constructed for each quarter of the state-level study that reflected the various stages of sampling described earlier in Section 1.2.2. 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: Selection of dwelling unit.
Three possible adjustments exist with this stage of selection:
(1) Subsegmentation inflation – by-product of counting and listing,
(2) Added dwelling unit – results from the half-open interval rule, and
(3) Release adjustment.

Stage 3: Selection of person within a dwelling unit.

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

1. *Nonresponse Adjustment at the Dwelling Unit Level.* This is 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 3.5.

2. *Dwelling Unit Level Post-Stratification.* This involves using screener data of demographic information (e.g., age, race, gender, etc.). Dwelling unit weights were adjusted to the

intercensal population estimates obtained from the U.S. Bureau of the Census' National Estimates and Projections Branch. In short, explanatory variables used during modeling consisted of counts of eligible persons within each dwelling unit that fell into the various demographic categories. Subsequently, these counts multiplied by the newly adjusted dwelling unit weight and summed across all dwelling units for various domains add to the Census control totals. This adjustment is necessary for the proper calculation of pairwise weights and allows us to achieve greater precision in subsequent adjustments. Screener level potential variables are listed in Table 3.6.

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 result in an unsatisfactory unequal weighting effect (i.e., variance between the dwelling unit level weights is too high as well as high frequency of extreme weights), then high weights were properly adjusted. This was implemented by doing another weighting calibration. The control totals are the dwelling unit level post-stratified weights, and the same explanatory variables as in dwelling unit level post-stratification 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 Post-Stratification to Roster Data.* This step utilizes control totals derived from the dwelling unit roster that are equal to the Census estimates. This assists in bias reduction and improves precision by taking advantage of the properties of a larger sample size. 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 variables for this adjustment are a combination of the variables presented in Table 3.7.

5. *Person Level Nonresponse Adjustment.* This adjustment allows for the correction of weights resulting from the failure of selected sample persons to complete the interview. Respondent sample weights were adjusted to the total weight sum of all selected persons, adjusted for post-stratification to the eligible roster of persons. Again, demographic information used in modeling is based solely on screener information. Potential variables for this adjustment are a combination of the variables presented in Table 3.7.

6. *Person Level Post-Stratification.* This step is to adjust the final person sample weights to the Census Bureau's estimates. These are the same outside control totals used in the second adjustment. However, explanatory variables for this adjustment are based on questionnaire data, not screener data as in adjustment 2. Data can differ between the two sources. Variables used in modeling are presented in Table 3.6.

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 3.5 Definitions of Levels for Potential Variables for Dwelling Unit Nonresponse Adjustment

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

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

Percent of Segments That are Black (% Black)
- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Percent of Segments That are Hispanic (% Hispanic)
- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Population Density
- 1: MSA $\geq$ 1,000,000
- 2: MSA less than 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

States

Interactions among the main effect variables are also considered.

Table 3.6 Definitions of Levels for Potential Variables for Dwelling Unit Post-Stratification and Respondent Post-Stratification at the Person Level

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

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

State

Interactions among the main effect variables are also considered.

Table 3.7 Definitions of Levels for Potential Variables for Selected Person Post-Stratification and Nonresponse Adjustment

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

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

Percent of Segments That are Black (% Black)
- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Percent of Segments That are Hispanic (% Hispanic)
- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Population Density
- 1: MSA $\geq$ 1,000,000
- 2: MSA less than 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

States

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

Gender
- 1: Male
- 2: Female

(continued)

Table 3.7 Definitions of Levels for Potential Variables for Selected Person Post-Stratification and Nonresponse Adjustment (continued)

Hispanicity
1: Hispanic
2: Non-Hispanic
Race
1: White
2: Black
3: Indian / Native American
4: Asian
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 adjustments for the 2001 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 Region Division definitions as shown in Table 3.8. 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 3.9 presents a flowchart of steps used in the weighting process and Table 3.10 displays all individual weight components.

Table 3.8 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 3.9 Flowchart of Sample Weighting Steps

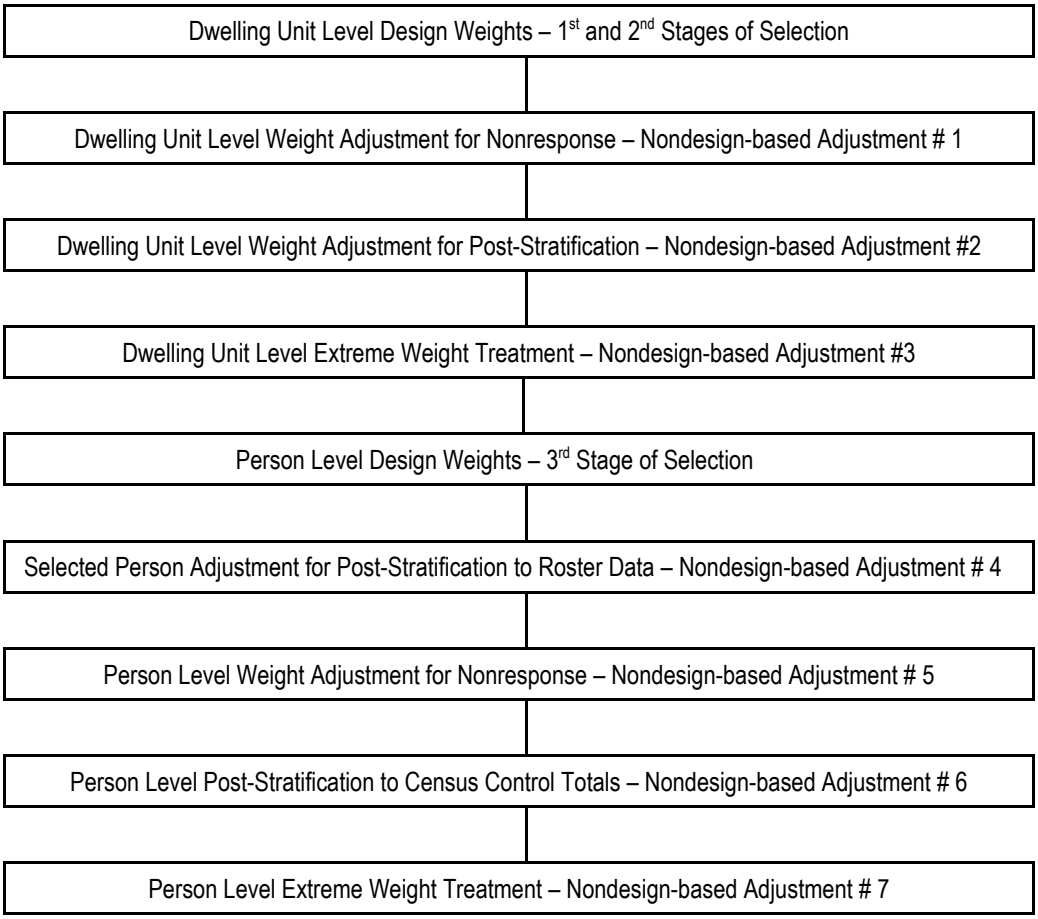

Table 3.10 Sample Weight Components

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 Dwelling Unit
6 Dwelling Unit Release Adjustment
7 Dwelling Unit Nonresponse Adjustment
8 Dwelling Unit Post-Stratification
9 Dwelling Unit Extreme Weight Treatment
Person Level Design Weight Components
10 Inverse Probability of Selecting a Person Within a Dwelling Unit
11 Roster Adjustment
12 Person Level Nonresponse Adjustment
13 Person Level Post-Stratification Adjustment
14 Person Level Extreme Weight Treatment

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 2001 NHSDA* (Chen, 2003).

This page intentionally left blank

References

Chen, P., Gordek, H., Murtha, C., Yu, D., Westlake, M., & Singh, A. (2003). Person-level sampling weight calibration for the 2001 NHSDA. In *2001 National Household Survey on Drug Abuse: 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. (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. (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 he 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. (1979). Sequential sample selection methods. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 401-406). Alexandria, VA: American Statistical Association.

Cochran, W. (1977). *Sampling techniques* (3rd ed.) (pp. 261-263). New York: John Wiley & Sons.

Eyerman, J. & Bowman, K. (2002). 2001 NHSDA Incentive Experiment: Combined Quarter 1 and Quarter 2 Analysis. Prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-98-9008. Research Triangle Park, NC: RTI.

Kish, L. (1965). *Survey sampling* (pp. 313-315). New York: John Wiley & Sons.

RTI (2000). 2001 National Household Survey on Drug Abuse: 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: Author.