2003 NATIONAL SURVEY ON DRUG USE AND HEALTH

SAMPLE DESIGN REPORT

Prepared for the 2003 Methodological Resource Book

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

Authors:

Project Director: Thomas G. Virag

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

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, NC 27709

January 2005

2003 NATIONAL SURVEY ON DRUG USE AND HEALTH

SAMPLE DESIGN REPORT

Prepared for the 2003 Methodological Resource Book

Contract No. 283-98-9008
RTI Project No. 7190

Phase V, Deliverable No. 10

Authors:

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

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, MD 20857

Prepared by:

RTI International
Research Triangle Park, NC 27709

January 2005

Acknowledgments

Table of Contents

List of Tables

Chapter 1: Overview

1.1 Target Population

The respondent universe for the 2003 National Survey on Drug Use and Health[1] (NSDUH) was the civilian, noninstitutionalized population aged 12 years or older residing within the United States and the District of Columbia. Consistent with the NSDUH designs since 1991, the 2003 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 2003 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 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: persons aged 12 to 17, persons aged 18 to 25, and persons aged 26 or older. This large sample size allows SAMHSA to continue reporting precise demographic subgroups at the national level without needing to oversample specially targeted demographics, as required in the past. This large sample is referred to as the "main sample" or the "CAI sample." The achieved sample for the 2003 CAI sample was 67,784 persons.

Beginning with the 2002 NSDUH and continuing with the 2003 NSDUH, survey respondents were given a $30 incentive payment for participation. As expected, the incentive had the effect of increasing response rates and requiring fewer selected households than previous surveys.

An additional design change was made in 2002 and continued in 2003. A new pair sampling strategy was implemented that increased the number of pairs selected in dwelling units (DUs) with older persons on the roster (Chromy & Penne, 2002). With the increase in the number of pairs came a moderate decrease in the response rate for older persons.

1.3 5-Year Design

A coordinated 5-year sample design was developed. The 2003 main sample is a subsample of the 5-year sample. Although there is no planned overlap with the 1998 sample, a

---

[1] This report presents information from the 2003 National Survey on Drug Use and Health (NSDUH). Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

coordinated design for 1999-2003 facilitated 50 percent 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 2003 survey. This sample size was considered adequate to support direct State estimates. The remaining 43 States[3] had samples designed to yield 900 respondents per State in the 2003 survey. In these 43 States, adequate data were available to support reliable State estimates based on small area estimation (SAE) methodology.

1.4 Stratification and First-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.

For the first stage of sampling, each of the FI regions was partitioned into noncompact clusters[4] of DUs by aggregating adjacent Census blocks. Consistent with the terminology used in previous NSDUHs, these geographic clusters of blocks are referred to as segments. A sample DU in NSDUH refers to either a housing unit or a group-quarters listing unit, such as a dormitory room or a shelter bed. 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 DUs[5] on average. In prior years, the average minimum segment DU 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

---

[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.

[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] DU 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).

area/socioeconomic status) indicator[6] and by the percentage of the population that is non-Hispanic and white. From this well-ordered sample frame, 96 segments[7] 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 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 2003 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 DUs needed in each segment to meet the targeted sample sizes for all age groups. Thus, listing unit sample sizes for the segment were determined using the age group with the largest sampling rate, which we refer to as the "driving" age group. Using 1990 decennial 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,[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 2001, 2002, and 2003 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 DUs found in the field during counting and listing activities. The product represents the segment's listing unit sample size.

Some constraints were put on the listing unit sample sizes. For example, to ensure adequate samples for 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.

---

[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.

[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 the 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.

Using a random start point and interval-based (systematic) selection, the actual listing units were selected from the segment frame. After DU selections were made, an interviewer visited each selected DU to obtain a roster of all persons residing in the DU. As in previous years, during the data collection period, if an interviewer encountered any new DU in a segment or found a DU that was missed during the original counting and listing activities, then the new or missed dwellings were selected into the 2003 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 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 2003, one feature that was included in the design was that any two survey-eligible persons within a DU had some chance of being selected (i.e., all survey-eligible pairs of persons had some nonzero chance of being selected). This design feature was of interest to NSDUH researchers because, for example, it allows analysts to examine how the drug use propensity of one individual in a family relates to the drug use propensity of other family members residing in the same DU (e.g., the relationship of drug use between a parent and his or her child).

---

[10] In summary, this technique states that, if a dwelling unit is selected for the 2003 study and an interviewer observes any new or missed DUs between the selected DU and the DU appearing immediately after the selection on the counting and listing form, then all new or missed dwellings falling in this interval will be selected. If a large number of new or missed DUs are encountered (generally greater than 10), then a sample of the new or missing DUs will be selected. For more information, please refer to Appendix C.

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 (DUs) that would achieve the following objectives:

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

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

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

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

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

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

- Simplify the size measure data requirements so that 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 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 DUs 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, 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 the 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 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

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 using clustered data analysis software packages such as SUDAAN (RTI, 2004). 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 2 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 (DU) sample sizes will be discussed. The within-DU age group-specific selection probabilities for the 2003 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 DUs necessary to achieve the various design requirements. In summary, this three-step optimization procedure proceeded as follows:

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

2. Using the m_{ha} determined from Step 1, the next step was to determine the optimal number of selected dwelling (D_{hj}) units (i.e., 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 DUs 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 DU.[12] A detailed discussion of the final step is given in Section 3.4. After calculating the required DUs and the selection probabilities, we applied sample size constraints[13] 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] Even though the 1999 survey was the last in the 5-year sample, constraints were applied to the required DU sample sizes under the assumption that some segments might be revisited in the 2004 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 DU sample size, approximately 25 percent of the yearly estimate is used.

y_{ha} = Estimated number of persons in the target population in State h and age group a. The 2003 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 2003 population (y^{*}_{ha}) is computed using the original formula and this time allowing x to be 13. Finally, the 2003 population is adjusted by the ratio of estimated eligible listed DUs to the Claritas DU counts (U_{hj}). This adjustment factor considers the number of added DUs expected to be obtained through the half-open interval rule (1.01) and the probability of a DU being eligible (ε_s), both determined via historic data. The coefficient adjustment of 1.01 is estimated using historical data and is the proportion of all screened DUs (includes added DUs) over the original total of selected DUs (excluding added DUs). So, $y_{ha} = \{[1.01 * \varepsilon_s * L_{hj} * (1/I_{hj}) / U_{hj}]\} * y^{*}_{ha}$, 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 = $Max\{f_{ha} / (\phi_h * \lambda_{ha} * \delta_{ha}), a = 1\text{-}5\}$.

P_{hj} = Inverse of the segment selection probability. DU sample sizes are computed on a quarterly basis, and segments are selected on a yearly basis. Because 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 to 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 DUs 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 DUs to select for screening in segment j to meet the targeted sample sizes for all age groups.

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

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

ε_h = State-specific DU eligibility rate. Derived from 2001 NHSDA Quarter 4 and 2002 NSDUH 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 DU eligibility rate (ε_h).

λ_{ha} = State- and age group-specific interview response rate. Using data from Quarter 4 of the 2001 NHSDA and Quarters 1 through 3 of the 2002 NSDUH, 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 DU. Calculated using 2001 NHSDA Quarter 4 and 2002 NSDUH Quarters 1 through 3 data by dividing the weighted total number of rostered persons in an age group by the weighted total number of complete screened DUs by State.

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

inflate the number of required DUs 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 needed in each of the 255 domains to minimize the costs associated with data collection, subject to multiple precision requirements established by SAMHSA. In summary, the precision requirements 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 included 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. Because 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 because these two measures reflect different types of drug abuse.

Drug or Alcohol Dependence

6. *Dependent on Illicit Drugs in the Past Year.* Dependent on the same drugs listed in class 3, *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 2003 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

	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	3.00	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 were not oversampled for the 2003 main study. However, consistent with previous NSDUHs, the 2003 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 DUs 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 DU frame is constructed is called counting and listing. In summary, a certified lister visits the selected area and lists a detailed and accurate address (or description, if no address is available) for each DU within the segment boundaries. The lister is given a series of maps on which to mark the locations of these DUs. The number of map pages per State and the average number of map pages per segment are summarized in Table 3.3. The resulting list of DUs is entered into a database and serves as the frame from which the second-stage sample is drawn.

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

that were subsegmented in the 2003 NSDUH sample is summarized in Table 3.4. For more information on the subsegmenting procedures, see Appendix B.

A minor error was discovered with some of the 2003 NSDUH segments that were subsegmented in-house. These segments were subsegmented prior to the segment kit envelope being ready, and the subsegmenting DU counts were never transferred to the front of the envelope. As a result, the counts were not entered into the database for subsequent stages of sample selection. A total of eight 2003 segments (one Quarter 1 segment, two Quarter 2 segments , four Quarter 3 segments, and one Quarter 4 segment) were effected. The correct counts were used when computing the design-based weights. However, because this information was not used during sample allocation, a small amount of unequal weighting may have been created. No other survey years were affected by this problem.

During counting and listing, the lister moves about the segment in a prescribed fashion called the "continuous path of travel." The lister attempts to move in a clockwise fashion, makes each possible right turn, makes U-turns at segment boundaries, and doesn't break street sections. Following these defined rules and always looking for DUs on the right-hand side of the street, the lister minimizes the chance of not listing a DU within the segment. Also, using a defined path of travel makes it easier for the FI assigned to the segment to locate the sampled DUs. Finally, the continuous path of travel lays the groundwork for the half-open interval procedure for recovering missed DUs, as described in Section 3.7 of this report. A detailed description of the counting and listing procedures is provided in the *2003 NSDUH: Counting and listing general manual* (RTI, 2002).

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	38,990	5.4
Alabama	96	590	6.1
Alaska	96	493	5.1
Arizona	96	528	5.5
Arkansas	96	753	7.8
California	384	1,287	3.4
Colorado	96	430	4.5
Connecticut	96	363	3.8
Delaware	96	407	4.2
District of Columbia	96	311	3.2
Florida	384	1,896	4.9
Georgia	96	463	4.8
Hawaii	96	324	3.4
Idaho	96	672	7.0
Illinois	384	1,799	4.7
Indiana	96	579	6.0
Iowa	96	765	8.0
Kansas	96	632	6.6
Kentucky	96	494	5.1
Louisiana	96	553	5.8
Maine	96	548	5.7
Maryland	96	376	3.9
Massachusetts	96	412	4.3
Michigan	384	1,909	5.0
Minnesota	96	562	5.9
Mississippi	96	659	6.9
Missouri	96	565	5.9
Montana	96	916	9.5
Nebraska	96	676	7.0
Nevada	96	411	4.3
New Hampshire	96	497	5.2
New Jersey	96	387	4.0
New Mexico	96	742	7.7
New York	384	1,569	4.1
North Carolina	96	568	5.9
North Dakota	96	1,331	13.9
Ohio	384	1,731	4.5
Oklahoma	96	682	7.1
Oregon	96	523	5.4
Pennsylvania	384	2,151	5.6
Rhode Island	96	477	5.0
South Carolina	96	579	6.0
South Dakota	96	995	10.4
Tennessee	96	491	5.1
Texas	384	1,774	4.6

(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	519	5.4
Vermont	96	570	5.9
Virginia	96	458	4.8
Washington	96	383	4.0
West Virginia	96	579	6.0
Wisconsin	96	593	6.2
Wyoming	96	1,018	10.6

Table 3.4   Segment and Dwelling Unit Summary

State	Total Segments	Subsegmented Segments	Listed Dwelling Units	Added Dwelling Units
Total Population	7,200	890	1,617,741	1,695
Alabama	96	5	21,063	12
Alaska	96	13	22,292	62
Arizona	96	12	20,585	33
Arkansas	96	4	20,192	7
California	384	40	83,569	36
Colorado	96	10	20,398	20
Connecticut	96	9	24,113	50
Delaware	96	13	24,066	39
District of Columbia	96	27	26,032	34
Florida	384	80	87,641	69
Georgia	96	25	20,850	21
Hawaii	96	26	22,019	40
Idaho	96	11	17,978	34
Illinois	384	42	83,415	70
Indiana	96	4	21,426	18
Iowa	96	3	19,616	6
Kansas	96	12	21,693	14
Kentucky	96	7	22,464	13
Louisiana	96	4	25,121	8
Maine	96	2	23,768	53
Maryland	96	18	22,031	18
Massachusetts	96	6	22,619	27
Michigan	384	32	89,500	76
Minnesota	96	11	19,895	14
Mississippi	96	5	20,718	7
Missouri	96	9	21,327	19
Montana	96	19	19,592	23
Nebraska	96	8	19,337	14
Nevada	96	23	20,774	3
New Hampshire	96	14	23,206	83
New Jersey	96	12	20,364	20
New Mexico	96	19	20,597	26

(continued)

Table 3.4 Segment and Dwelling Unit Summary (continued)

State	Total Segments	Subsegmented Segments	Listed Dwelling Units	Added Dwelling Units
New York	384	61	92,092	115
North Carolina	96	9	23,055	18
North Dakota	96	23	20,426	27
Ohio	384	31	85,034	46
Oklahoma	96	8	22,524	30
Oregon	96	12	20,362	14
Pennsylvania	384	30	84,484	60
Rhode Island	96	8	22,297	51
South Carolina	96	9	21,996	7
South Dakota	96	17	19,927	23
Tennessee	96	11	22,071	7
Texas	384	68	88,048	79
Utah	96	10	20,675	38
Vermont	96	6	22,551	98
Virginia	96	17	23,233	18
Washington	96	19	20,498	30
West Virginia	96	9	20,828	25
Wisconsin	96	10	19,757	28
Wyoming	96	7	19,622	12

3.3.2 Determining Dwelling Unit Sample Size

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

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

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

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

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

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

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

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

A major area of interest for the survey is to ensure that an adequate sample of eligible DUs remain within each segment. This sample surplus is needed to provide for the yearly 50-percent overlap across segments[14] 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 < ½L_{hj}. Adjustments were made by adjusting the D_{hj} counts to equal the minimum of 100 or ½L_{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:

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

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

$$5 / D^*_{hj} . \tag{8}$$

Similarly, the interview adjustment is computed as

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

---

[14] While the 2003 survey was the last in the 5-year design, constraints were put on the DU sample under the assumption that a 50-percent overlap with the 2004 survey was a possibility.

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

Both constraints 1 and 3 reduce the second-stage sample, which could in turn 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 DU sample is iterative until the expected person sample sizes are met.

3.6   Dwelling Unit Selection and Release Partitioning

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

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), and
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 DU 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 DU and between that DU and the next listed DU for any unlisted units. When found in these specific locations, the unlisted units became part of the sample (added DUs). If the number of added DUs linked to any particular sample DU did not exceed five, or if the number for the entire segment was less than or equal to 10, the FI was instructed to consider these DUs as part of their assignment. If either of these limits was exceeded, special subsampling procedures were implemented, as described in Appendix C. The number of added DUs in the 2003 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, because 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	44,983	41,022	91%	48,332	43,009	89%
Northeast						
Connecticut	773	702	91%	698	602	86%
Maine	751	684	91%	837	761	91%
Massachusetts	606	550	91%	679	525	77%
New Hampshire	640	587	92%	608	442	73%
New Jersey	679	618	91%	685	623	91%
New York	2,587	2,361	91%	2,732	2,607	95%
Pennsylvania	2,523	2,253	89%	2,774	2,646	95%
Rhode Island	620	561	90%	628	514	82%
Vermont	638	580	91%	633	633	100%
Midwest						
Illinois	2,223	2,021	91%	2,398	2,183	91%
Indiana	567	514	91%	565	516	91%
Iowa	569	517	91%	624	541	87%
Kansas	481	438	91%	540	540	100%
Michigan	2,405	2,189	91%	2,618	2,147	82%
Minnesota	578	523	90%	651	592	91%
Missouri	659	598	91%	696	573	82%
Nebraska	530	482	91%	574	521	91%
North Dakota	581	527	91%	574	522	91%
Ohio	2,306	2,096	91%	2,546	2,312	91%
South Dakota	507	507	100%	592	565	95%
Wisconsin	557	505	91%	632	545	86%

(continued)

Table 3.5   Quarterly Sample Sizes and Percentages Released (continued)

State	Quarter 1			Quarter 2		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
Total Population	44,983	41,022	91%	48,332	43,009	89%
South						
Alabama	645	585	91%	626	538	86%
Arkansas	572	522	91%	668	606	91%
Delaware	641	641	100%	648	620	96%
District of Columbia	908	822	91%	978	843	86%
Florida	2,851	2,592	91%	3,188	2,604	82%
Georgia	516	472	91%	585	531	91%
Kentucky	667	606	91%	724	561	77%
Louisiana	592	537	91%	626	568	91%
Maryland	431	377	87%	475	474	100%
Mississippi	537	488	91%	662	600	91%
North Carolina	564	513	91%	616	531	86%
Oklahoma	625	568	91%	653	653	100%
South Carolina	637	581	91%	640	521	81%
Tennessee	611	553	91%	641	611	95%
Texas	2,001	1,817	91%	2,130	2,130	100%
Virginia	584	533	91%	596	460	77%
West Virginia	736	672	91%	788	720	91%
West						
Alaska	606	548	90%	660	574	87%
Arizona	638	579	91%	657	595	91%
California	1,997	1,809	91%	2,167	1,961	90%
Colorado	546	546	100%	581	529	91%
Hawaii	600	546	91%	627	546	87%
Idaho	486	442	91%	564	539	96%
Montana	683	622	91%	669	576	86%
Nevada	636	579	91%	639	495	77%
New Mexico	554	500	90%	724	691	95%
Oregon	666	603	91%	684	497	73%
Utah	385	385	100%	416	378	91%
Washington	691	628	91%	778	567	73%
Wyoming	597	543	91%	638	580	91%

(continued)

29

Table 3.5   Quarterly Sample Sizes and Percentages Released (continued)

State	Quarter 3			Quarter 4		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
Total Population	49,110	44,608	91%	44,424	40,367	91%
Northeast						
Connecticut	719	686	95%	640	583	91%
Maine	767	696	91%	696	633	91%
Massachusetts	741	641	87%	671	670	100%
New Hampshire	553	427	77%	526	476	90%
New Jersey	735	704	96%	661	599	91%
New York	2,870	2,607	91%	2,509	2,283	91%
Pennsylvania	2,725	2,601	95%	2,615	2,261	86%
Rhode Island	684	684	100%	580	445	77%
Vermont	662	662	100%	665	665	100%
Midwest						
Illinois	2,770	2,770	100%	2,332	2,119	91%
Indiana	583	531	91%	513	467	91%
Iowa	563	488	87%	506	483	95%
Kansas	522	500	96%	550	550	100%
Michigan	2,623	2,265	86%	2,562	2,323	91%
Minnesota	607	495	82%	559	405	72%
Missouri	746	673	90%	661	632	96%
Nebraska	548	525	96%	454	454	100%
North Dakota	602	519	86%	551	477	87%
Ohio	2,456	2,234	91%	2,292	2,186	95%
South Dakota	542	542	100%	517	517	100%
Wisconsin	637	578	91%	626	626	100%
South						
Alabama	601	491	82%	542	445	82%
Arkansas	654	535	82%	588	588	100%
Delaware	748	612	82%	657	507	77%
District of Columbia	1,001	949	95%	1,044	1,044	100%
Florida	3,199	2,750	86%	2,827	2,436	86%
Georgia	624	568	91%	575	520	90%
Kentucky	629	515	82%	600	571	95%
Louisiana	587	481	82%	538	490	91%
Maryland	540	540	100%	475	475	100%
Mississippi	688	591	86%	625	510	82%
North Carolina	670	637	95%	620	540	87%
Oklahoma	684	652	95%	607	552	91%
South Carolina	638	608	95%	565	488	86%
Tennessee	622	592	95%	527	527	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	49,110	44,608	91%	44,424	40,367	91%
Texas	2,193	2,092	95%	1,955	1,782	91%
Virginia	692	692	100%	601	465	77%
West Virginia	865	785	91%	792	721	91%
West						
Alaska	638	638	100%	568	492	87%
Arizona	691	506	73%	491	446	91%
California	2,274	2,072	91%	1,990	1,809	91%
Colorado	659	629	95%	551	501	91%
Hawaii	690	569	82%	616	558	91%
Idaho	546	471	86%	512	512	100%
Montana	640	611	95%	577	552	96%
Nevada	630	516	82%	525	478	91%
New Mexico	690	535	78%	560	508	91%
Oregon	659	567	86%	580	421	73%
Utah	437	419	96%	426	403	95%
Washington	672	617	92%	665	633	95%
Wyoming	594	540	91%	539	539	100%

Quarter 2

Design:

An additional 10 percent reserve sample was added to the 104 percent quarterly sample to allow for supplemental releases where needed. Thus, the total Quarter 2 sample was increased to the 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 2002 NSDUH[15] were used to compute average yields, DU eligibility, screening response, and interviewer response rates.

Dwelling Unit Selection:

The Quarter 2 D_{hj} sample was partitioned into the following release percentages:

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

---

[15] The fraudulent cases in NM/NV/MS were dropped from the 2002 file because this experience was not representative of what we expected in 2003.

31

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); and
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 2002 NSDUH[16] and Quarter 1 of the 2003 NSDUH were used to compute State-specific average yields, DU eligibility, screening response, and interviewer response rates. In addition, the maximum-of-two rule adjustment was updated using all four quarters of data from the 2002 NSDUH.

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); and
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 Quarters 1 and 2 and the projected number of completes from

---

[16] The fraudulent cases in NM/NV/MS were dropped from the 2002 file because this experience was not representative of what we expected in 2003.

Quarter 3. An additional 10 percent sample was also included, bringing the total Quarter 4 adjusted sample to the 105.6 percent level.

Procedural: Data from Quarters 3 and 4 of the 2002 NSDUH[17] and Quarters 1 and 2 of the 2003 NSDUH were used to compute State-specific average yields, DU eligibility, screening response, and interviewer response rates.

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

Release 1: 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); and
Release 5: 9 percent of entire sample (10/110, main sample + 10 percent reserve).

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: Selection of DU.
Three possible adjustments exist with this stage of selection:
(1)   Subsegmentation inflation: By-product of counting and listing,
(2)   Added DU: Results from the half-open interval rule when subsampling is needed, and
(3)   Release adjustment.

---

[17] The fraudulent cases in NM/NV/MS were dropped from the 2002 file because this experience was not representative of what we expected in 2003.

Stage 3:   Selection of person within a DU.

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:

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 DU 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.). DU 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 DU that fell into the various demographic categories. Consequently, these counts, multiplied by the newly adjusted DU weight and summed across all DUs 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 DU roster that were already post-stratified 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 DU level and account for third-stage sampling) were adjusted to the DU weight sums of all eligible rostered persons. Any demographic information used in modeling is based solely on screener information, because this is the only information available for all rostered persons. Potential explanatory variables for this adjustment are a combination of the variables presented in Table 3.8.

5. *Person-Level Nonresponse Adjustment.* This adjustment 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 screener data as in adjustments 2, 4, and 5. Potential explanatory variables used in modeling are presented in Table 3.7.

7. *Extreme Weight Treatment at the Person Level.* This was implemented in the same manner as described 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+[18]

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.

---

[18] For person-level respondent 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 Person-Level 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 2003 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 was used for the 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

Dwelling Unit-Level Design Weights–1st and 2nd Stages of Selection

Dwelling Unit-Level Weight Adjustment for Nonresponse–Nondesign-Based Adjustment # 1

Dwelling Unit-Level Weight Adjustment for Poststratification–Nondesign-Based Adjustment #2

Dwelling Unit-Level Extreme Weight Treatment–Nondesign-Based Adjustment #3

Person-Level Design Weights–3rd Stage of Selection

Selected Person Adjustment for Poststratification to Roster Data–Nondesign-Based Adjustment # 4

Person-Level Weight Adjustment for Nonresponse–Nondesign-Based Adjustment # 5

Person-Level Poststratification to Census Control Totals–Nondesign-Based Adjustment # 6

Person-Level Extreme Weight Treatment–Nondesign-Based Adjustment # 7

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 2003 NSDUH* (Chen et al., 2004).

References

Chen, P., Gordek, H., Dai, L., Singh, A., Shi, W., & Westlake, M. (2004). Person-level sampling weight calibration for the 2003 NSDUH. In *2003 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.

Research Triangle Institute (RTI). (2002). 2003 *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.

Research Triangle Institute (RTI). (2004). SUDAAN (Release 9.0) [Computer Software]. Research Triangle Park, NC: RTI.

Appendix A

1999-2003 NHSDA/NSDUH Field Interviewer Regions
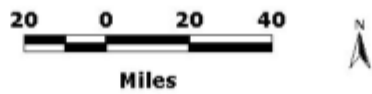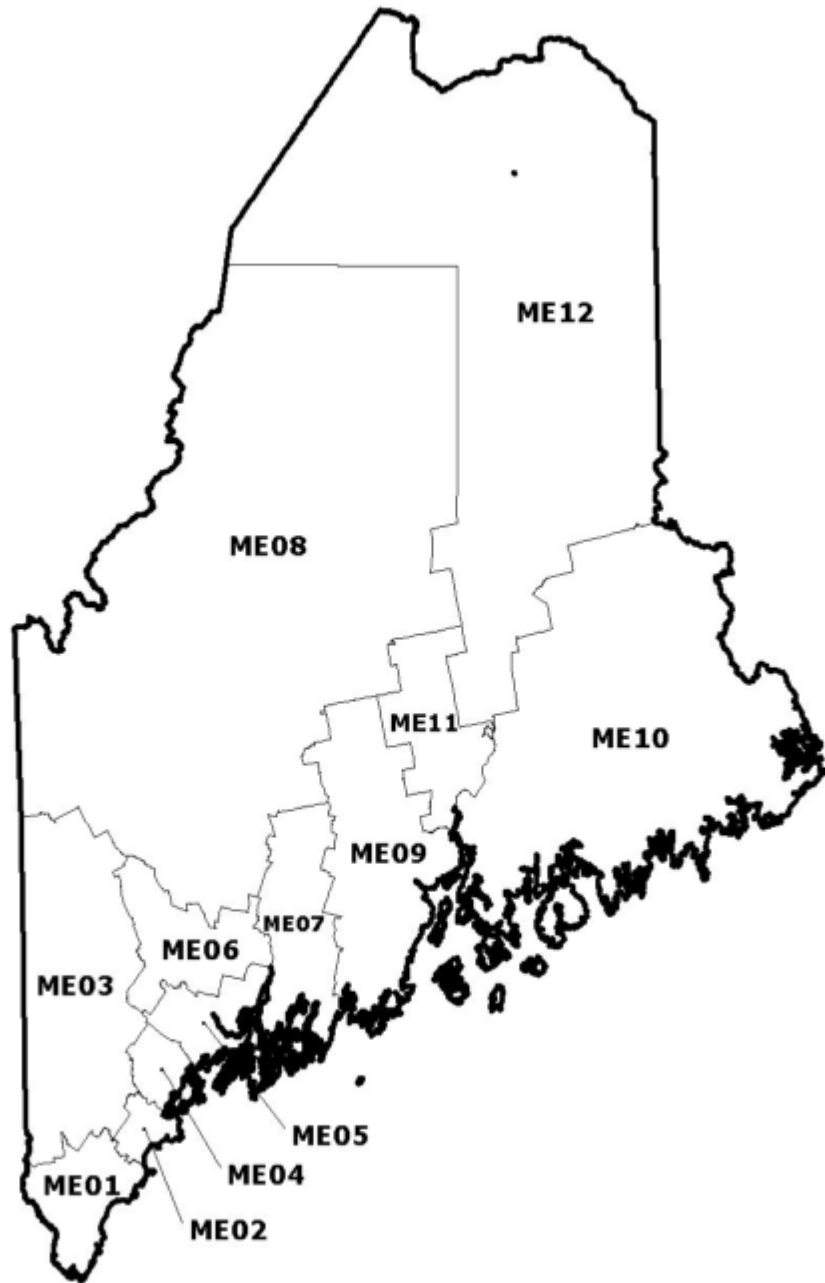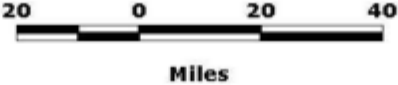
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Alabama
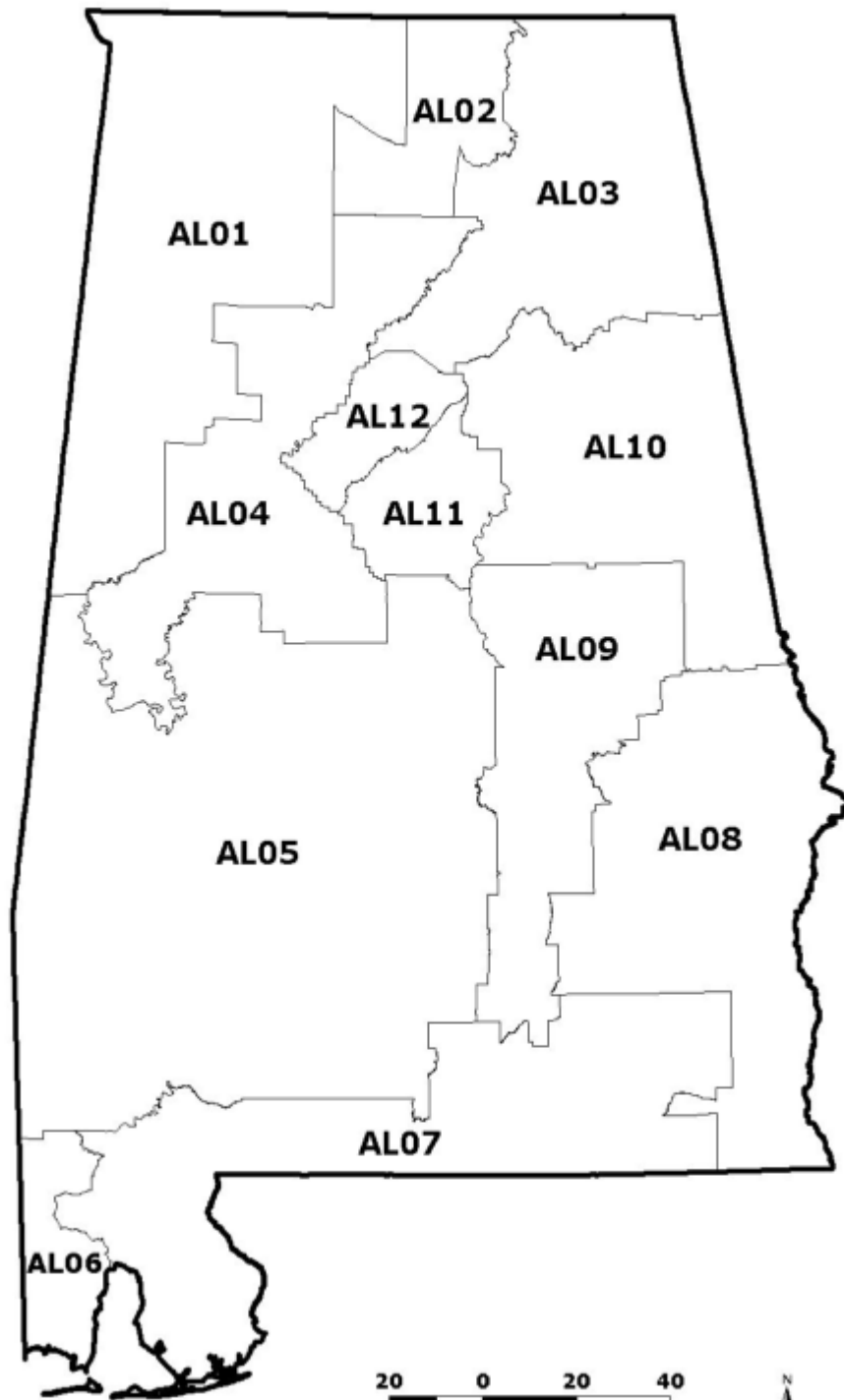
AL02

AL03

AL01

AL12

AL10

AL04

AL11

AL09

AL05

AL08

AL07

AL06

20   0   20   40

Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Alaska

AK04

AK04

AK05

AK06

AK04

AK07

AK03

AK02

See Inset
next page

AK04

AK02

AK03

See Inset
next page

AK03

AK03

AK03

AK03

100   0   100   200

Miles

N
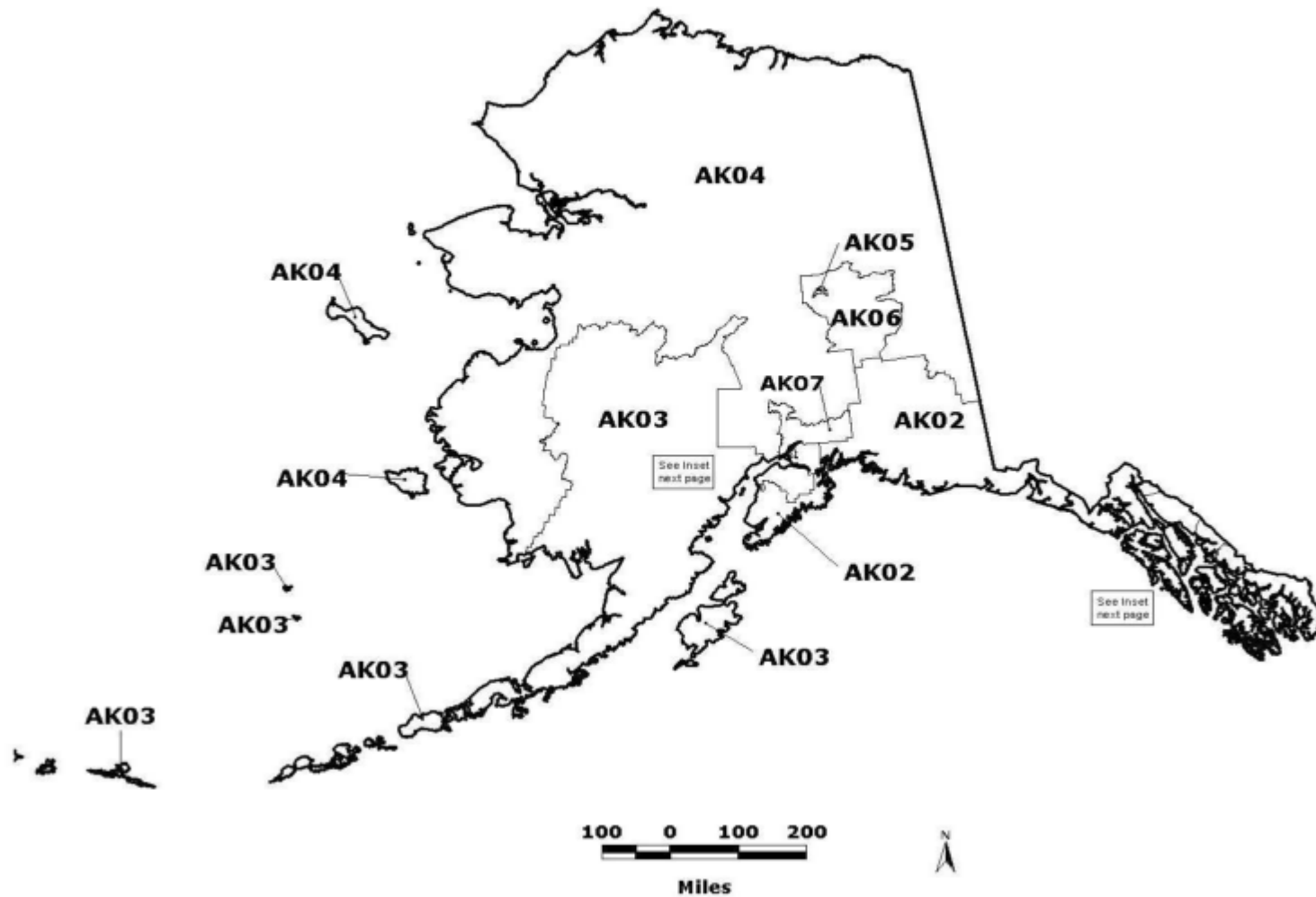
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Alaska Insets

Anchorage Area

AK07

AK04

AK10

AK09

AK08

AK11

AK12

AK03

AK02

AK02

10   0   10   20

Miles

Alaska Panhandle

AK02

AK02

AK01

AK02

AK02

AK01

AK02

AK02

AK01

AK02

AK01

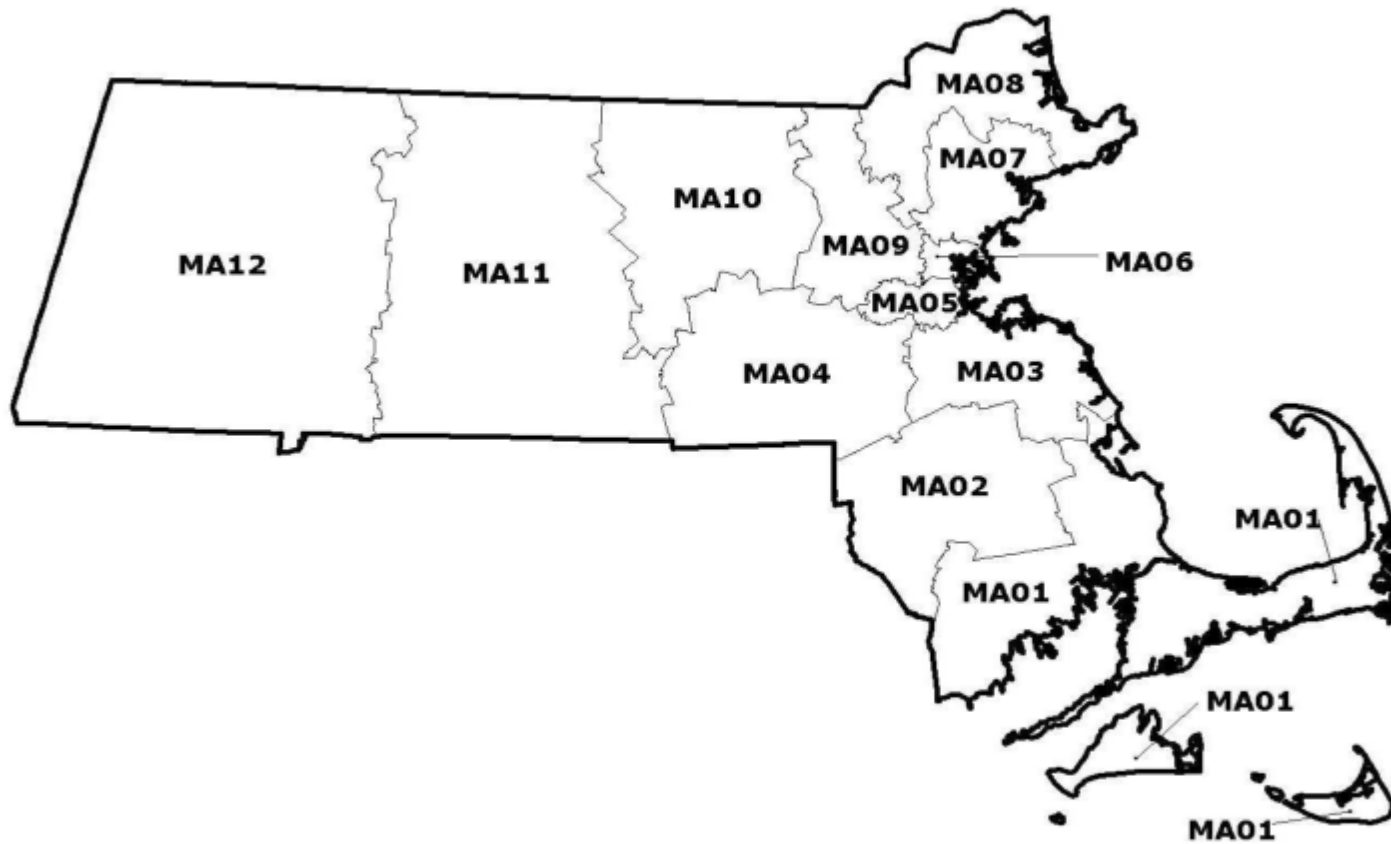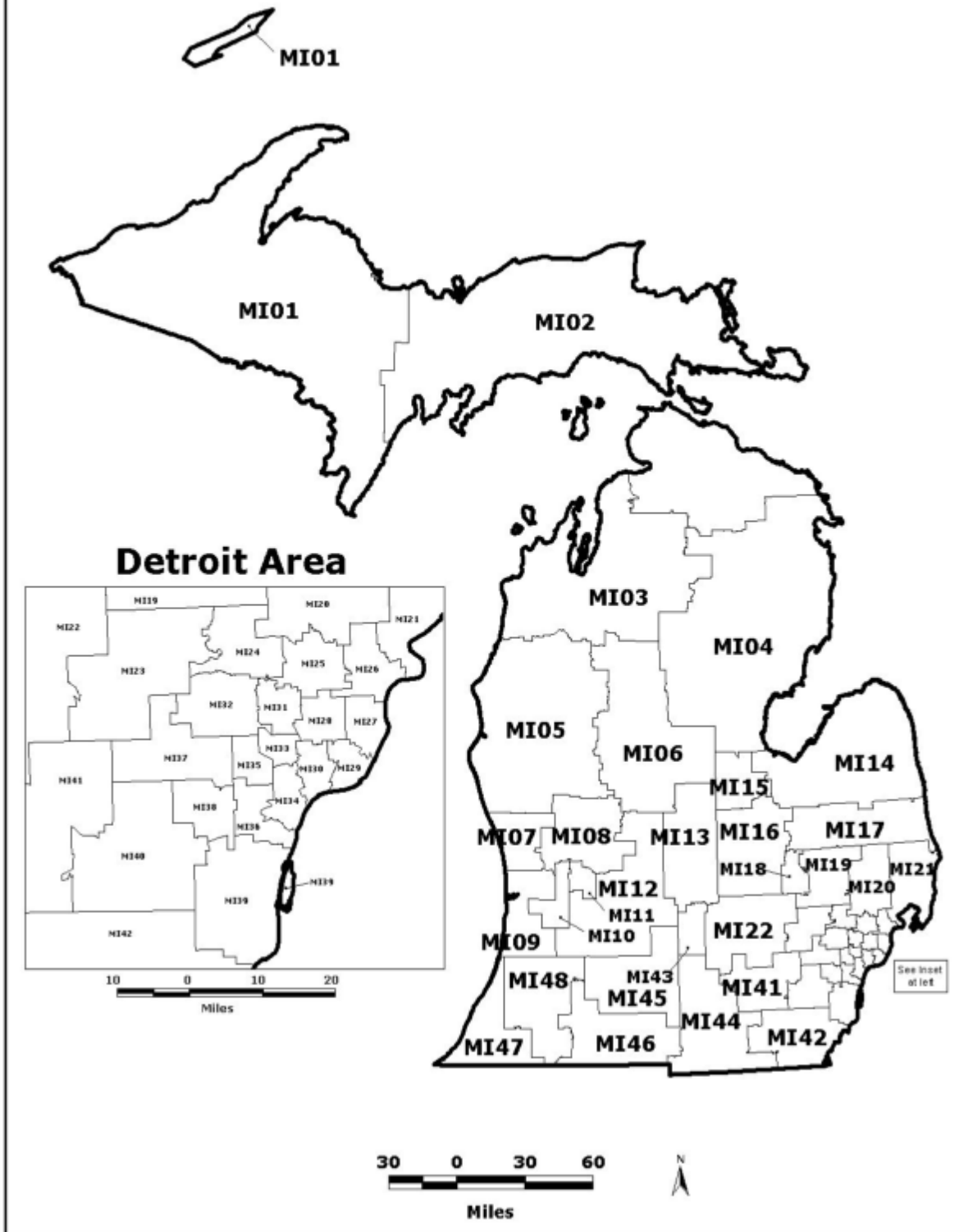AK01

AK01

80   0   80   160

Miles

A-3
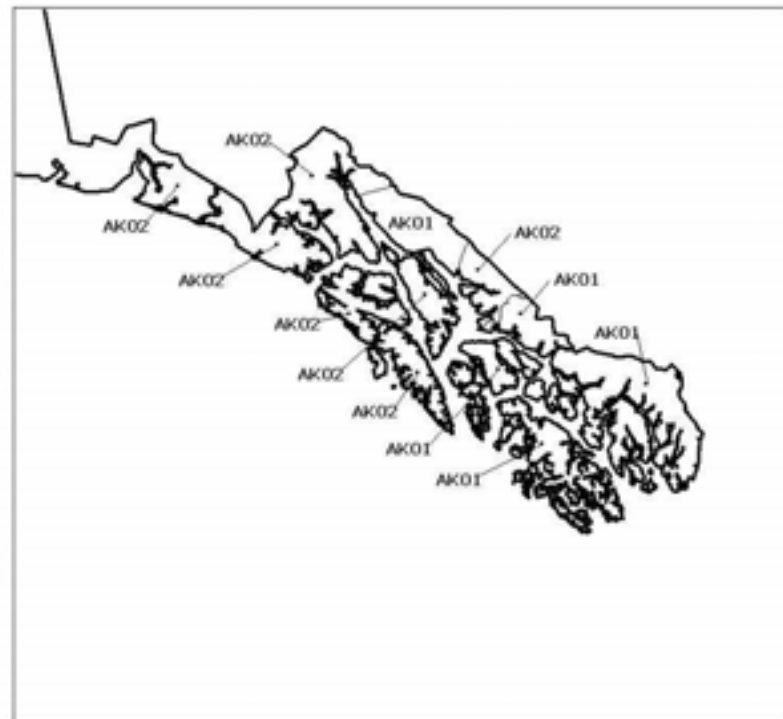
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Arizona

A-4
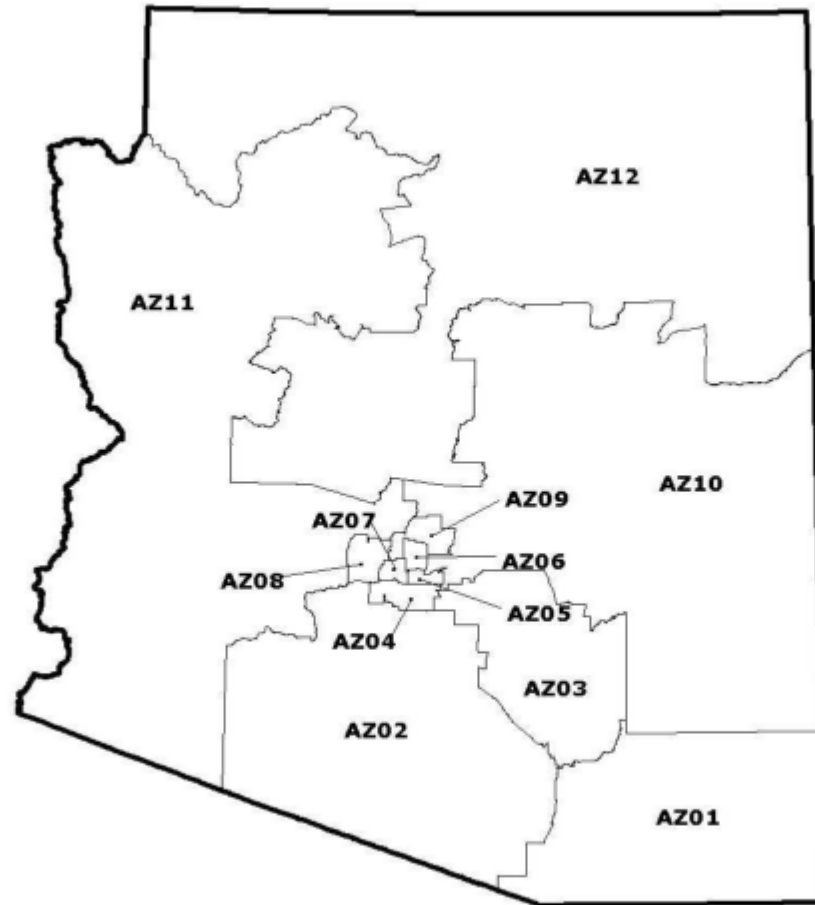
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Arkansas
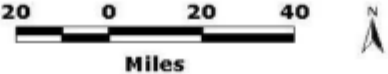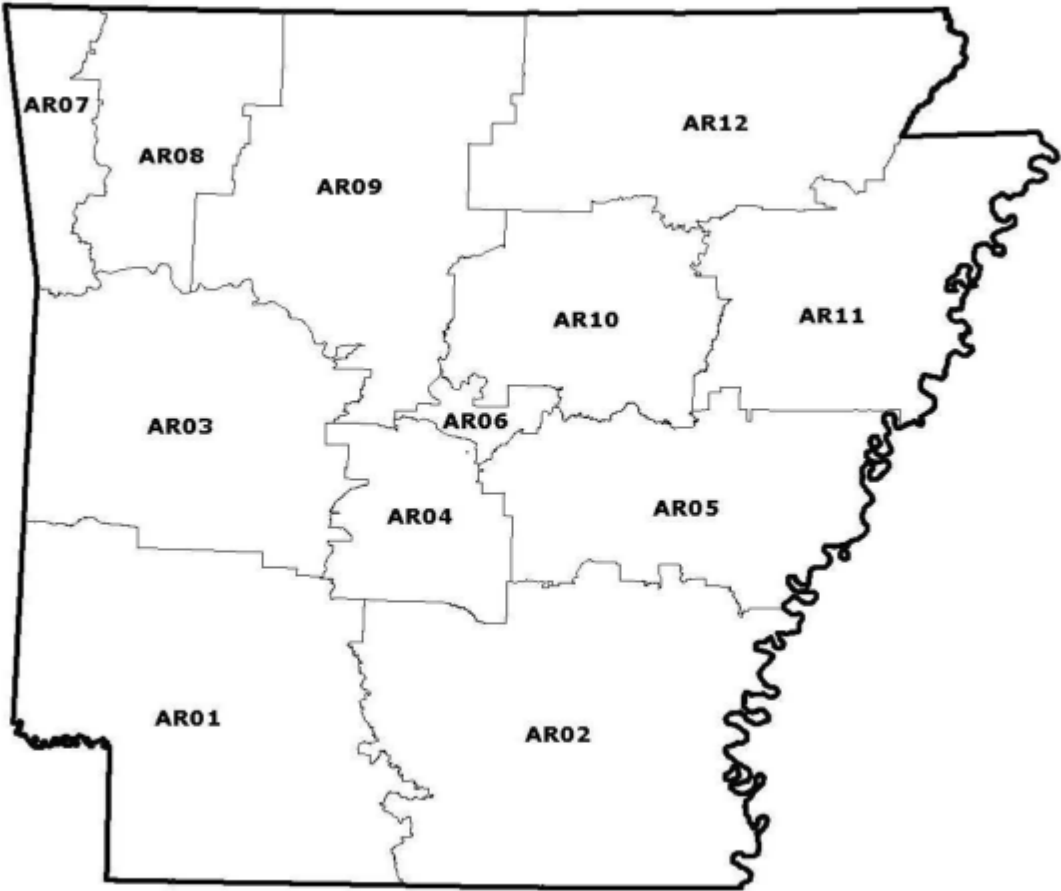
AR07

AR08

AR12

AR09

AR10

AR11

AR03

AR06

AR04

AR05

AR01

AR02

20   0   20   40

N

Miles

1999-2003 NHSDA/NSDUH Field Interviewer Regions: California

CA01

CA02

CA07

CA15

CA18

CA17

CA16

CA19

CA22

CA23

CA20

CA20

CA20

CA21

CA29

CA24

CA21

CA26

CA39

CA28

CA25

CA27

See Inset next page

See Inset next page

40   0   40   80

Miles

N
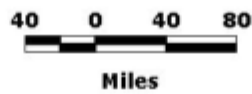
1999-2003 NHSDA/NSDUH Field Interviewer Regions: California Insets

San Francisco Area
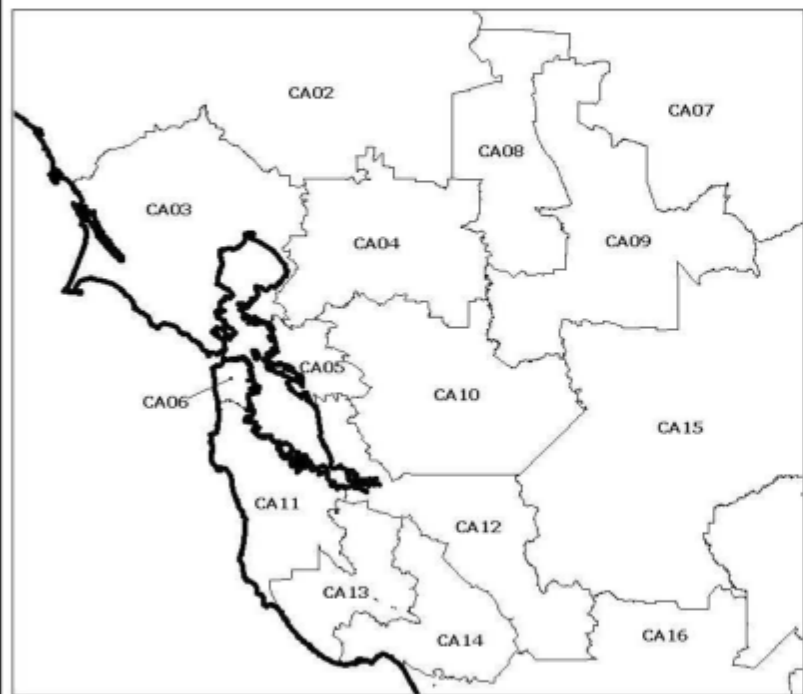
Los Angeles Area

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Colorado
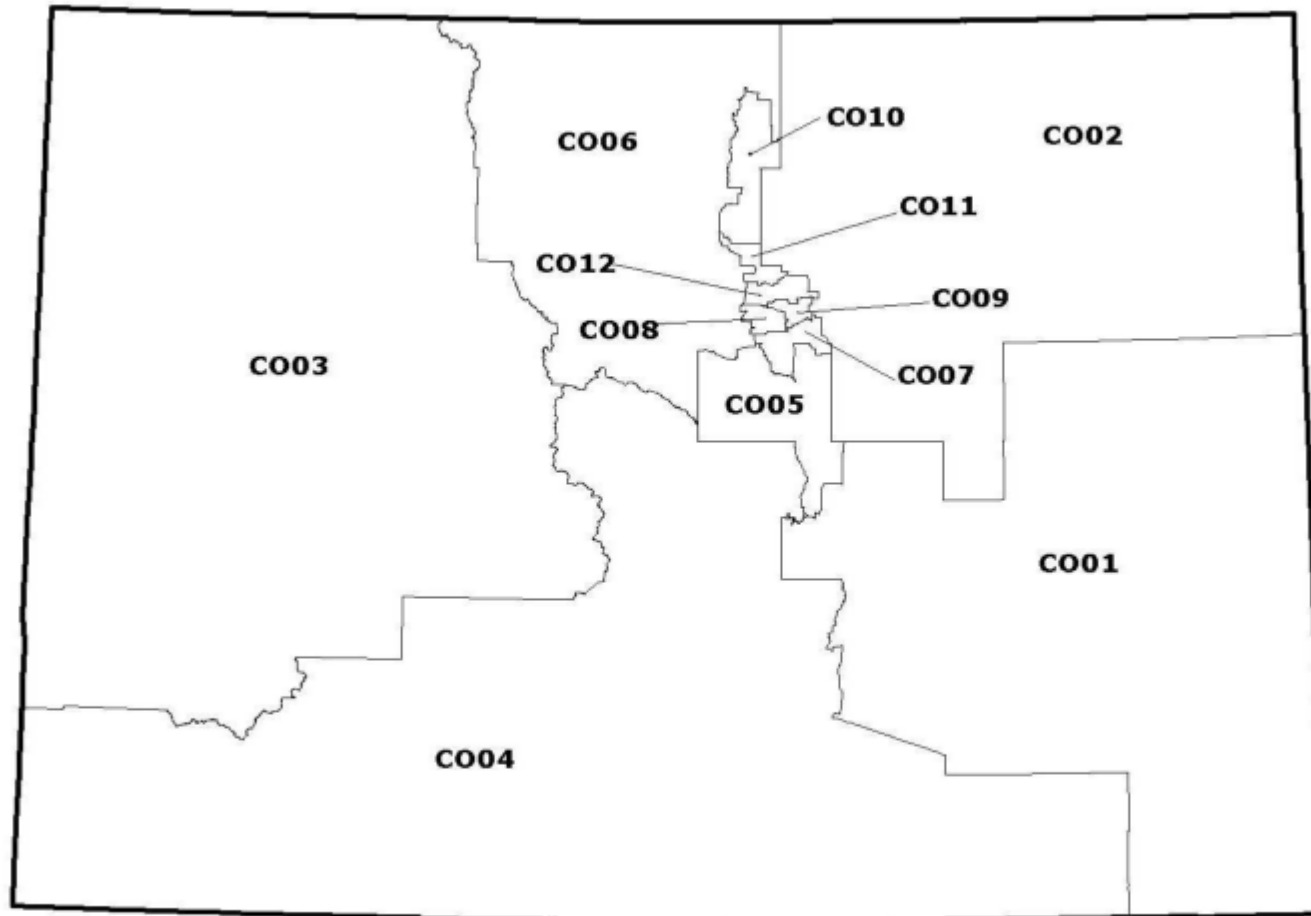
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Connecticut

CT06

CT09

CT11

CT08

CT10

CT12

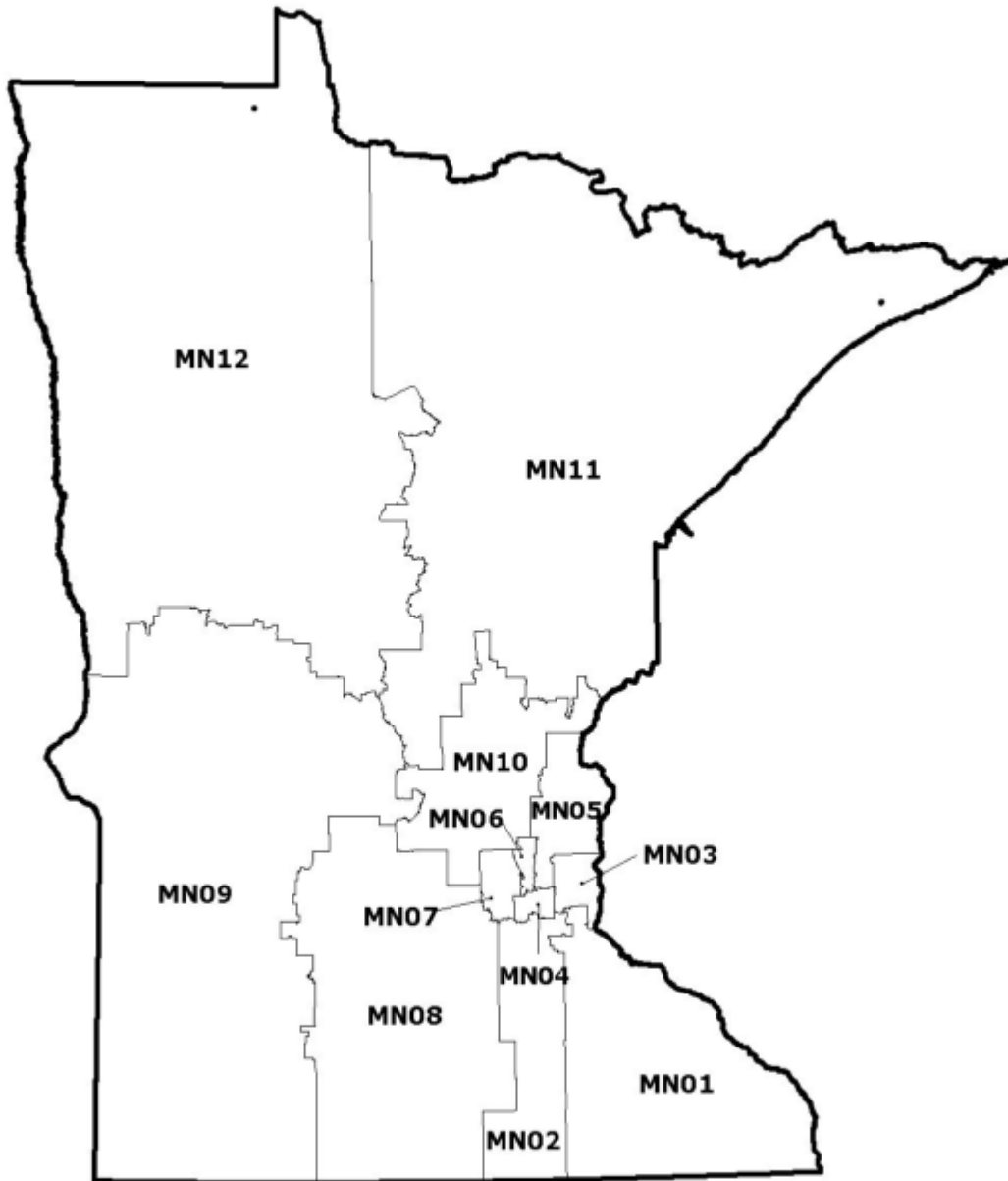CT05

CT04

CT07

CT03
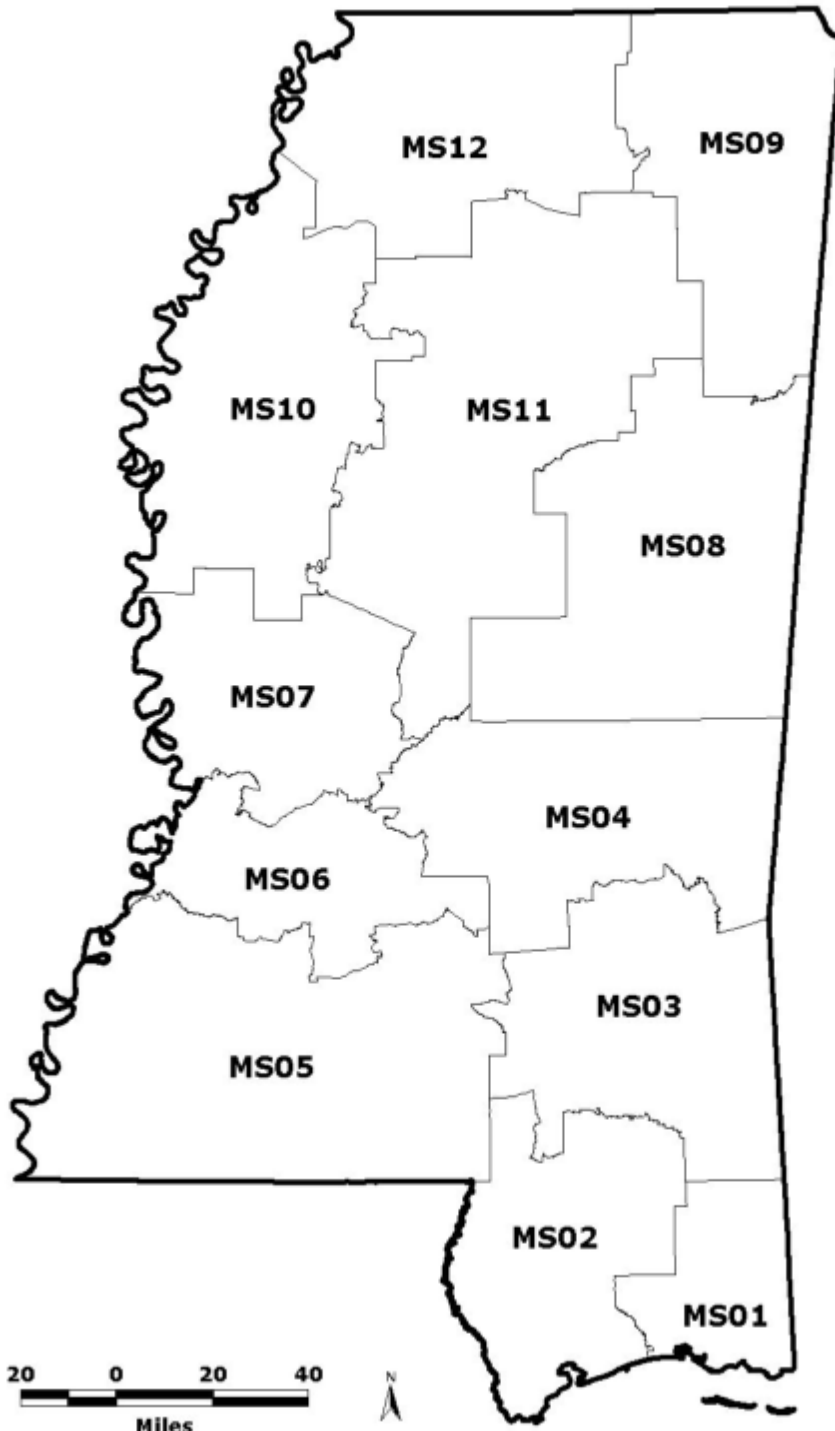
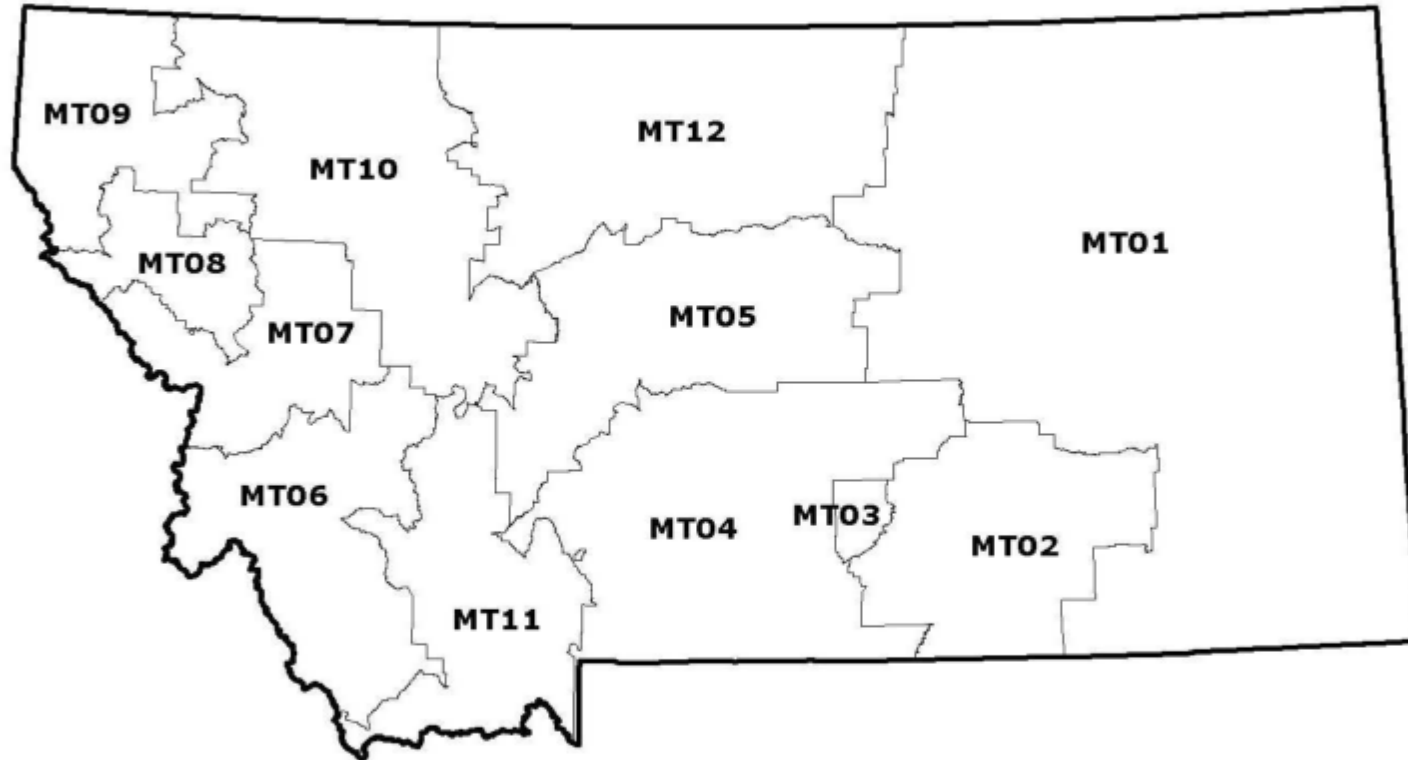CT02

CT01

8   0   8   16

Miles

N
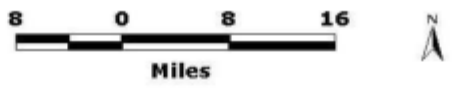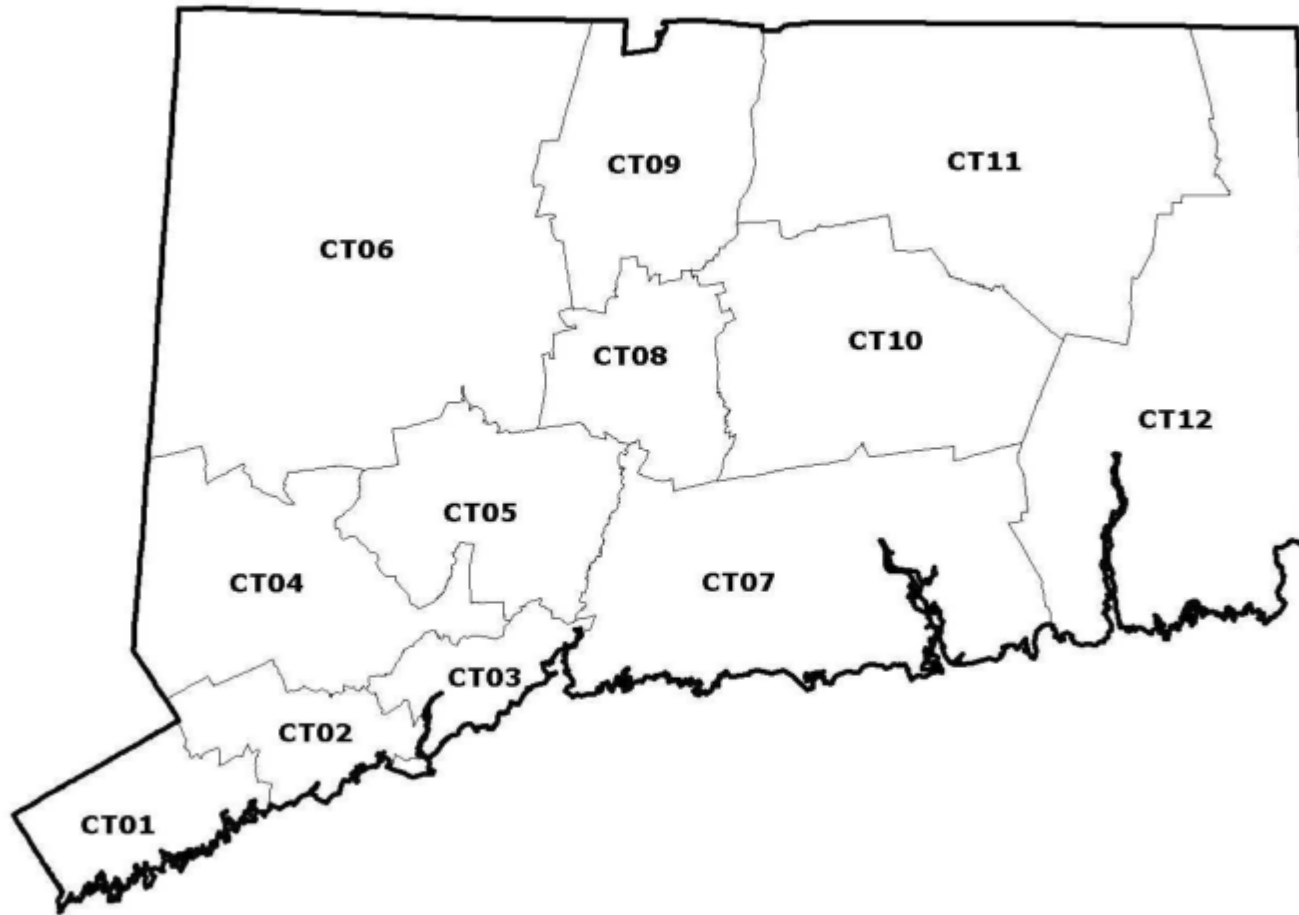
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Delaware
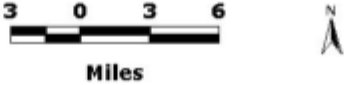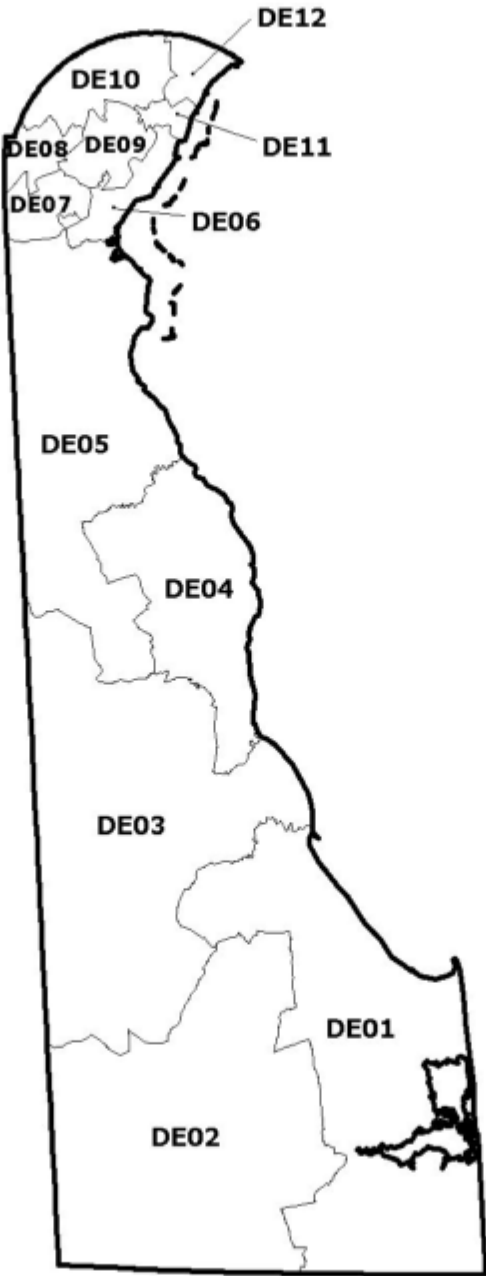
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Florida
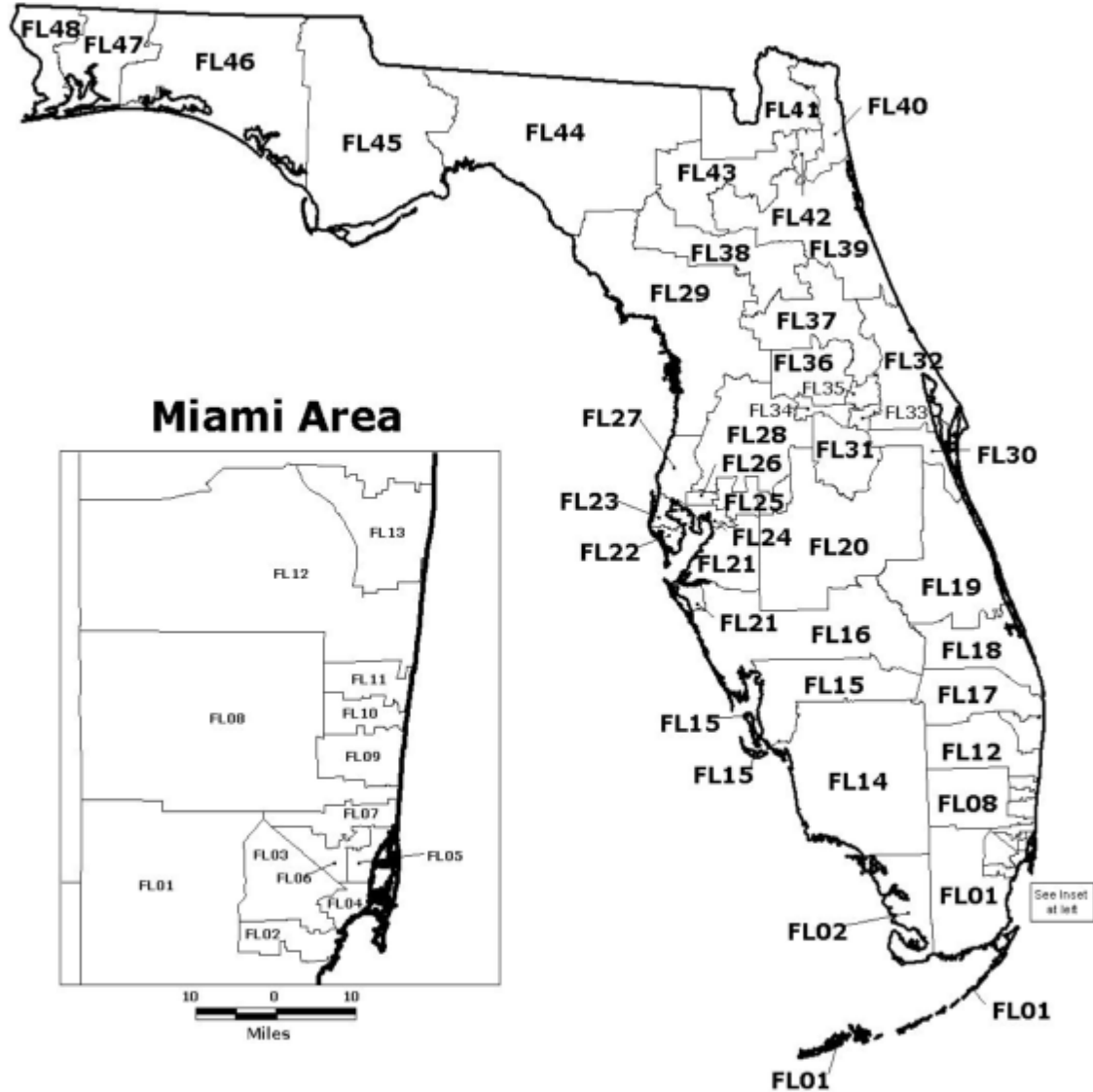
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Georgia

GA11

GA12

GA07

GA10

GA09

GA08

GA06

GA05

GA03

GA04

GA02

GA01

20 0 20 40

N

Miles
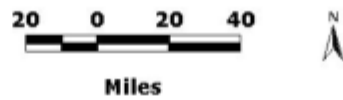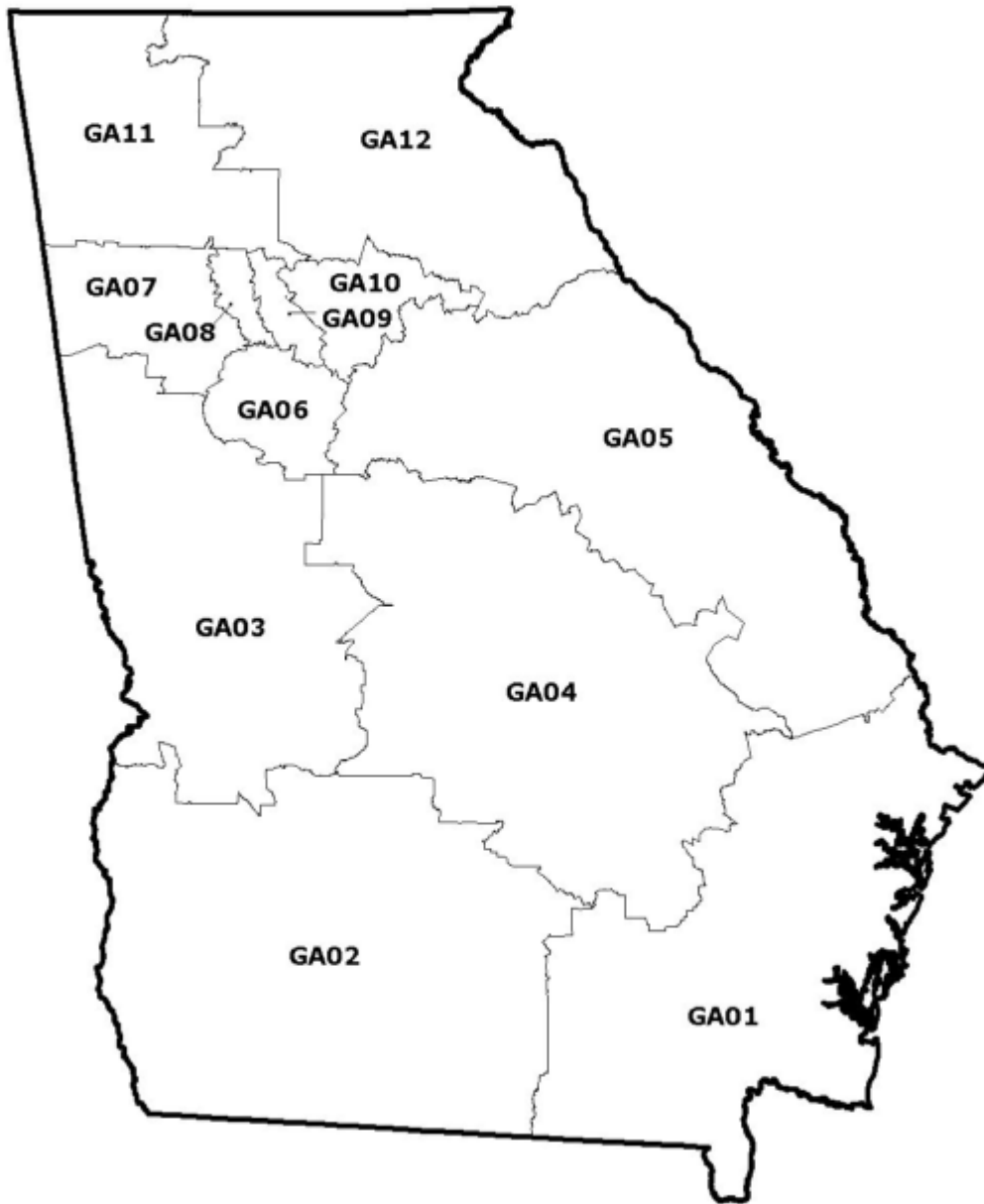
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Hawaii
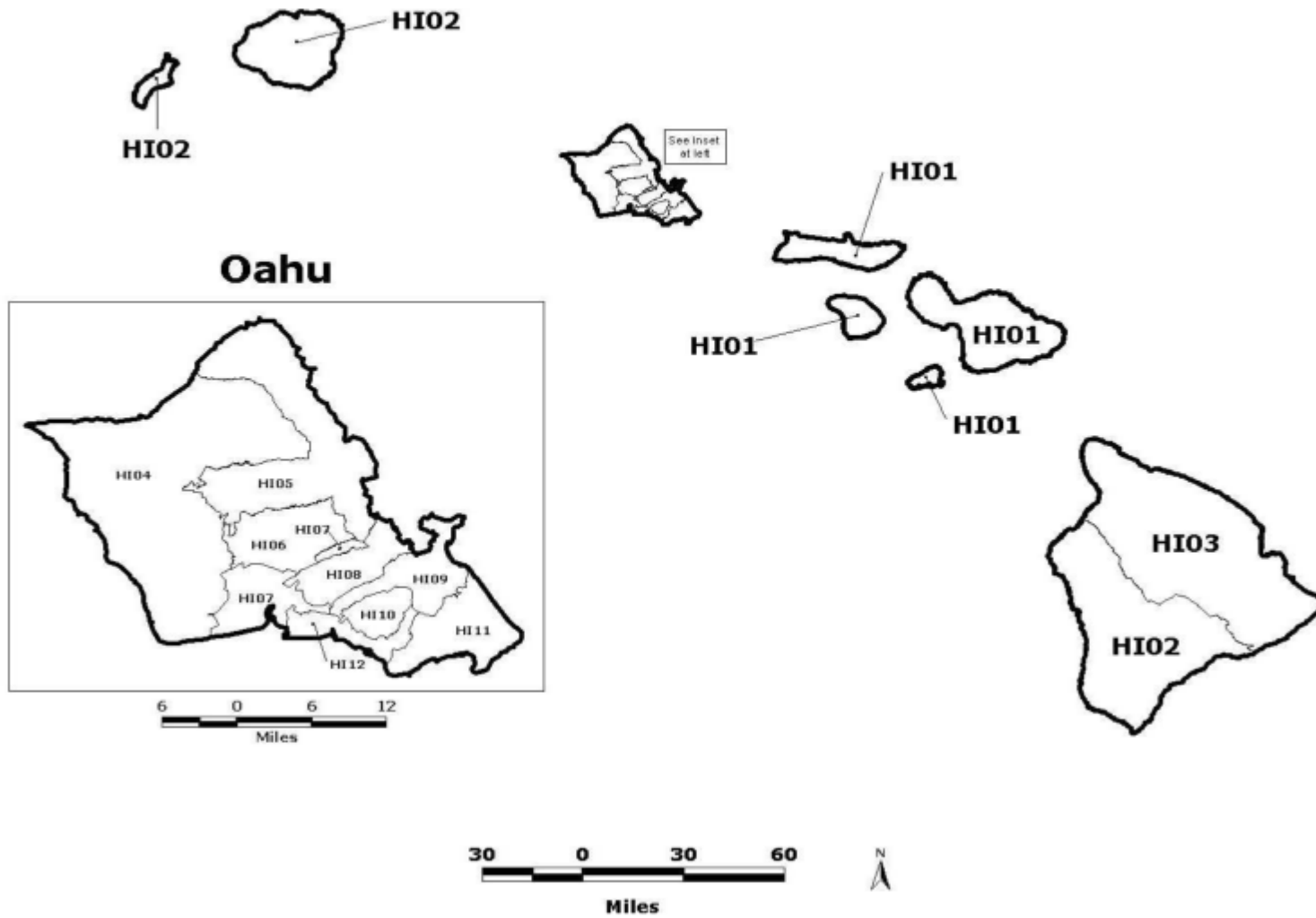
HI02

HI02

See Inset
at left

HI01

HI01

HI01

Oahu

HI04

HI05

HI07

HI06

HI08

HI09

HI07

HI10

HI11

HI12

HI03

HI02

6 0 6 12
Miles

30 0 30 60
Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Idaho

ID11

ID12

ID10

ID04

ID06

ID09

ID01

ID07

ID08

ID02

ID05

ID03

30   0   30   60

N

Miles
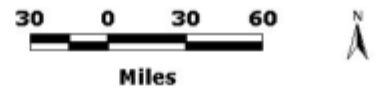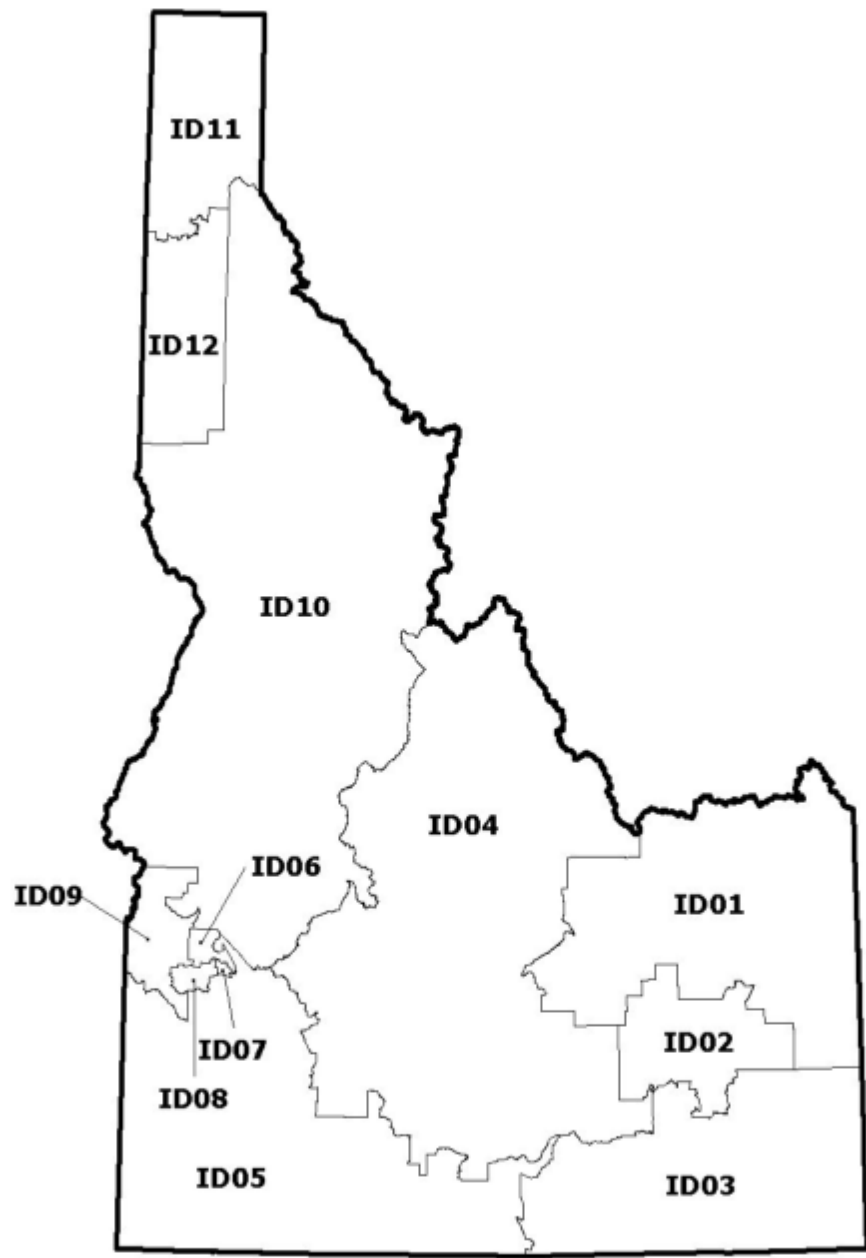
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Illinois

IL15

IL44

IL16

IL17

IL13

IL14

IL18

See inset
next page

IL12

IL10

IL09

IL11

IL08

IL07

IL04

IL06

IL05

IL03

IL01

IL02

20   0   20   40

N

Miles
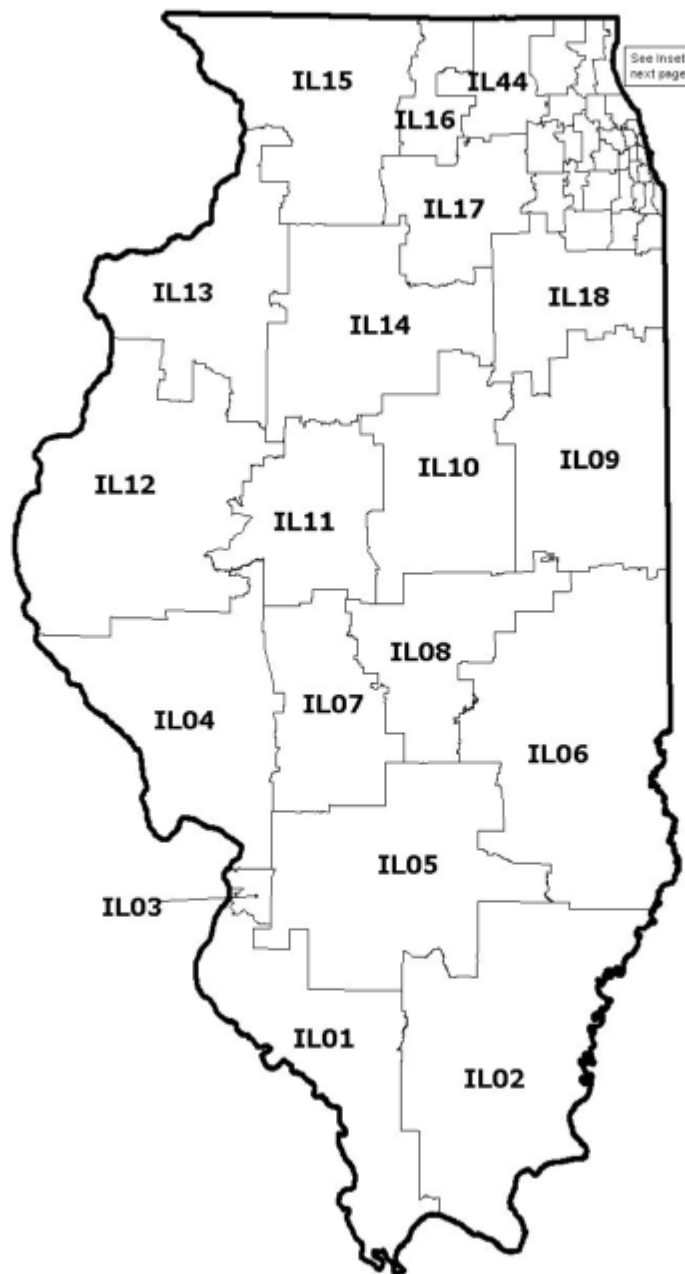
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Chicago Area

IL47

IL46

IL44

IL48

IL45

IL40

IL39

IL36

IL34

IL37

IL33

IL42

IL41

IL32

IL43

IL38

IL31

IL17

IL29

IL35

IL23

IL30

IL24

IL28

IL25

IL22

IL26

IL27

IL21

IL20

IL19

IL18

30   0   30   60

Miles

N
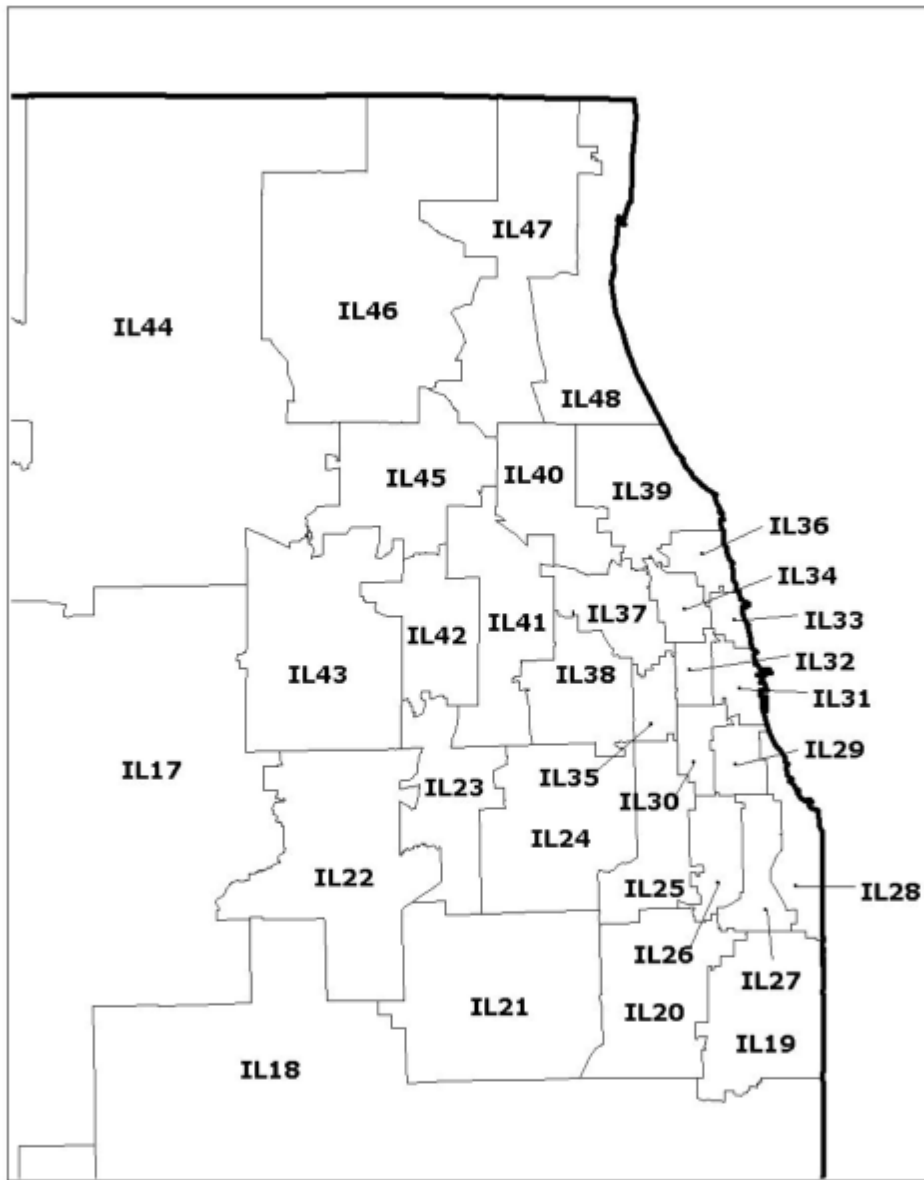
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Indiana
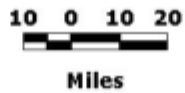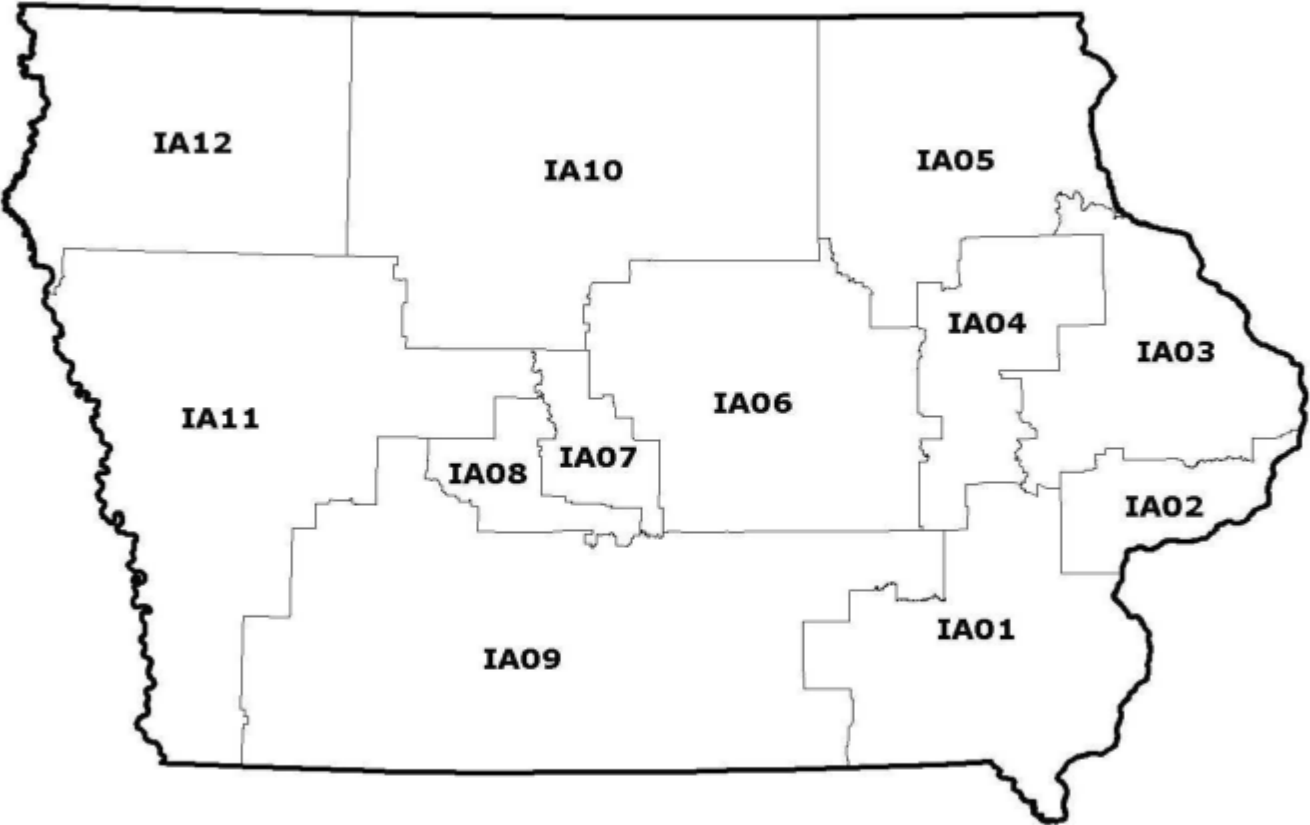
IN11

IN12

IN10

IN08

IN09

IN07

IN06

IN05

IN04

IN03

IN02

IN01

10 0 10 20

Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Iowa

IA12

IA10

IA05

IA04

IA03

IA11

IA06

IA08 IA07

IA02

IA09

IA01

30   0   30   60

Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Kansas

KS04

KS06

KS01

KS12

KS05

KS02

KS10

KS03

KS07

KS09

KS08

KS11

30   0   30   60

N

Miles
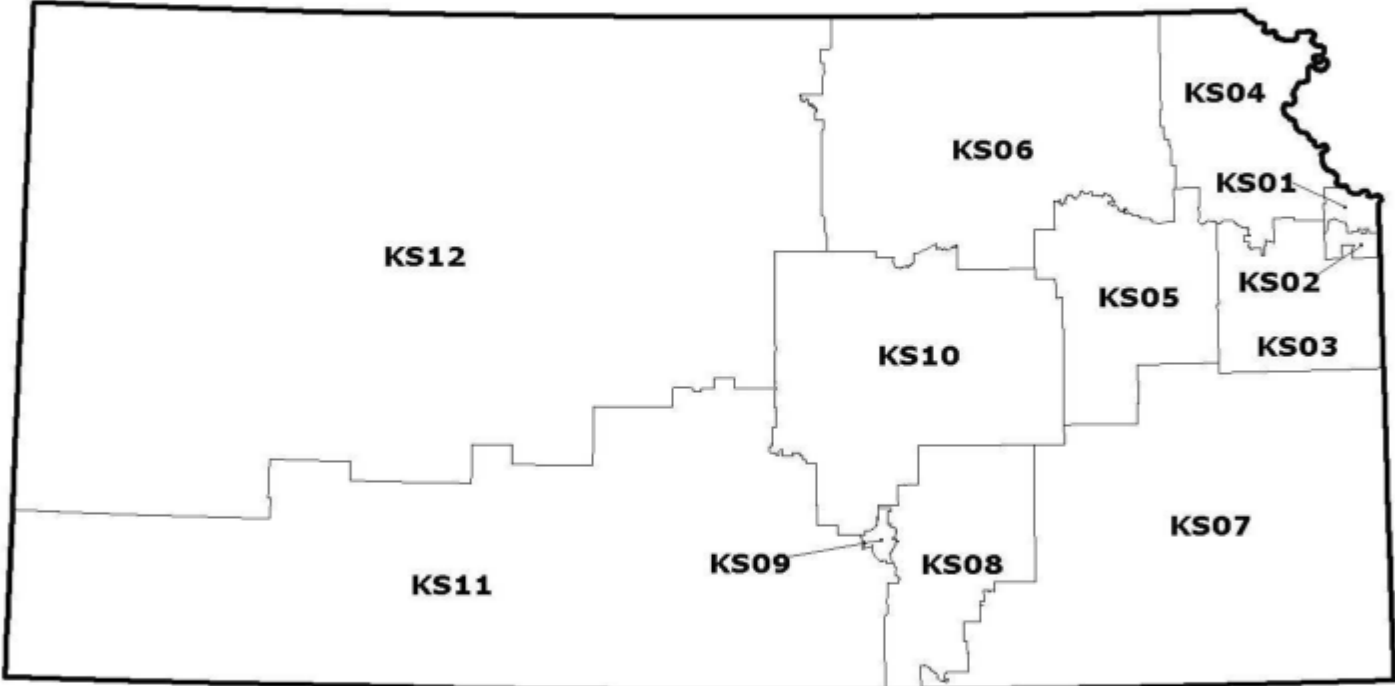
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Kentucky

KY04

KY08

KY06

KY03

KY02

KY07

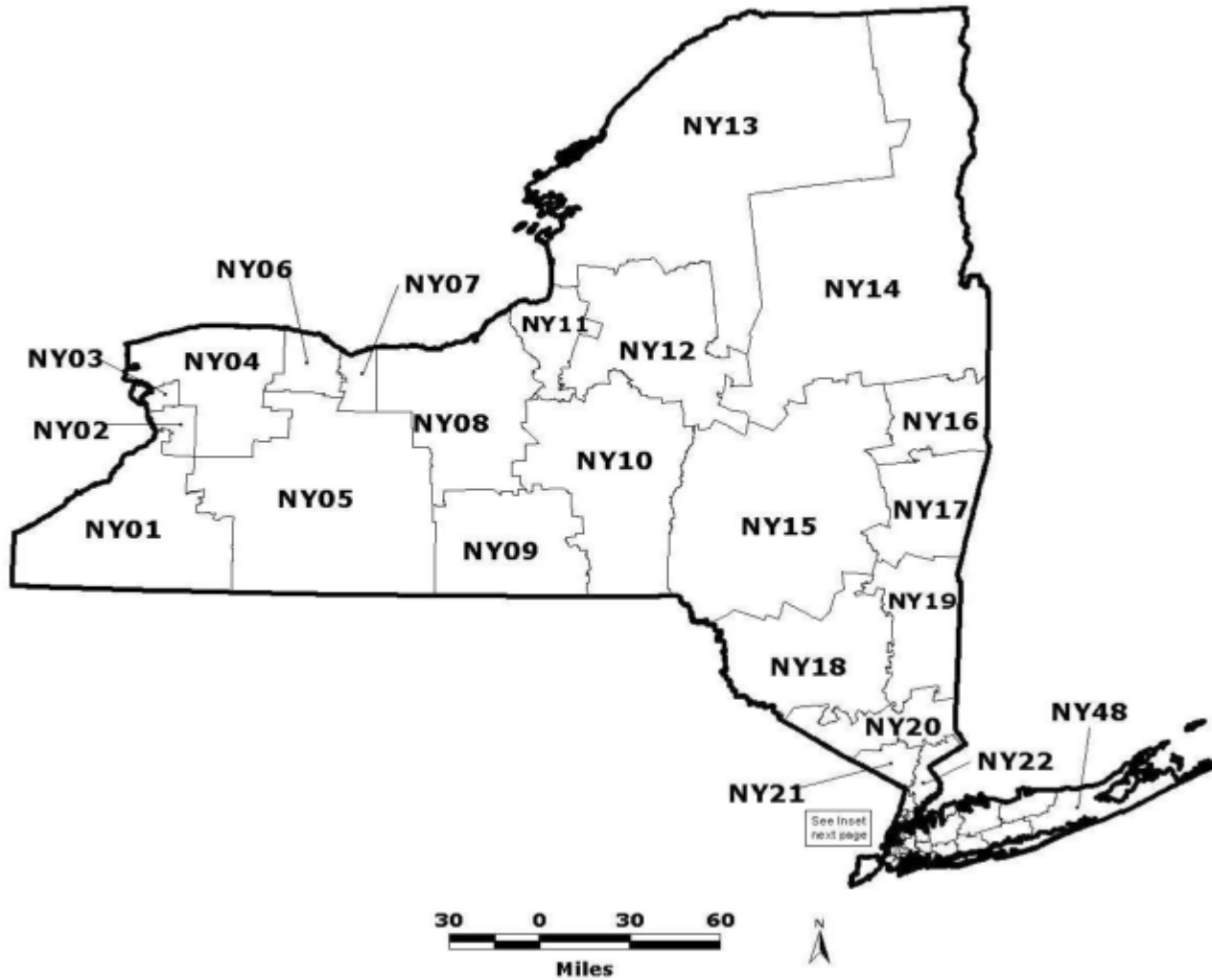KY09

KY05

KY11

KY12

KY10

KY01

30 0 30 60
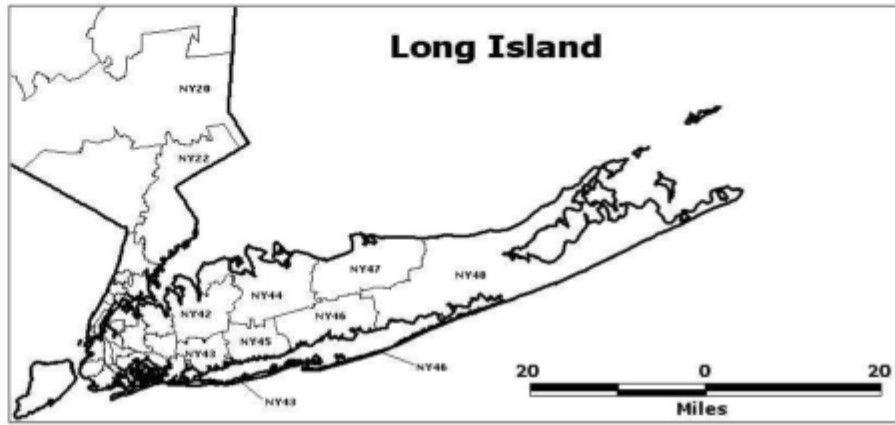
N

Miles
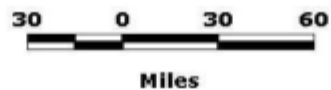
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Louisiana

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Maine

ME12

ME08

ME11

ME10

ME09

ME07

ME06

ME03

ME05

ME04

ME01

ME02

```
20    0    20    40
```
Miles

N

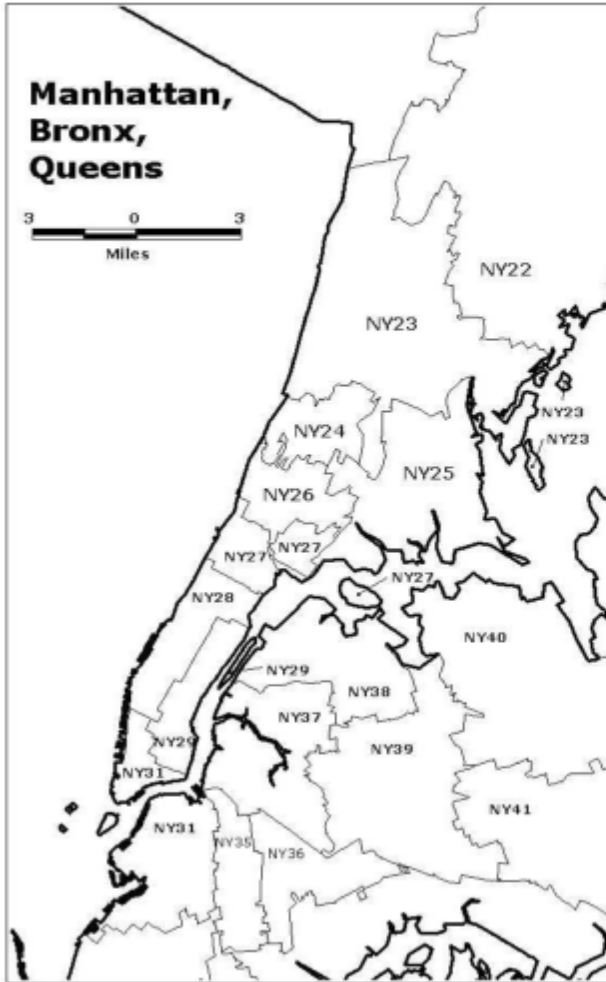# 1999-2003 NHSDA/NSDUH Field Interviewer Regions: Maryland

MD01

MD04 MD03

MD06 MD05

MD07

MD08

MD09

MD10

MD11

MD02

MD12

20 0 20 40

N

Miles
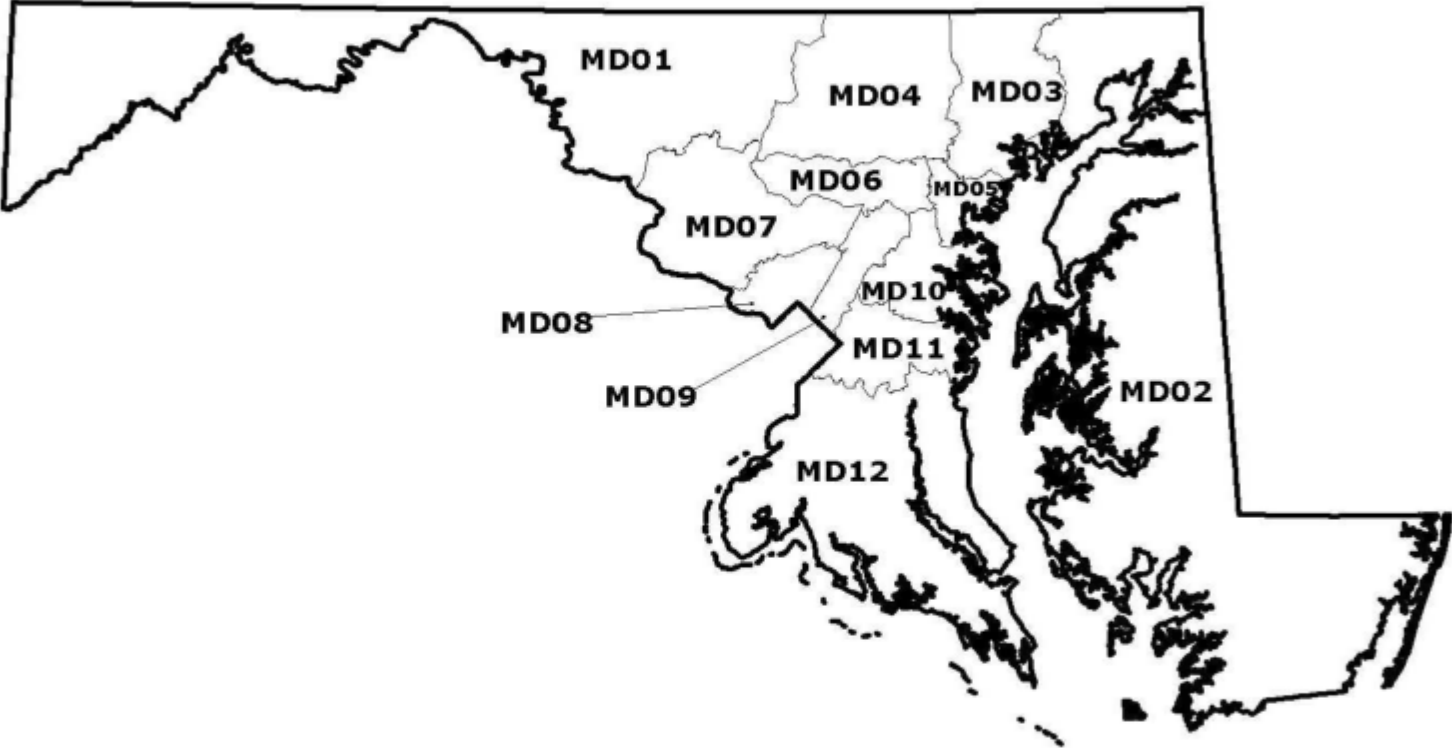
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Massachusetts

MA08

MA07

MA10

MA09

MA12

MA11

MA06

MA05

MA04

MA03

MA02

MA01

MA01

MA01

MA01

10 0 10 20

N

Miles

1999 -2003 NHSDA/NSDUH Field Interviewer Regions: Michigan

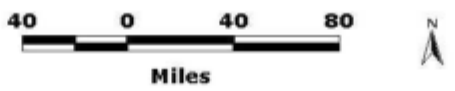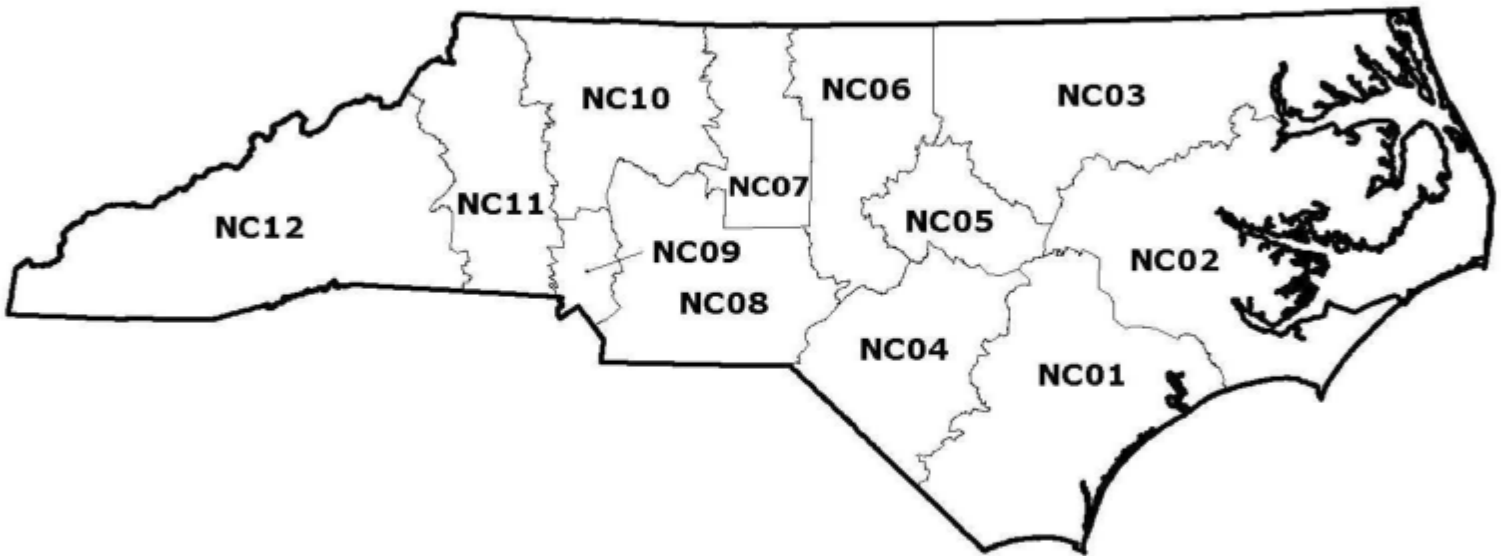MI01

MI01

MI02

MI03

MI04

MI05

MI06

MI14

MI15

MI07

MI08

MI13

MI16

MI17

MI18

MI19

MI21

MI20

MI12

MI11

MI22

MI09

MI10

MI48

MI43

MI45

MI41

MI44

MI42

MI47

MI46

See Inset
at left

Detroit Area

MI19

MI20

MI21

MI22

MI23

MI24

MI25

MI26

MI32

MI31

MI28

MI27

MI37

MI33

MI35

MI30

MI29

MI41

MI34

MI36

MI40

MI39

MI39

MI42

10   0   10   20

Miles

30   0   30   60

Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Minnesota

MN12

MN11

MN10

MN06

MN05

MN03

MN09

MN07

MN04

MN08

MN01

MN02

30   0   30   60

Miles

N
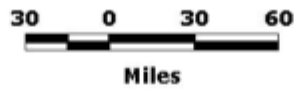
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Mississippi

MS12

MS09

MS10

MS11

MS08

MS07

MS04

MS06

MS03

MS05

MS02

MS01

20   0   20   40
Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Missouri

MO11

MO09

MO12

MO06

MO05

MO04

MO03

MO10

MO02

MO07

MO01

MO08

30 0 30 60

Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Montana
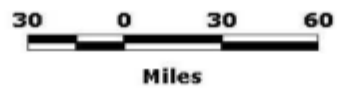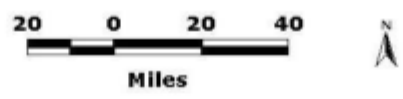
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Nebraska
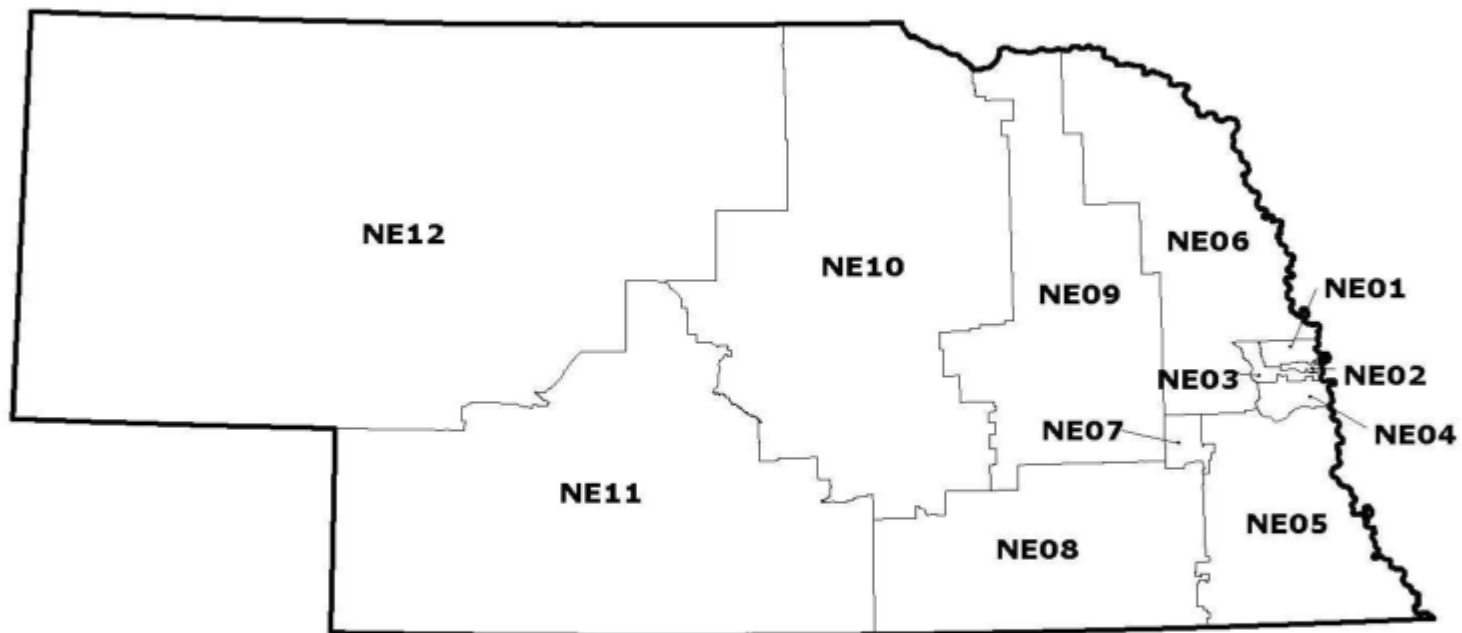
NE12

NE10

NE06

NE09

NE01

NE03 NE02

NE07 NE04

NE11 NE05

NE08

40 0 40 80

Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Nevada

NV12

NV11

NV10

NV09

NV01

NV08

NV05

NV08

NV04

NV06

NV03

NV07

NV02

30　0　30　60

Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: New Hampshire

NH01

NH12

NH08

NH09

NH02

NH06

NH04

NH03

NH11

NH10

NH07

NH05

8 0 8 16
Miles

N
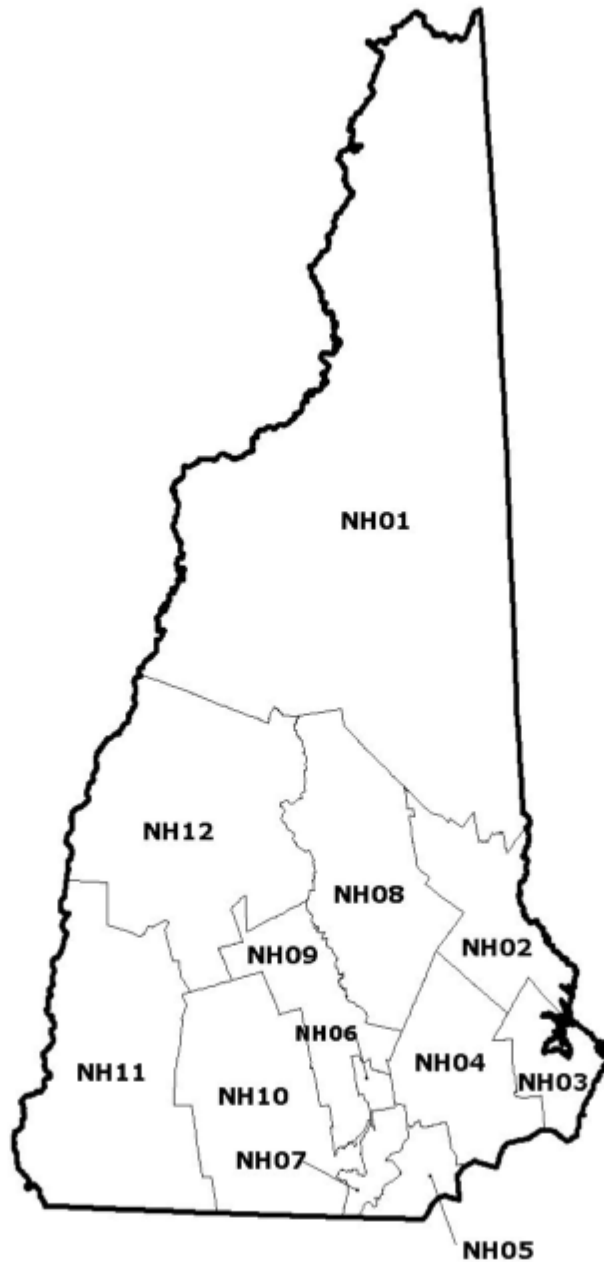
1999-2003 NHSDA/NSDUH Field Interviewer Regions: New Jersey

NJ09

NJ11

NJ12

NJ10

NJ08

NJ07

NJ06

NJ05

NJ04

NJ03

NJ02

NJ01

8   0   8   16
Miles

N
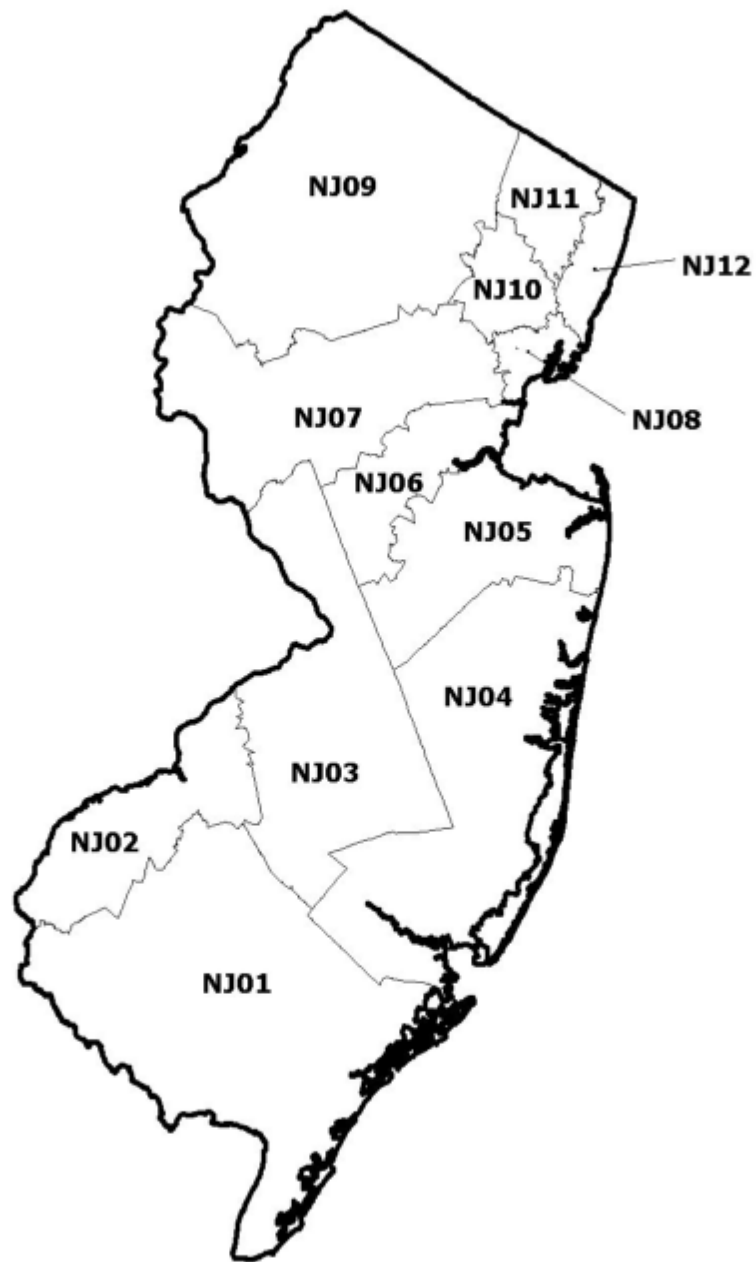
1999-2003 NHSDA/NSDUH Field Interviewer Regions: New Mexico

NM09

NM04

NM10

NM05
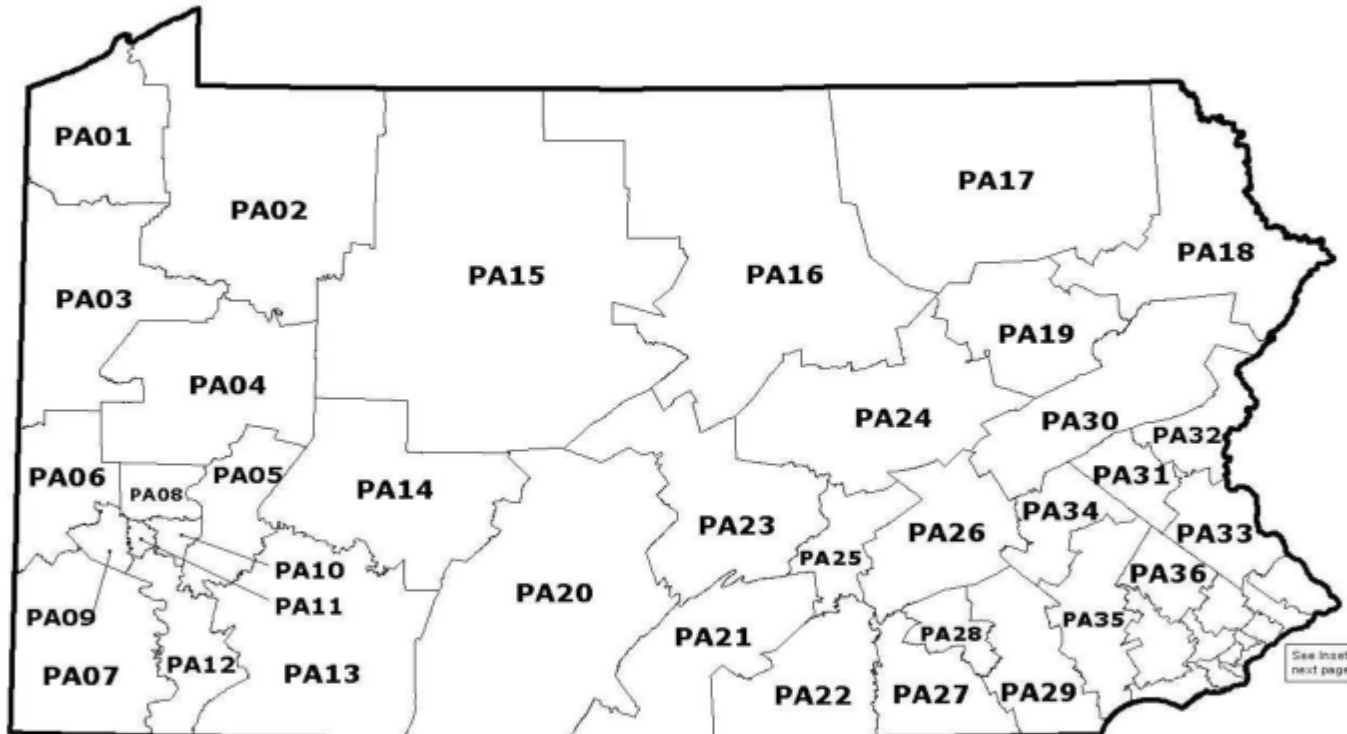
NM02

NM08

NM06

NM07

NM03

NM12

NM01

NM11

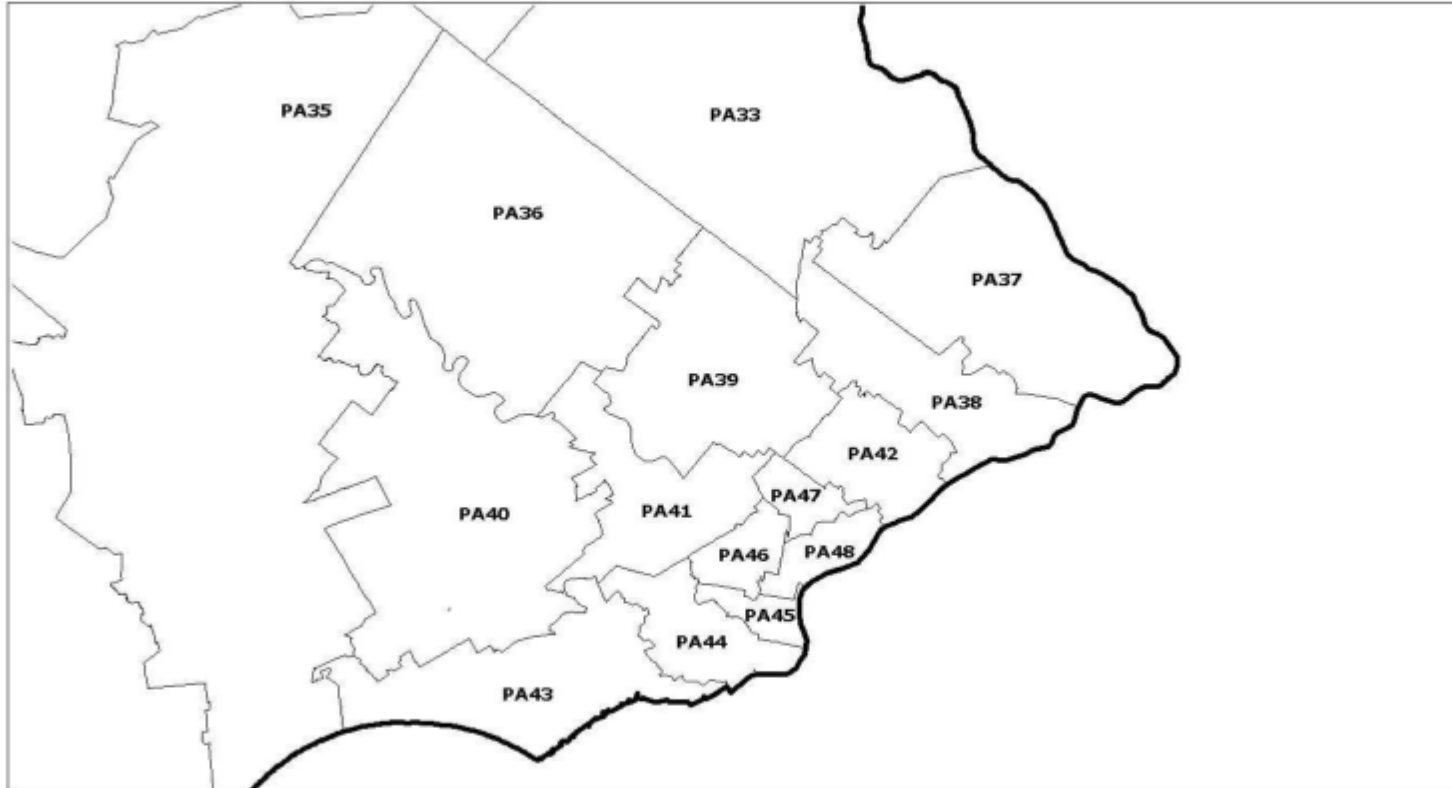40   0   40   80

Miles
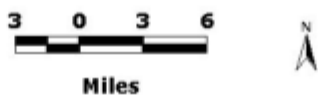
N
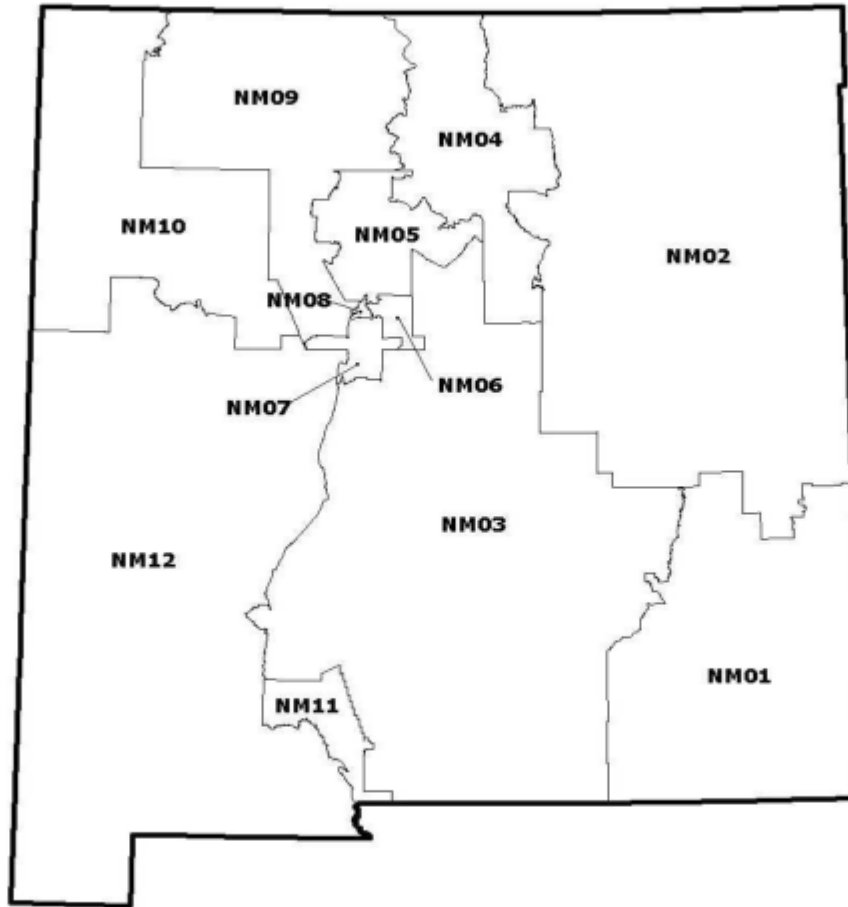
1999-2003 NHSDA/NSDUH Field Interviewer Regions: New York

1999-2003 NHSDA/NSDUH Field Interviewer Regions: New York City Area Insets
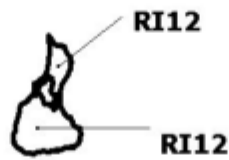
Manhattan, Bronx, Queens

3   0   3
Miles

NY22
NY23
NY23
NY23
NY24
NY25
NY26
NY27 NY27
NY27
NY28
NY40
NY29
NY38
NY37
NY39
NY29
NY31
NY31
NY41
NY35 NY36

Staten Island, Brooklyn

NY27
NY28
NY38
NY40
NY42
NY31
NY29
NY37
NY39
NY41
NY31
NY35 NY36
NY43
NY32
NY34
NY34
NY33
NY34
NY43
NY30

4   0   4
Miles

Long Island

NY28
NY22
NY47 NY48
NY44
NY42
NY46
NY43
NY45
NY46
NY43

20   0   20
Miles

A-36
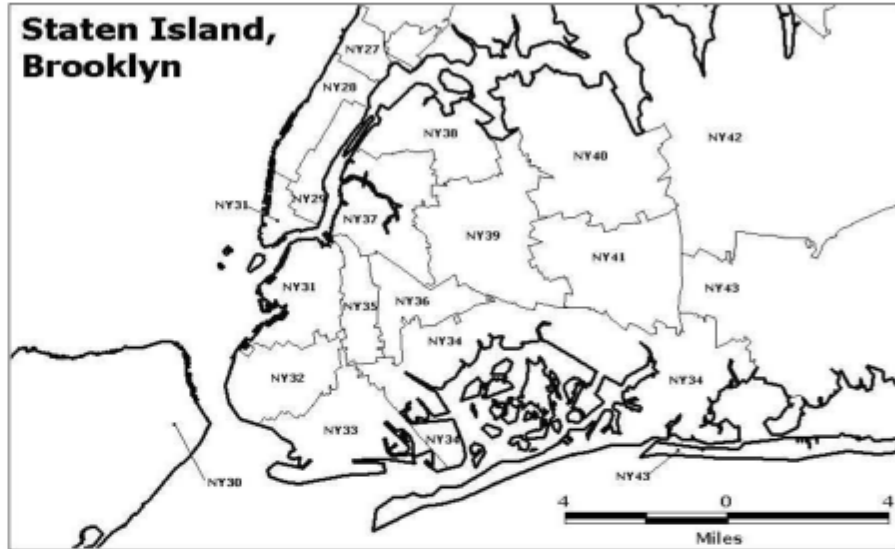
1999-2003 NHSDA/NSDUH Field Interviewer Regions: North Carolina

1999-2003 NHSDA/NSDUH Field Interviewer Regions: North Dakota
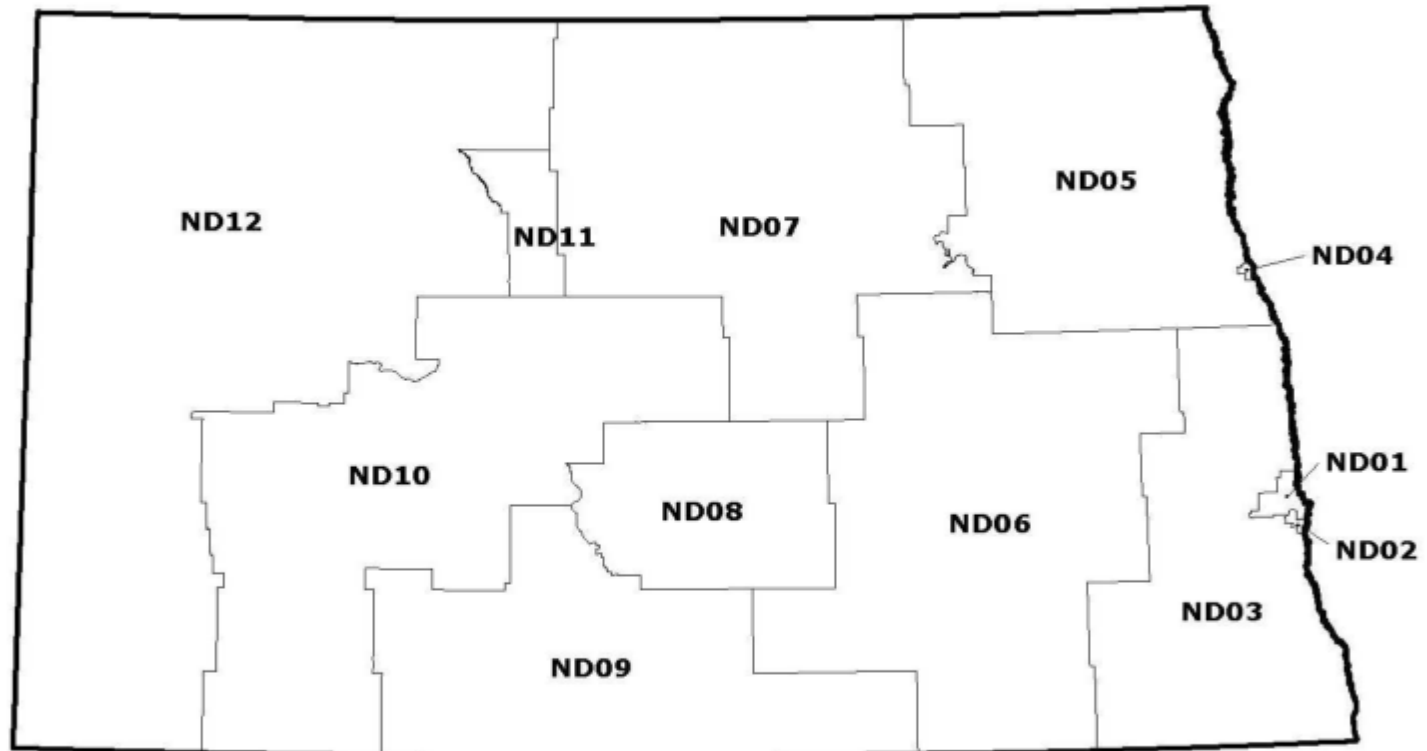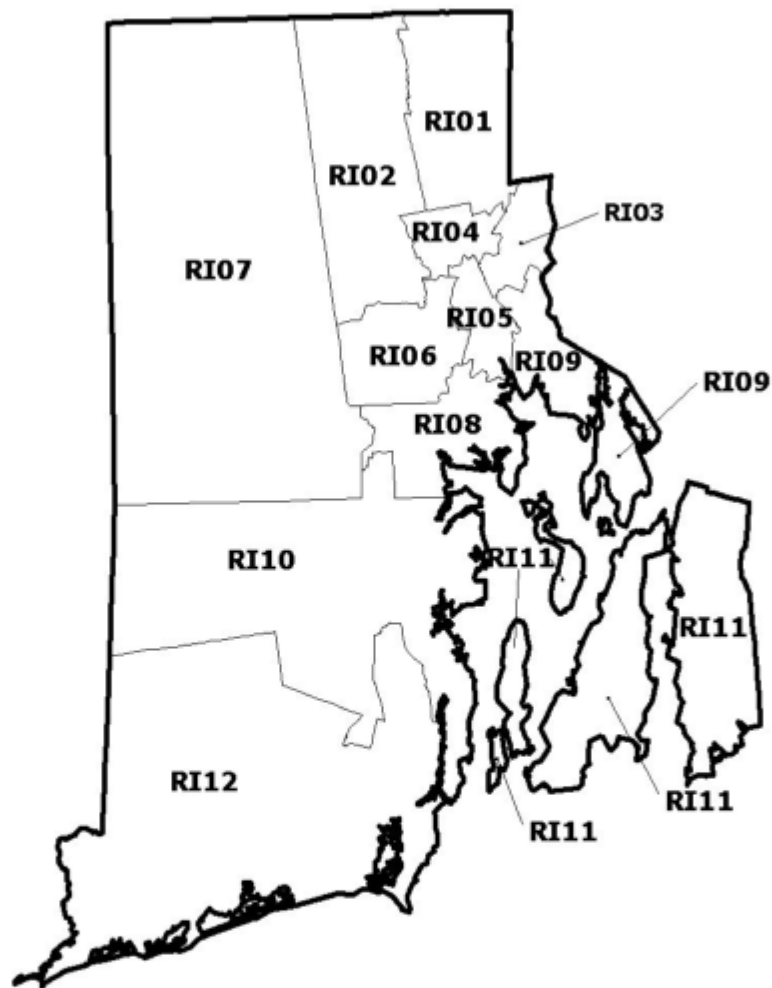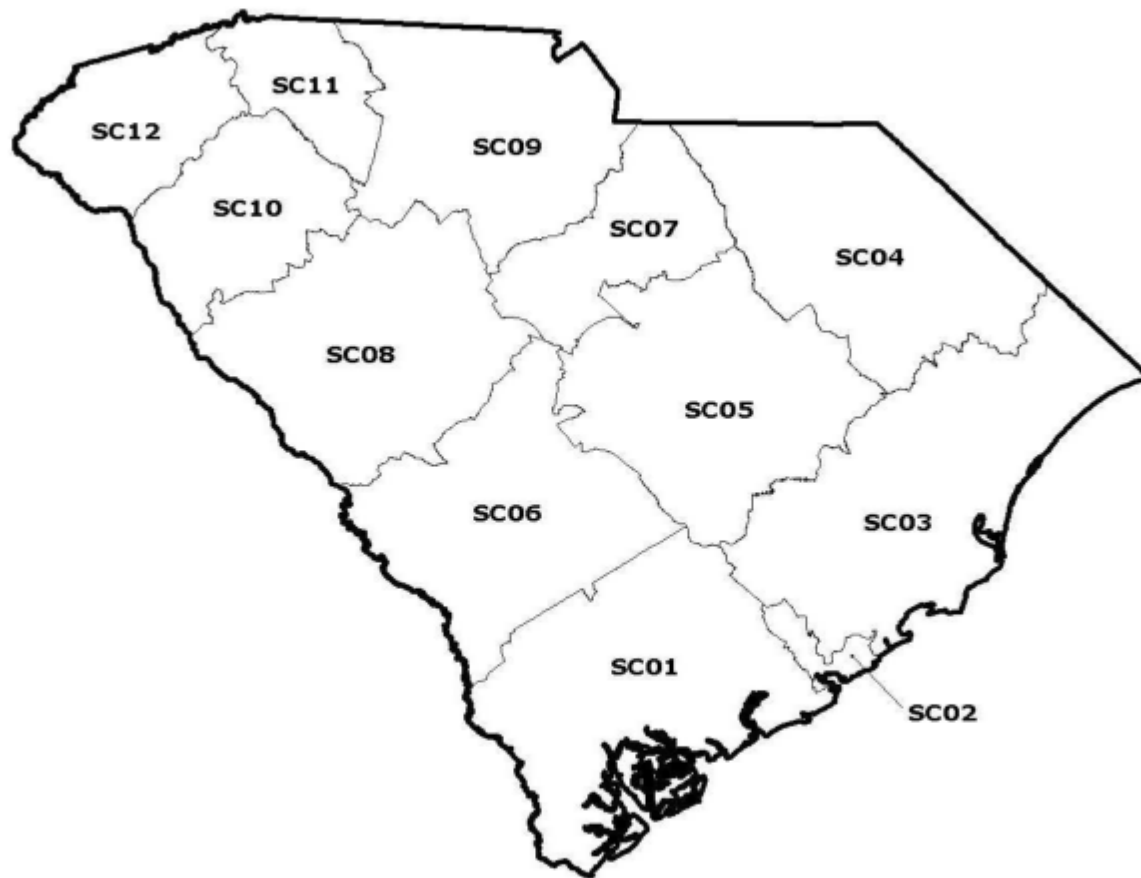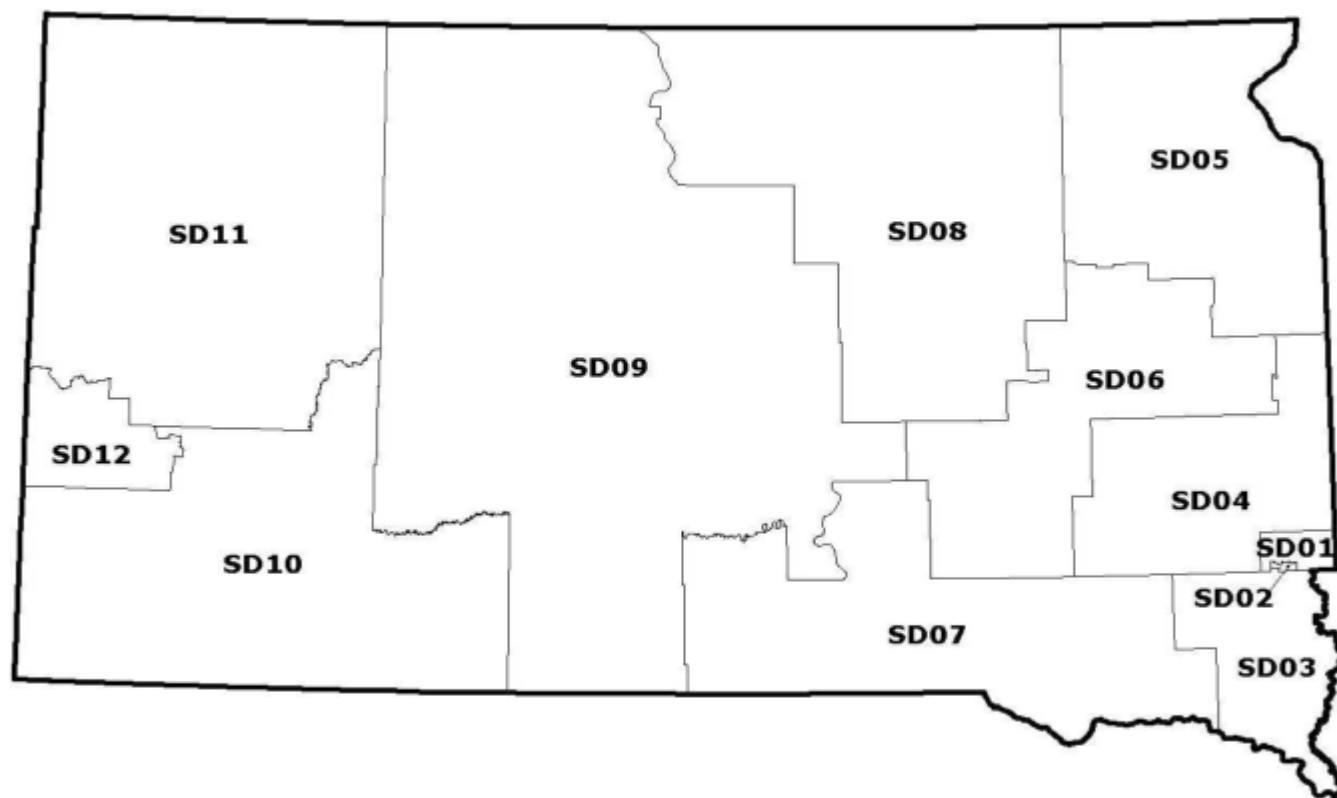
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Ohio

A-39

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Oklahoma

OK11

OK03

OK01

OK10

OK09

OK02

OK04

OK06

OK07

OK12

OK05

OK08

40   0   40   80   Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Oregon

OR01

OR02

OR03

OR05

OR04

OR09

OR10

OR07

OR06

OR11

OR12

OR08

30   0   30   60

N

Miles

A-41

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Pennsylvania

PA01

PA02

PA17

PA03

PA15

PA16

PA18

PA19

PA04

PA24

PA30

PA32

PA06

PA05

PA14

PA31

PA08

PA23

PA26

PA34

PA33

PA10

PA25

PA36

PA11

PA20

PA35

PA09

PA21

PA28

PA07

PA12

PA13

PA22

PA27

PA29

See Inset
next page

20   0   20   40

N

Miles

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Philadelphia Area

PA35

PA33

PA36

PA37

PA39

PA38

PA42

PA40

PA41

PA47

PA46

PA48

PA45

PA44

PA43

30 0 30 60

N

Miles

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Rhode Island

RI01

RI02

RI03

RI04

RI07

RI05

RI06

RI09

RI09

RI08

RI10

RI11

RI11

RI11

RI12

RI11

RI12

RI12

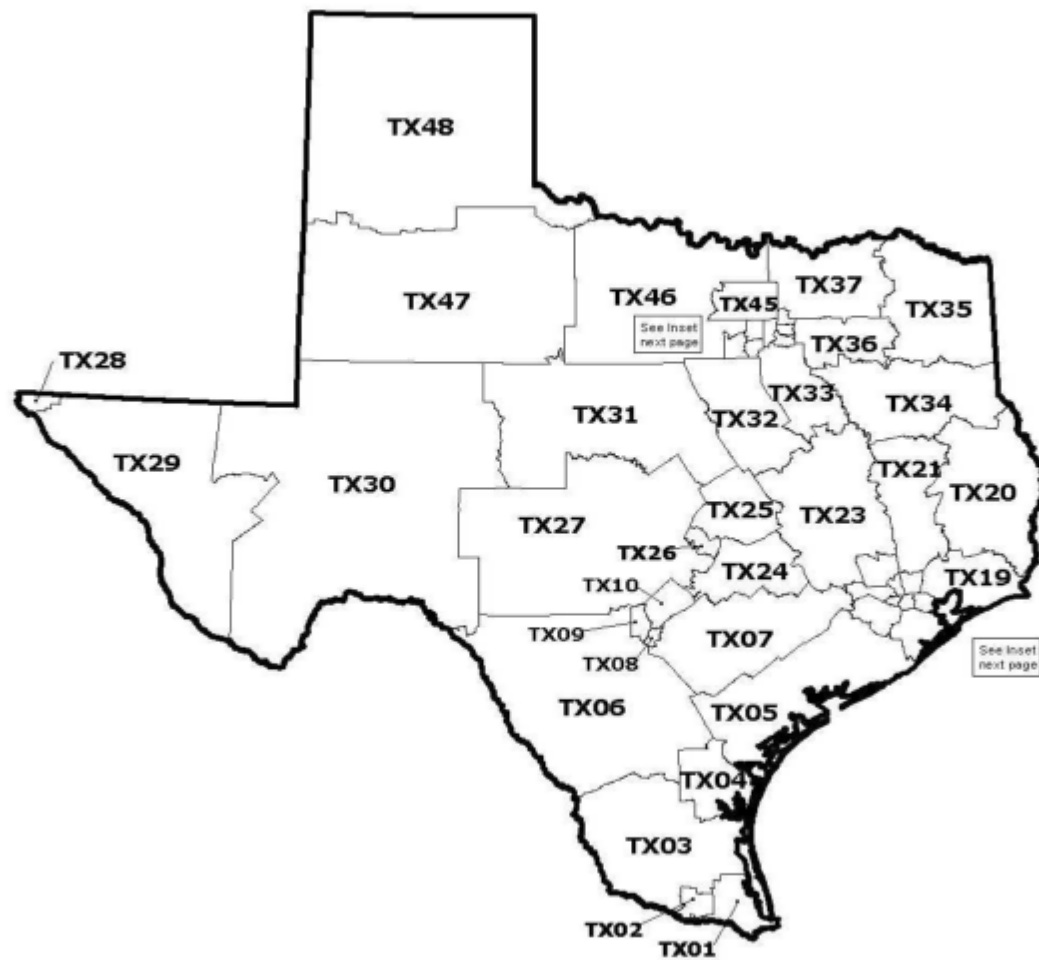3 0 3 6

Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: South Carolina

1999-2003 NHSDA/NSDUH Field Interviewer Regions: South Dakota

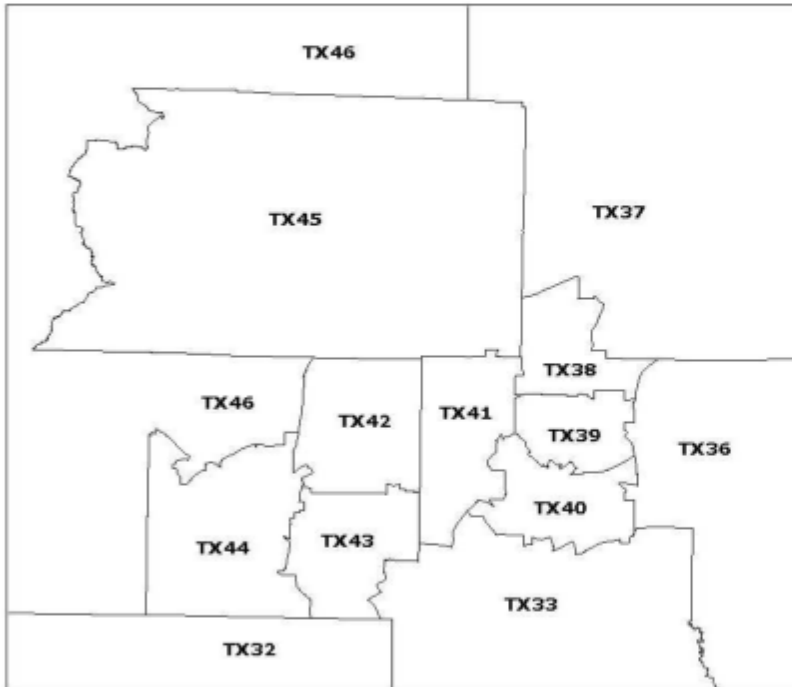# 1999-2003 NHSDA/NSDUH Field Interviewer Regions: Tennessee

TN05

TN06

TN12

TN03

TN09

TN11

TN10

TN04

TN07

TN02

TN08

TN01

40   0   40   80

Miles
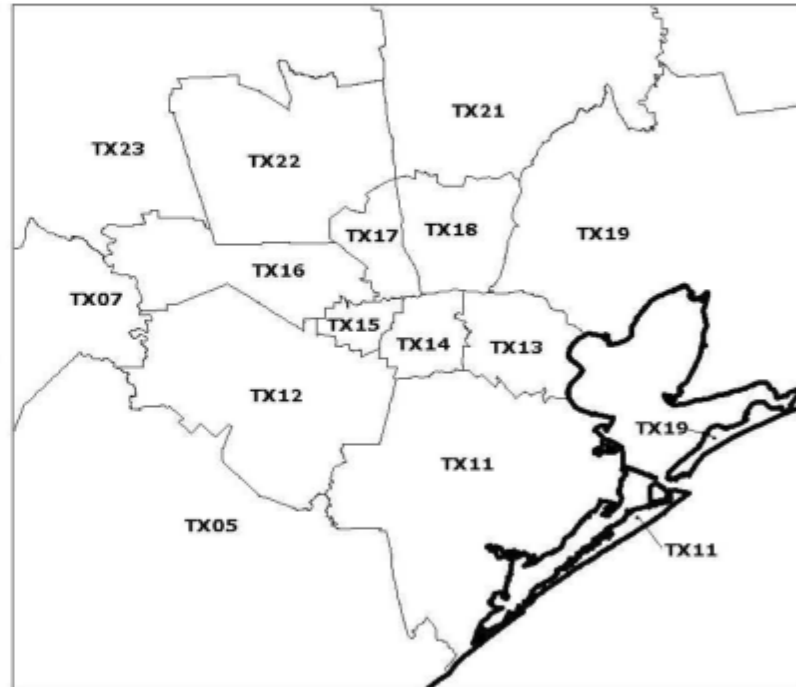
N
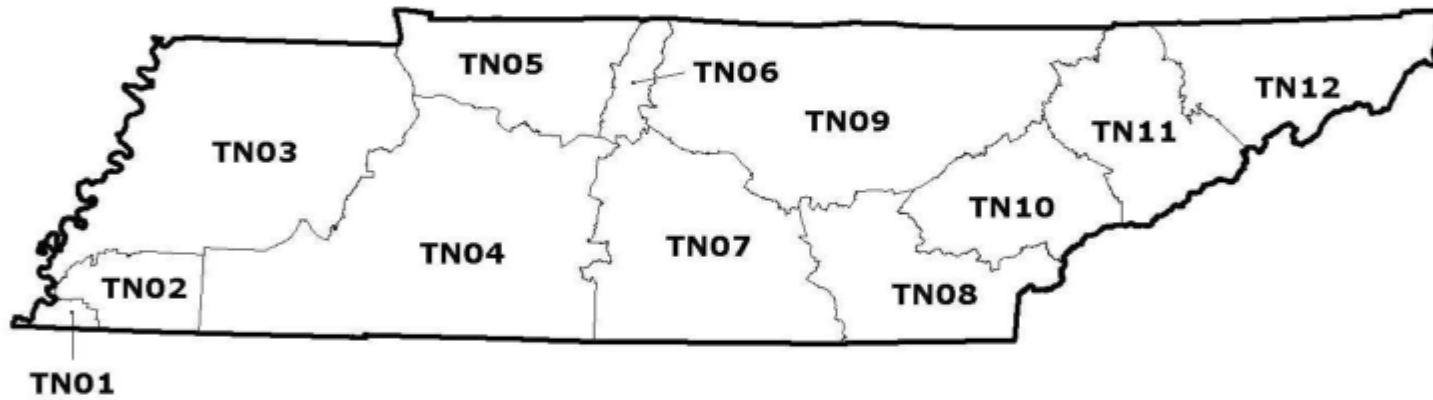
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Texas

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Texas Insets

Dallas Area

TX46

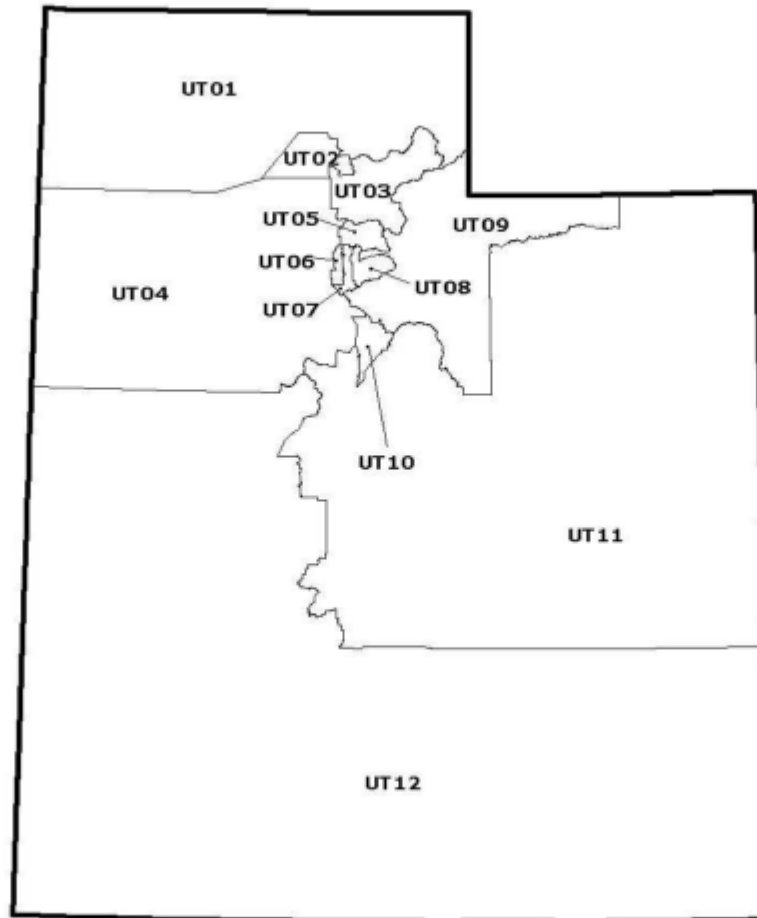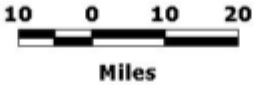TX45

TX37

TX46

TX42

TX41

TX38

TX39

TX36

TX44

TX43

TX40

TX33

TX32

10 0 10 20
Miles

Houston Area

TX23

TX22

TX21

TX17 TX18 TX19

TX16

TX07

TX15

TX14 TX13

TX12

TX19

TX11

TX05

TX11

10 0 10 20
Miles

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Utah

UT01

UT02

UT03

UT05

UT09

UT06

UT08

UT04

UT07

UT10

UT11

UT12

40   0   40   80
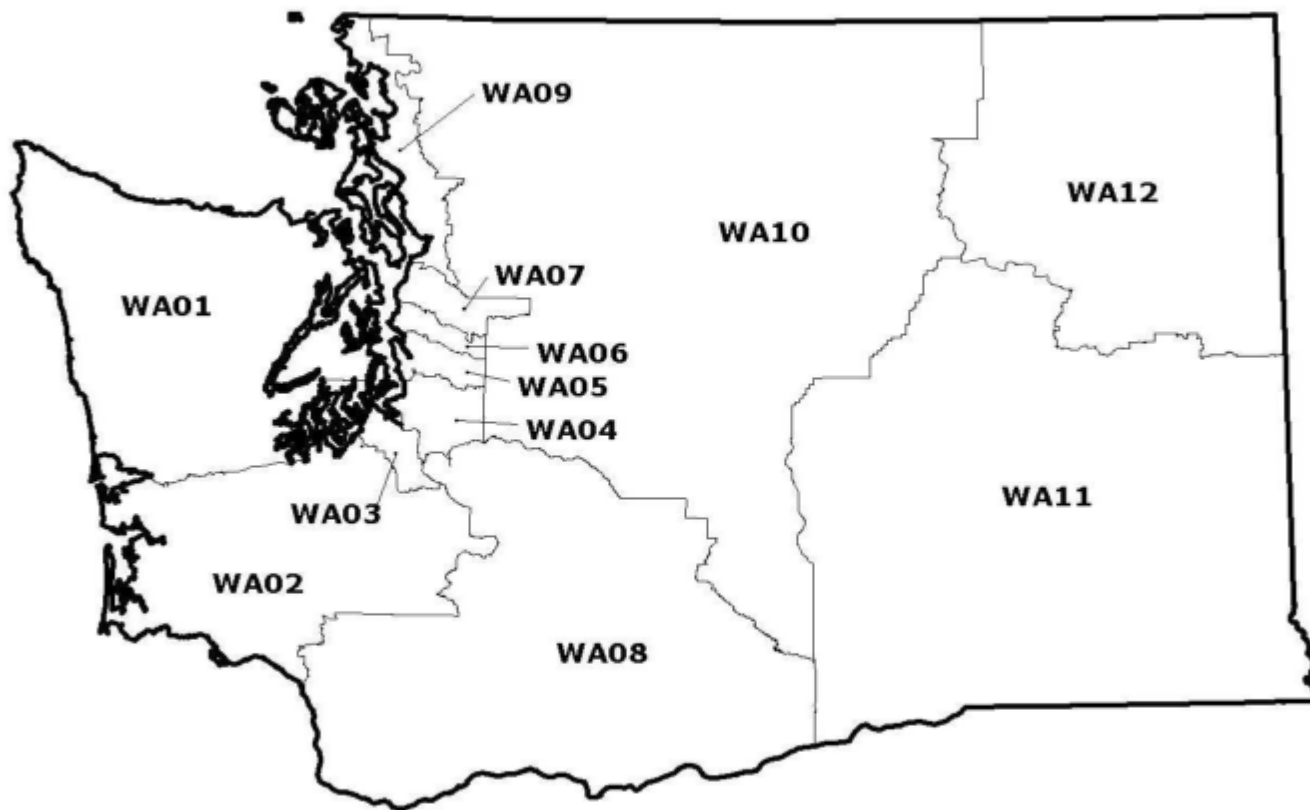
Miles

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Vermont

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Virginia
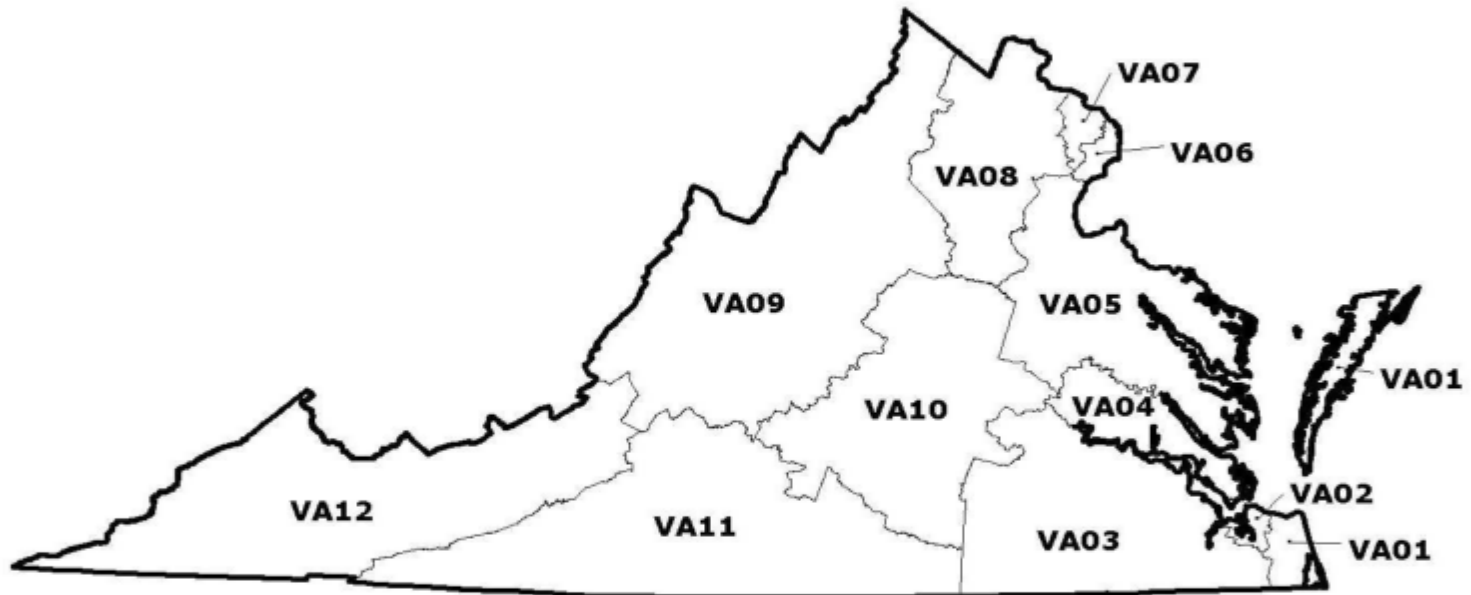
VA07

VA06

VA08

VA09

VA05

VA01

VA10

VA04

VA12

VA02

VA11

VA03

VA01

40   0   40   80

N

Miles

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Washington

WA09

WA12

WA01

WA10

WA07

WA06

WA05

WA04

WA03

WA11

WA02

WA08

30   0   30   60

Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Washington, D.C.

DC09

DC11

DC10

DC08 DC07

DC12

DC02

DC03 DC01

DC06

DC04

DC05

Miles
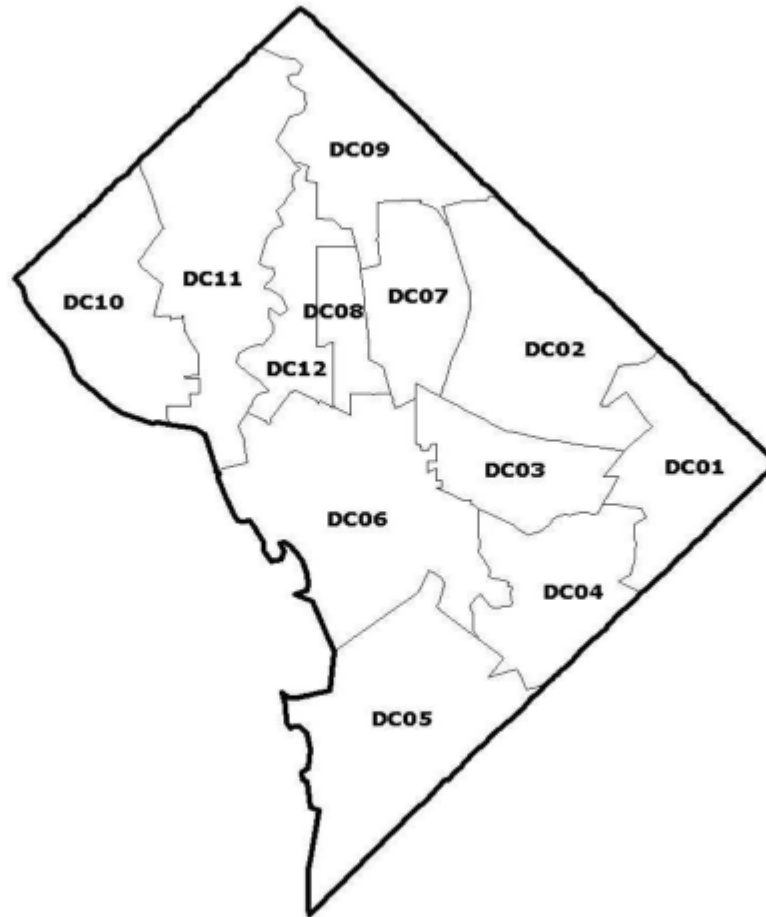
1999-2003 NHSDA/NSDUH Field Interviewer Regions: West Virginia

WV03

WV04

WV05

WV02

WV01

WV08

WV09

WV10

WV12

WV06

WV11

WV07

20   0   20   40

Miles

N

1999-2003 NHSDA/NSDUH Field Interviewer Regions: Wisconsin

WI12

WI09

WI10

WI11

WI08

WI07

WI03

WI06

WI02

WI04

WI05

WI01

20 0 20 40

Miles

N
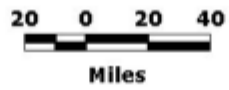
1999-2003 NHSDA/NSDUH Field Interviewer Regions: Wyoming

WY06

WY05

WY11

WY08

WY04

WY07

WY10

WY09

WY02

WY03

WY12

WY01

30   0   30   60

Miles

N

A-57
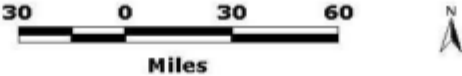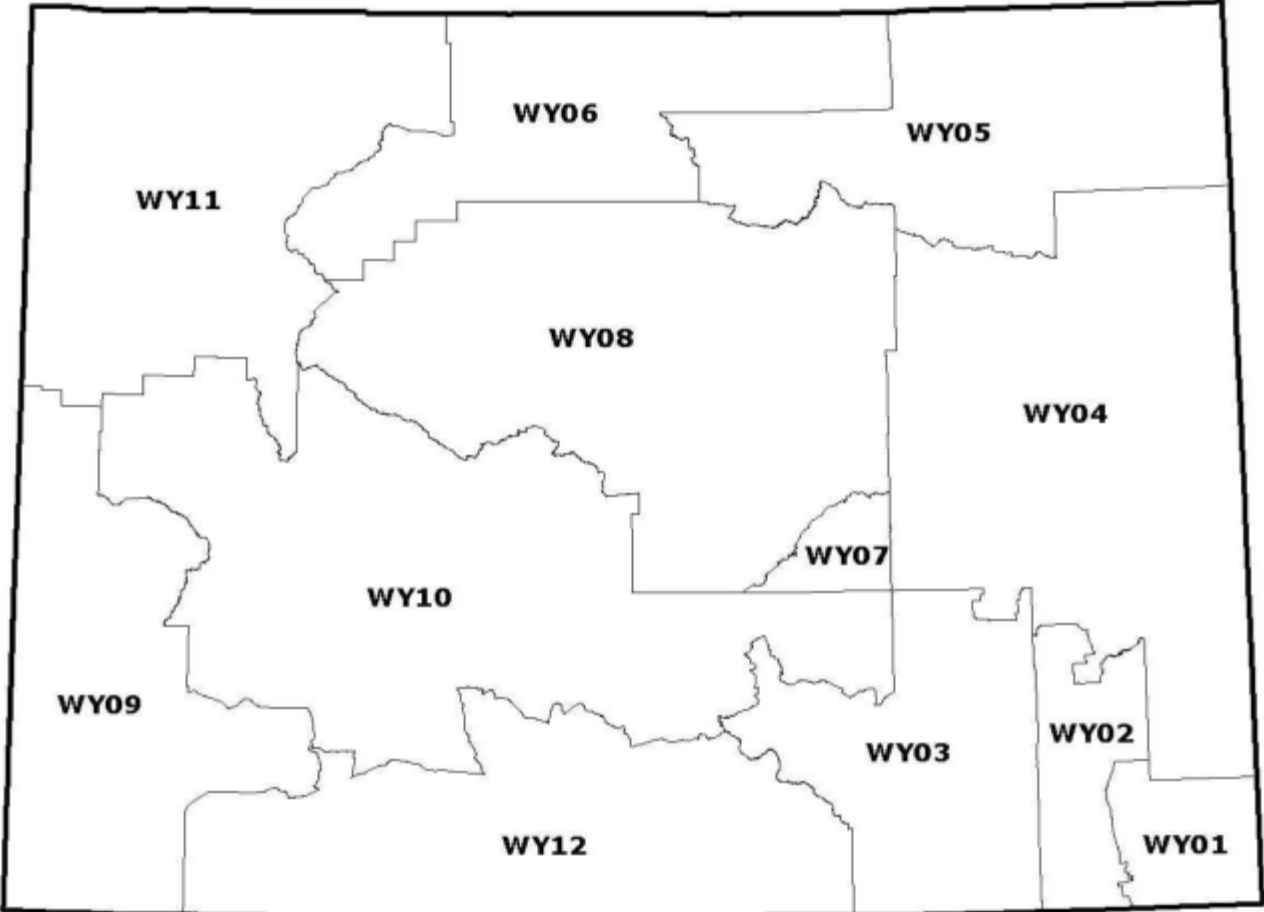
Appendix B

2003 NSDUH Procedure for Subsegmenting

1. Introduction

Subsegmenting is a statistical process used to reduce the size of the sample which reduces time and cost spent in the field for counting and listing. The precise and accurate application of subsegmenting procedures is most feasible when boundaries of subsegments can be formed using actual surface features such as streets, rivers, railroads, etc. When such features cannot be used, listing the entire area segment is considered. Because subsegmenting is a sampling function, it must be carried out with the same high degree of scientific precision exercised in the other stages of sample development.

2. Determining Subsegmenting While in the Field

If a certified lister is counting a segment and determines that the dwelling unit count is greater than 400, the segment is too large and must be subsegmented. The lister then mails the segment materials back to RTI. Once the segment is in house, standard subsegmenting procedures are followed using the street segment counts obtained by the lister.

3. Standard Subsegmenting Procedures

Once it is determined that subsegmenting is required, the following procedures are used:

Step 1   On the basis of the count, the segment is divided into areas (list units) containing not less than 100 dwelling units. If available, actual surface features are used to form new boundaries between divisions. An attempt to maintain balance between divisions is made (the largest list unit should not contain more than 1½ times the number of dwelling units contained in the smallest unit).

Step 2   After properly dividing the segment into list units, the units are lettered consecutively with Arabic capitals (A, B, C, ...) starting with the list unit including the northeast (or most appropriate) corner of the segment and continuing clockwise around the segment.

Step 3   Using a Subsegmenting Worksheet, one of the list units is randomly selected to be listed. In summary, the number of dwelling units in each list unit are recorded and accumulated. A random number generated for each segment is multiplied by the total accumulated dwelling units. The product is then rounded up, and the list unit whose cumulative dwelling units is greater than or equal to the product is selected for listing.

Once the segment materials have been returned to the field, only the selected unit is listed. All counts used in the subsegmenting process are retained so that weights can be adjusted to reflect the entire area segment.

Appendix C

2003 NSDUH Procedure for Adding Missed Dwelling Units

1.   Introduction

The 2003 National Survey on Drug Use and Health (NSDUH) requires field interviewers (FIs) to visit sample segments and screen and interview dwelling units (DUs) that were selected from an ordered list. The list of DUs, which includes housing units and group quarters, was constructed by the counting and listing staff during the summer and fall of 2001 for the overlapping segments and the summer and fall of 2002 for the replacement segments. Because the listing was done a short time before the 2003 screening and interviewing activities began, no major discrepancies were expected. However, factors such as new construction, demolition, and inaccurate listing may be present in some cases. More commonly, DUs may have been "hidden" and therefore overlooked by the counter and lister.

In order for all DUs to be given a chance of being selected, the NSDUH has a procedure for locating and adding missed DUs. It requires FIs to look on the property of selected DUs and between that DU and the next listed DU (half-open interval rule). In 2000, the rule was modified such that the half-open interval is closed on each map page. Therefore, if the selected DU is the last on a page, the "next listed DU" will be the first one listed on the same page. If the number of added DUs linked to any particular DU does not exceed five or if the number for the entire segment is less than or equal to ten, the FI is instructed to consider these DUs as part of their assignment. However, if either of these limits is exceeded, the FI will contact RTI for subsampling to be considered.

This document outlines the proposed procedures for RTI to use when discrepant segments are found in the field. For this document, procedures for adding missed DUs will be classified into three categories: adding housing units (HUs), adding group quarter units, and "busts."

2.   Motivation

Prior to the 1999 survey, if the number of added DUs exceeded the defined limits, the added DUs were subsampled at the same rate of the original selection for the segment. To maintain unequal weighting effect and to control costs associated with adding DUs, a new subsampling procedure was implemented:

Number of Added DUs	Sampling Rate
0	No action
1 to 10	Automatic (all DUs added to the sample)
11 to 25	1/2
26 to 40	1/3
41 to 50	1/4
50 or more	1/5

3.   Procedure for Adding Housing Units

This section refers to HUs that are obtained through the half-open interval rule. This method of dealing with added HUs is preferable to all others because it is probability-based and maintains the integrity of the sample. When possible, this methodology will be used to resolve added DU problems.

1.   Once the limit of five (or ten) rule is exceeded, the FI should stop screening and interviewing activities on added HUs and contact RTI. The FI will be instructed to do a quick check of the segment to see if any other listing problems might arise. At this time, the FI will complete a paper list of added HUs for the entire segment.

2.   Once the final list of added HUs has been received by RTI:

      a)   Sampling will examine the added HUs and determine whether they are linked to a sample dwelling unit (SDU);

      b)   If the number of added HUs linked to any *one* SDU exceeds 50, these units will be treated as a "bust" (see Section 6);

      c)   If the number of added HUs linked to any *one* non-sampled DU exceeds 50, these units will also be treated using the procedure for "busts" (see Section 6);

      d)   Sampling will calculate the total number of added DUs by adding the number of sampling units obtained through the "bust" procedure to the number of added DUs obtained through the half-open interval rule;

      e)   If the total number of added DUs exceeds 10, a subsampling rate will be determined using the criteria above.

3.   RTI will add the HUs to the system and subsample if necessary:

      a)   Data entry of the added HUs will be done. Lines will be entered for all units that collectively qualify as a "bust" and units obtained through the half-open interval rule—not for all missed DUs found in the segment. The link number will then be entered and a line number will be assigned. For lines obtained through the "bust" procedure, the sampling link number (SLN) will also be recorded. Finally, it will be necessary to check that none of the lines have already been entered in the Newton so that lines don't appear in the system twice.

      b)   Select lines from the added HUs at the rate defined above. Record the subsampling rate in a data field.

      c)   Bring over probabilities of selection as appropriate for the segment.

      d)   Add a random number for the Newton selection algorithm.

4. Selected lines will be added to the FI's assignment during the next transmission.

4.   Procedure for Adding Group Quarter Structures

In the case of an entire group quarter (GQ) structure not being listed (or erroneously being listed as an HU), the half-open interval rule will be applied. For example, if the DU preceding the GQ was selected, or if the HU that is really a GQ was selected, the entire GQ structure will be added to the sample. The exception to this rule will be if the number of GQ

units in the missed GQ structure exceeds 50. In this last case, the "bust" procedure will be applied (see Section 6).

5. Procedure for Adding Group Quarter Units

In the case of discrepant GQ listings, we will know in advance the number of sampling units (rooms, persons, or beds) and the number of selected units. If the actual number of sampling units equals the amount listed in advance, the Newton will only need to be notified of the new unit type in order to function properly. However, if the actual units do not equal the advance units, two approaches will be taken.

5.1. Number of Actual GQ Units Less Than Number of Advance GQ Units

In the case that there are extra GQ units listed, the units at the end of the list will be assigned an ineligible code such as "Not A DU." All other units will remain eligible.

5.2. Number of Actual GQ Units Greater Than Number of Advance GQ Units

If there are more GQ units in the structure than were previously listed, a complete list will be made and the units will be consecutively numbered. Assume, for example, that 11 units were listed and 45 were actually found. Also, assume that units 1, 5, and 10 were selected for Screening and Interviewing (indicated in bold).

Original list: **1**
2
3
4
5
6
7
8
9
10
11

Then, the additional units will be numbered consecutively and an SLN corresponding to each of the originally listed units will be assigned. Next, the added GQ units with SLNs corresponding to the original selected units will be added to the sample.

Unit Number	SLN
12	**1**
13	2
14	3
15	4
16	**5**
17	6

18	7
19	8
20	9
21	**10**
22	11
23	**1**
24	2
25	3
26	4
27	**5**
28	6
29	7
30	8
31	9
32	**10**
33	11
34	**1**
35	2
36	3
37	4
38	**5**
39	6
40	7
41	8
42	9
43	**10**
44	11
45	**1**

6.   "Busts"

Any segment listing with a major discrepancy (defined by 150 or more total unlisted units or 50 or more added DUs linked to any one SDU) or that is completely unrepresentative of what is actually found is called a "bust." In the case of a fictitious listing, RTI will relist the segment as quickly as possible. Otherwise, the following approach will be employed. First, if any DUs have disappeared since the time of the listing, all selected "disappears" will be assigned an "ineligible" final screening code. Then, any new DUs will be listed consecutively, assigned a SLN, and added to the sample if the SLN corresponds to the line number of an originally selected DU. Note that if the DU was coded as ineligible in the first step, the new DUs having its line number as the SLN will still be added. This procedure is identical to the procedure for adding extra GQ units, however the list can contain any combination of HUs and GQ units in this case. Again, if the number of DUs added is greater than 10, then resampling will occur from all non-finalized DUs as in Section 3.

7. **Quality Control**

In order to ensure quality, RTI will employ several quality control checks:

- Mapping will ensure that the correct information has been keyed by data entry,
- Checks within the computing division will be performed, and
- Sampling will check the number of selected lines and the person probabilities of selection assigned to each DU selected in the subsampling routine.