

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

SAMPLE DESIGN REPORT

Prepared for the 2012 Methodological Resource Book

RTI Project No. 0212800.001.103
Contract No. HHSS283201000003C
Phase I, Deliverable 8

Authors:

Katherine B. Morton
Peilan C. Martin
Bonnie E. Shook-Sa
James R. Chromy
Erica L. Hirsch

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

January 2013

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

SAMPLE DESIGN REPORT

Prepared for the 2012 Methodological Resource Book

RTI Project No. 0212800.001.103
Contract No. HHSS283201000003C
Phase I, Deliverable No. 8

Authors:

Katherine B. Morton
Peilan C. Martin
Bonnie E. Shook-Sa
James R. Chromy
Erica L. Hirsch

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

January 2013

Acknowledgments

This publication was developed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality (CBHSQ), by RTI International (a trade name of Research Triangle Institute), Research Triangle Park, North Carolina, under Contract No. HHSS283201000003C. Significant contributors at RTI include Katherine B. Morton, Peilan C. Martin, Bonnie E. Shook-Sa, James R. Chromy, Erica L. Hirsch, Jeremy Aldworth, Patrick Chen, David Chrest, Phillip Kott, Richard Straw, Debbie Bond, Valerie Garner, Teresa Bass, and Thomas G. Virag (Project Director).

DISCLAIMER

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

Table of Contents

Chapter	Page
List of Tables	v
List of Exhibits.....	vii
1. Overview.....	1
1.1 Target Population.....	1
1.2 Design Overview	1
1.3 5-Year Design and the 2010-2011 and 2012-2013 Sample Extensions	2
1.4 Stratification and First- and Second-Stage Sample Selections.....	2
1.5 Sample Dwelling Units and Persons.....	4
2. Extending the Coordinated 5-Year Sample	7
2.1 Formation of and Objectives for Using the Composite Size Measures	7
2.2 Stratification.....	8
2.3 First- and Second-Stage Sample Selection	9
2.4 Survey Year and Quarter Assignment	11
2.5 Creation of Variance Estimation Strata and Replicates.....	11
2.6 Other Sampling-Related Variables	13
3. General Sample Allocation Procedures for the Main Study.....	15
3.1 Notation.....	16
3.2 Determining Person Sample Sizes by State and Age Group	18
3.3 Third-Stage Sample Allocation for Each Segment.....	22
3.3.1 Dwelling Unit Frame Construction—Counting and Listing.....	22
3.3.2 Determining Dwelling Unit Sample Size	26
3.4 Determining Fourth-Stage Sample (Person) Selection Probabilities for Each Segment.....	26
3.5 Sample Size Constraints: Guaranteeing Sufficient Sample for Additional Studies and Reducing Field Interviewer Burden	26
3.6 Dwelling Unit Selection and Release Partitioning.....	27
3.7 Half-Open Interval Rule and Procedure for Adding Dwelling Units	28
3.8 Quarter-by-Quarter Deviations	33
3.9 Sample Weighting Procedures.....	35
3.9.1 Main Study Sampling Weights	35
3.9.2 Quality Control Measures in Design-Based Weighting Procedures.....	43
4. General Sample Allocation Procedures for the Mental Health Surveillance Study and Expanded Mental Health Surveillance Study	45
4.1 Respondent Universe and Sampling Methods for the Mental Health Surveillance Study and Expanded Mental Health Surveillance Study	45
4.2 Person Sample Allocation Procedures for the Mental Health Surveillance Study and Expanded Mental Health Surveillance Study	46

Table of Contents (continued)

Chapter		Page
5.	General Sample Procedures for the Questionnaire Field Test	51
5.1	Target Population.....	51
5.2	Selection of State Sampling Regions and Segments.....	51
5.3	Selection of Dwelling Units.....	53
5.4	Age Group Allocations	54
5.5	Selection of Persons.....	54
5.6	Creation of Variance Estimation Strata and Replicates.....	55
	References.....	57
Appendix		
A	2005 through 2013 NSDUH State Sampling Regions.....	59
B	2012 NSDUH Procedure for Subsegmenting	111
C	2012 NSDUH Procedure for Adding Missed Dwelling Units.....	113

List of Tables

Table	Page
2.1 Number of Census Tracts and Segments on Sampling Frame, by State.....	10
2.2 Segment Identification Number Suffixes and Quarter Assignment	12
3.1 Expected Relative Standard Errors, by Age Group	20
3.2 Expected Main Study Sample Sizes, by State and Age Group.....	20
3.3 Number of Map Pages, by State and Segment.....	23
3.4 Segment and Dwelling Unit Summary	24
3.5 Quarterly Sample Sizes and Percentages Released	29
3.6 Definitions of Levels for Potential Variables for Dwelling Unit Nonresponse Adjustment.....	37
3.7 Definitions of Levels for Potential Variables for Dwelling Unit Poststratification and Respondent Poststratification at the Person Level	38
3.8 Definitions of Levels for Potential Variables for Selected Person Poststratification and Person-Level Nonresponse Adjustment.....	39
3.9 Model Group Definitions (Census Divisions)	40
4.1 MHSS/EMHSS Age-Related Factors	46
4.2 Projected Yields of Predicted Positive Cases	47
4.3 2012 MHSS/EMHSS Sample Allocation, by K6 Group	48
4.4 2012 MHSS/EMHSS Sample Allocation, by WHODAS Score.....	48
5.1 Number of 2012 Questionnaire Field Test State Sampling Regions and Achieved Sample Sizes, by State	52
5.2 Summary of the 2012 Questionnaire Field Test Sample Results	54

List of Exhibits

Exhibit	Page
1	Flowchart of Sample Weighting Steps41
2	Sample Weight Components.....42
3	MHSS/EMHSS Sample Weight Components49

1. Overview

1.1 Target Population

The respondent universe for the 2012 National Survey on Drug Use and Health¹ (NSDUH) was the civilian, noninstitutionalized population aged 12 years or older residing within the United States and the District of Columbia. Consistent with the NSDUH designs since 1991, the 2012 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 2012 universe included those with no fixed household address (e.g., homeless and/or transient persons not in shelters) and residents of institutional group quarters, such as jails and hospitals.

1.2 Design Overview

Beginning in 1999 and continuing through subsequent years, the Substance Abuse and Mental Health Services Administration (SAMHSA) implemented major changes in the way NSDUH would be conducted. The surveys are conducted using computer-assisted interviewing (CAI) methods and provide improved State estimates based on minimum sample sizes per State. The national 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." The achieved sample for the 2012 NSDUH was 68,309 persons.

Beginning with the 2002 NSDUH and continuing through the 2012 NSDUH, survey respondents were given a \$30 incentive payment for participation. As expected, the incentive had the effect of increasing response rates, thereby requiring fewer selected households than previous surveys. In recent years, however, response rates have been slowly declining, which has required the number of selected households to increase.

An additional design change was made in 2002 and continued through 2012. A new pair-sampling strategy was implemented that increased the number of pairs selected in dwelling units (DUs) with older persons on the roster (Chromy & Penne, 2002). With the increase in the number of pairs came a moderate decrease in the response rate for older persons.

A Mental Health Surveillance Study (MHSS) was conducted to monitor the efficacy of a new mental health screening measure by conducting follow-up telephone psychiatric interviews using the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP) with selected respondents (First, Spitzer, Gibbon, & Williams,

¹ This report presents information from the 2012 National Survey on Drug Use and Health (NSDUH). Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

2002).² The MHSS (which began in 2008) was conducted as an embedded study within the 2012 NSDUH. The sample design for the MHSS is described in Chapter 4.

Finally, a Questionnaire Field Test (QFT) was conducted in late 2012 to test a redesigned questionnaire protocol for implementation in the 2015 NSDUH. Chapter 5 describes the sample design for the 2012 QFT.

1.3 5-Year Design and the 2010-2011 and 2012-2013 Sample Extensions

A coordinated sample design was developed for the 2005 through 2009 NSDUHs. The 2010-2011 and 2012-2013 samples are extensions of the 5-year sample. Although there is no planned overlap with the 1999 through 2004 samples, the coordinated design for 2005 through 2009 facilitated 50 percent overlap in second-stage units (area segments) within each successive 2-year period from 2005 through 2009. This design was intended to increase the precision of estimates in year-to-year trend analyses, using the expected positive correlation resulting from the overlapping sample between successive NSDUH years. The 2012 NSDUH main sample continues the 50 percent overlap by retaining half of the second-stage units from the 2011 survey. The 2012 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 and as a reporting variable. Eight States, referred to as the "large" States,³ had samples designed to yield 3,600 respondents per State for the 2012 main study. This sample size was considered adequate to support direct State estimates. The remaining 43 States⁴ had samples designed to yield 900 respondents per State in the 2012 main study. In these 43 States, adequate data were available to support reliable State estimates based on small area estimation (SAE) methodology. Reliable direct State estimates are also possible (in any State) by pooling multiple years of data.

1.4 Stratification and First- and Second-Stage Sample Selections

Within each State, State sampling (SS) regions were formed. Based on a composite size measure, States were geographically partitioned into roughly equal-sized regions according to population. In other words, regions were formed such that each area yielded, in expectation, roughly the same number of interviews during each data collection period. The smaller States were partitioned into 12 SS regions, whereas the 8 large States were divided into 48 SS regions. Therefore, the partitioning of the United States resulted in the formation of a total of 900 SS regions. Maps for these regions, which will be used through the 2013 NSDUH, can be found in Appendix A.

Unlike the 1999 through 2001 NHSDAs and the 2002 through 2004 NSDUHs, the first stage of selection for the 2005 through 2012 NSDUHs was census tracts.⁵ This stage was

² "DSM-IV-TR" stands for the *Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Text Revision* (American Psychiatric Association [APA], 2000).

³ The large States are California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas.

⁴ For reporting and stratification purposes, the District of Columbia is treated the same as a State, and no distinction is made in the discussion.

⁵ A census tract is a small, relatively permanent statistical subdivision of a county or equivalent entity that contains between 1,200 and 8,000 people, with an optimum size of 4,000 people (U.S. Census Bureau, Redistricting Data Office, 2009).

included to contain sample segments within a single census tract to the extent possible.⁶ In prior years, segments that crossed census tract boundaries made merging to external data sources difficult.

The first stage of selection began with the construction of an area sample frame that contained one record for each census tract in the United States. If necessary, census tracts were aggregated within SS regions until each tract⁷ had, at a minimum, 150 DUs⁸ in urban areas and 100 DUs in rural areas.⁹

Before selecting census tracts, additional implicit stratification was achieved by sorting the first-stage sampling units by a CBSA/SES¹⁰ (core-based statistical area/socioeconomic status) indicator¹¹ and by the percentage of the population that is non-Hispanic and white.¹² From this well-ordered sample frame, 48 census tracts per SS region were sequentially selected with probabilities proportionate to a composite size measure and with minimum replacement (Chromy, 1979).

Because census tracts generally exceed the minimum DU requirement, one smaller geographic region was selected within each sampled census tract. For this second stage of sampling, each selected census tract was partitioned into compact clusters¹³ of DUs by aggregating adjacent census blocks.¹⁴ 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

⁶ Some census tracts had to be aggregated in order to meet the minimum DU requirement.

⁷ For the remainder of the discussion, first-stage sampling units are referred to as "census tracts" even though each first-stage sampling unit contains one or more census tracts.

⁸ DU counts were obtained from the 2000 census data supplemented with revised population counts from Nielsen Claritas.

⁹ The basis for the differing minimum DU requirement in urban and rural areas is that it is more difficult to meet the requirement in rural areas, and 100 DUs are sufficient to support one field test and two main study samples.

¹⁰ CBSAs include metropolitan and micropolitan statistical areas as defined by the Office of Management and Budget (2003).

¹¹ Four categories are defined as (1) CBSA/low SES, (2) CBSA/high SES, (3) non-CBSA/low SES, and (4) non-CBSA/high SES. To define SES, census tract-level median rents and property values obtained from the 2000 Census Summary File 3 were given a rank (1,...,5) based on State and CBSA quintiles. The rent and value ranks then were averaged, weighted by the percentages of renter- and owner-occupied DUs, respectively. If the resulting score fell in the lower 25th percentile by State and CBSA, the area was considered "low SES"; otherwise, it was considered "high SES."

¹² Although the large sample size eliminates the need for the oversampling of specially targeted demographic subgroups as was required prior to the 1999 NHSDA, sorting by a CBSA/SES indicator and by the percentage of the population that is non-Hispanic and white ensures dispersion of the sample with respect to SES and race/ethnicity. Implicit stratification also has the potential to lower sampling error by reducing the selection of neighboring and possibly similar segments than if the selection was done completely at random.

¹³ Although the entire cluster is compact, the final sample of DUs represents a noncompact cluster. Noncompact clusters (selection from a list) differ from compact clusters in that not all units within the cluster are included in the sample. Although compact cluster designs are less costly and more stable, a noncompact cluster design was used because it provides for greater heterogeneity of dwellings within the sample. Also, social interaction (contagion) among neighboring dwellings is sometimes introduced with compact clusters (Kish, 1965).

¹⁴ A census block is a small statistical area bounded by visible features (streets, roads, streams, railroad tracks, etc.) and nonvisible boundaries (e.g., city, town, and county limits). A block group is a cluster of census blocks within the same census tract and generally contains between 300 and 6,000 people (U.S. Census Bureau, Redistricting Data Office, 2009).

or a shelter bed. Similar to census tracts, segments were formed to contain a minimum of 150 DUs in urban areas and 100 DUs in rural areas. This minimum DU requirement will support the overlapping sample design and any special supplemental samples or field tests that SAMHSA may wish to conduct.

Prior to selection, the segments were sorted in the order they were formed (i.e., geographically), and one segment was selected within each sampled census tract using Chromy's method of sequential random sampling (with probability proportionate to size and minimum replacement) (Chromy, 1979). The 48 selected segments then were randomly assigned to a survey year and quarter of data collection as described in Section 2.4.

1.5 Sample Dwelling Units and Persons

After sample segments for the 2012 NSDUH were selected, specially trained field household listers visited the areas and obtained complete and accurate lists of all eligible DUs within the sample segment boundaries. These lists served as the frames for the third stage of sample selection.

The primary objective of the third stage of sample selection (listing units) was to determine the minimum number of DUs needed in each segment to meet the targeted sample sizes for all age groups. Thus, listing unit sample sizes for the segment were determined using the age group with the largest sampling rate, which is referred to as the "driving" age group. Using 2000 census data adjusted to more recent data from Claritas, State- and age-specific sampling rates were computed. These rates then were adjusted by the segment's probability of selection; the subsegmentation inflation factor,¹⁵ 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.¹⁶ In addition to these factors, historical data from the 2010, 2011, and 2012 NSDUHs were used to compute predicted screening and interviewing response rate adjustments. The final adjusted sampling rate then was multiplied by the actual number of DUs found in the field during counting and listing activities. The product represents the segment's listing unit sample size.

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

Using a random start point and interval-based (systematic) selection, the actual listing units were selected from the segment frame. After DU selections were made, an interviewer visited each selected DU to obtain a roster of all persons residing in the DU. As in previous

¹⁵ 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. In some cases, a second-level subsegmenting was required if the census totals used in the initial subsegmenting were off and the selected subsegment was still too large for listing. The subsegmentation inflation factor accounts for reducing the size of the segment.

¹⁶ 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.

years, during the data collection period, if an interviewer encountered any new DU in a segment or found a DU that was missed during the original counting and listing activities, the new or missed dwellings were selected into the 2012 NSDUH using the half-open interval selection technique.¹⁷ This selection technique eliminates any frame bias that might be introduced because of errors and/or omissions in the counting and listing activities, and it also eliminates any bias that might be associated with using "old" segment listings.

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

One advantage of using an electronic screening instrument in NSDUH is the ability to impose a more complicated person-level selection algorithm on the fourth stage of the NSDUH design. Similar to the 1999 through 2011 designs, one feature that was included in the 2012 design was that any two survey-eligible persons within a DU had some chance of being selected (i.e., all survey-eligible pairs of persons had some nonzero chance of being selected). This design feature was of interest to NSDUH researchers because, for example, it allows analysts to examine how the drug use propensity of one individual in a family relates to the drug use propensity of another family member residing in the same DU (e.g., the relationship of drug use between a parent and his or her child).

¹⁷ In summary, this technique states that, if a DU is selected for the 2012 study and an interviewer observes any new or missed DUs between the selected DU and the DU appearing immediately after the selection on the counting and listing form, all new or missed dwellings falling in this interval will be selected. These added DUs are assigned the same probability of selection as the selected DU. If a large number of new or missed DUs are encountered (greater than 10), a sample of the new or missing DUs will be selected, and the sample weight will be adjusted accordingly. For more information, refer to Appendix C.

2. Extending the Coordinated 5-Year Sample

As was mentioned previously, the sample design was developed simultaneously for each of the 2005 through 2009 National Surveys on Drug Use and Health (NSDUHs), and the design was extended for the 2010-2011 and 2012-2013 NSDUHs. Starting with a census block-level frame, first- and second-stage sampling units (census tracts and area segments, respectively) were formed. A sufficient number of segments then were selected within sampled census tracts to support the 5-year design, several years beyond 2009, and any supplemental studies the Substance Abuse and Mental Health Services Administration (SAMHSA) chose to field.

2.1 Formation of and Objectives for Using the Composite Size Measures

The composite size measure procedure is used to obtain self-weighting¹⁸ samples for multiple domains in multistage designs. The NSDUH sample design has employed the composite size measure methodology since 1988. The 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

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

- Constrain the maximum number of selections per DU at a specified value; specifically, the total number of within-DU selections was limited 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 DUs to select from within each segment. As a consequence, all DUs within a specific stratum were selected with approximately the same probability and, therefore, approximately equalized DU sampling weights. This feature minimizes the variance inflation that results from unnecessary variation in sampling weights.
- Equalize the expected number of sample persons per cluster to balance the interviewing workload and to facilitate the assignment of interviewers to regions and segments. This feature also minimizes adverse effects on precision resulting from extreme cluster size variations.
- Simplify the size measure data requirements so that census data (block-level counts) are adequate to implement the method.

¹⁸ Self-weighting implies equal weights within domains defined by State and age group.

Using the 2000 census data supplemented with revised population projections, a composite size measure was computed for each census block defined within the United States. The composite size measure began by defining the rate $f_h(d)$ at which each age group domain d ($d = 1, \dots, 5$ for 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 years or older) was to be sampled from State h .

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

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

The composite size measure for segment j was calculated as

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

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

2.2 Stratification

Because the NSDUH design provides for estimates by State in all 50 States plus the District of Columbia, States may be viewed as the first level of stratification. The objective of the next level of stratification was to distribute the number of interviews, in expectation, equally among SS regions. Within each State, census tracts were joined to form mutually exclusive and exhaustive SS regions of approximately equal sizes. Prior to forming the SS regions, composite size measures were scaled so that the aggregate composite size measure was roughly 100 per region. This scaling made it easier for the technician when forming the regions. Without scaling, the composite size measures would sum to approximately 75 (the expected sample size per region). Using desktop computer mapping software, the regions were formed, taking into account geographical boundaries, such as mountain ranges and rivers, to the extent possible. Therefore, the resulting regions facilitated ease of access and distributed the workload evenly among regions. A total of 900 SS regions were formed for the coordinated 2005-2009 design, and these strata definitions will remain the same through the 2013 survey.

In each of the 43 small States, which include the District of Columbia (see footnote 3), 12 SS regions were formed; however, 48 SS regions were formed in each of the 8 large States (i.e., in California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas). The number of SS regions to create in the small States versus the large States was determined as follows. The design called for 300 persons in each of three age groups (12 to 17, 18 to 25, and 26 or older) equally allocated to four quarters within each small sample State. Based on an analysis of the cost variance tradeoffs, an average cluster size of 3.125 persons in each of the three age groups (or an average of 9.375 persons over the three age groups combined) was considered near optimal. When applied to the small States, a quarterly sample of 75 persons per quarter per age group could be

obtained from 24 clusters or area segments. For unbiased variance estimation purposes, at least two observations are required per stratum (Chromy, 1981); maximum geographic stratification was obtained by defining 12 strata (SS regions) with 2 area segments each per quarter. Two additional segments were selected for each of the other three quarters, yielding 8 area segments per stratum, or 96 area segments per small sample State. This approach supported a target sample size for the small States of 300 persons per age group, or a total of 900 for the year. In the large sample States, 4 times as large a sample was required. Optimum cluster size configuration and maximum stratification given the need for unbiased variance estimation were maintained by simply quadrupling the number of strata (SS regions) to 48 per large sample State, yielding a sample 300 persons per age group per quarter, 1,200 per age group over four quarters, and 3,600 per year over all three age groups.

2.3 First- and Second-Stage Sample Selection

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

To form segments within sampled census tracts, adjacent census blocks were collapsed until the total number of DUs within the area was at least 150 in urban areas and 100 in rural areas. In order to obtain geographic ordering of the blocks within tracts, block centroids were serpentine-sorted by latitude and longitude.¹⁹ If a portion of a block fell between two other blocks but its centroid did not, the block was not combined with the other two blocks, and the resulting segment contained multiple pieces. However, the majority of segments consisted of contiguous blocks.

To control the geographic distribution of the sample, segments were sorted in the order they were formed, and one segment was selected per sampled census tract using the probability proportional to size sequential sampling method. As [Table 2.1](#) indicates, 48 census tracts/segments per SS region were chosen for a total of 576 segments in each State, except in the large States where a total of 2,304 segments were chosen. Although only 24 segments per SS region were needed to support the 5-year study from 2005 through 2009, an additional 24 segments were selected to serve as replacements when segment DUs are depleted and/or to support any supplemental studies embedded within NSDUH. These 24 segments constitute the "reserve" sample and are available for use in 2010-2011 and 2012-2013 NSDUHs.

¹⁹ The latitude and longitude for each census block were obtained from the Census 2000 Summary File 1.

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

State	State Abbreviation	State FIPS Code	Number of Census Tracts on Sampling Frame	Total Number of Census Tracts/Segments Selected	Number of Segments on Sampling Frame	Number Selected for 5-Year Sample	Unique Segments in 5-Year Sample
Total U.S.			64,505	43,200	382,598		
Northeast							
Connecticut	CT	09	807	576	5,095	288	287
Maine	ME	23	343	576	3,533	288	287
Massachusetts	MA	25	1,355	576	6,163	288	288
New Hampshire	NH	33	272	576	3,076	288	286
New Jersey	NJ	34	1,914	576	5,657	288	288
New York	NY	36	4,738	2,304	19,057	1,152	1,149
Pennsylvania	PA	42	3,088	2,304	21,704	1,152	1,150
Rhode Island	RI	44	233	576	2,305	288	283
Vermont	VT	50	179	576	1,648	288	285
Midwest							
Illinois	IL	17	2,901	2,304	20,733	1,152	1,147
Indiana	IN	18	1,408	576	6,863	288	287
Iowa	IA	19	790	576	5,366	288	288
Kansas	KS	20	719	576	5,120	288	288
Michigan	MI	26	2,689	2,304	18,765	1,152	1,148
Minnesota	MN	27	1,293	576	5,955	288	288
Missouri	MO	29	1,303	576	7,193	288	287
Nebraska	NE	31	495	576	4,075	288	288
North Dakota	ND	38	215	576	1,618	288	279
Ohio	OH	39	2,902	2,304	20,342	1,152	1,149
South Dakota	SD	46	212	576	2,001	288	284
Wisconsin	WI	55	1,310	576	6,773	288	288
South							
Alabama	AL	01	1,079	576	6,958	288	288
Arkansas	AR	05	618	576	6,128	288	288
Delaware	DE	10	196	576	1,721	288	282
District of Columbia	DC	11	179	576	1,049	288	270
Florida	FL	12	3,140	2,304	25,374	1,152	1,150
Georgia	GA	13	1,609	576	7,682	288	288
Kentucky	KY	21	992	576	6,301	288	288
Louisiana	LA	22	1,099	576	5,841	288	288
Maryland	MD	24	1,204	576	5,477	288	288
Mississippi	MS	28	601	576	6,448	288	287
North Carolina	NC	37	1,550	576	8,708	288	287
Oklahoma	OK	40	977	576	5,654	288	286
South Carolina	SC	45	862	576	7,365	288	288
Tennessee	TN	47	1,246	576	7,534	288	288
Texas	TX	48	4,351	2,304	26,096	1,152	1,152
Virginia	VA	51	1,513	576	6,448	288	286
West Virginia	WV	54	466	576	4,319	288	287

(continued)

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

State	State Abbreviation	State FIPS Code	Number of Census Tracts on Sampling Frame	Total Number of Census Tracts/Segments Selected	Number of Segments on Sampling Frame	Number Selected for 5-Year Sample	Unique Segments in 5-Year Sample
West							
Alaska	AK	02	154	576	1,348	288	283
Arizona	AZ	04	1,089	576	6,759	288	287
California	CA	06	6,978	2,304	22,973	1,152	1,152
Colorado	CO	08	1,050	576	6,231	288	288
Hawaii	HI	15	274	576	1,784	288	285
Idaho	ID	16	277	576	3,224	288	285
Montana	MT	30	256	576	2,417	288	288
Nevada	NV	32	474	576	3,919	288	288
New Mexico	NM	35	431	576	3,839	288	288
Oregon	OR	41	752	576	6,219	288	288
Utah	UT	49	485	576	4,024	288	288
Washington	WA	53	1,312	576	6,425	288	288
Wyoming	WY	56	125	576	1,291	288	283

FIPS = Federal information processing standards.

2.4 Survey Year and Quarter Assignment

The 48 sampled segments per SS region were randomly assigned to survey years by drawing equal probability subsamples of 4 segments. Prior to selecting the second subsample, the first subsample segments were removed from the pool of eligible segments. The second subsample then was selected from the remaining segments. This process was initially repeated 5 times until the 48 sampled segments were assigned to 6 subsamples of 4 and a "reserve" sample of 24 segments; for the 2010-2011 surveys, 2 additional subsamples of 4 segments were selected from the reserve sample. Similarly, for the 2012-2013 surveys, 2 more subsamples of 4 segments were chosen from the reserve sample.

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

Using the survey year and quarter assignments, a sequential segment identification number (SEGID) then was assigned. [Table 2.2](#) describes the relationship between SEGIDs and quarter assignment. The last two digits in the SEGID are called the "segment suffix." The 2012 main survey corresponds to segment suffixes 29 through 36.

2.5 Creation of Variance Estimation Strata and Replicates

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

Table 2.2 Segment Identification Number Suffixes and Quarter Assignment

Segment Suffix	2005 NSDUH	2006 NSDUH	2007 NSDUH	2008 NSDUH	2009 NSDUH	2010 NSDUH	2011 NSDUH	2012 NSDUH	2013 NSDUH	Variance Replicate
01	x (Q1)									1
02	x (Q2)									1
03	x (Q3)									1
04	x (Q4)									1
05	x (Q1)	x (Q1)								2
06	x (Q2)	x (Q2)								2
07	x (Q3)	x (Q3)								2
08	x (Q4)	x (Q4)								2
09		x (Q1)	x (Q1)							1
10		x (Q2)	x (Q2)							1
11		x (Q3)	x (Q3)							1
12		x (Q4)	x (Q4)							1
13			x (Q1)	x (Q1)						2
14			x (Q2)	x (Q2)						2
15			x (Q3)	x (Q3)						2
16			x (Q4)	x (Q4)						2
17				x (Q1)	x (Q1)					1
18				x (Q2)	x (Q2)					1
19				x (Q3)	x (Q3)					1
20				x (Q4)	x (Q4)					1
21					x (Q1)	x (Q1)				2
22					x (Q2)	x (Q2)				2
23					x (Q3)	x (Q3)				2
24					x (Q4)	x (Q4)				2
25						x (Q1)	x (Q1)			1
26						x (Q2)	x (Q2)			1
27						x (Q3)	x (Q3)			1
28						x (Q4)	x (Q4)			1
29							x (Q1)	x (Q1)		2
30							x (Q2)	x (Q2)		2
31							x (Q3)	x (Q3)		2
32							x (Q4)	x (Q4)		2
33								x (Q1)	x (Q1)	1
34								x (Q2)	x (Q2)	1
35								x (Q3)	x (Q3)	1
36								x (Q4)	x (Q4)	1
37									x (Q1)	2
38									x (Q2)	2
39									x (Q3)	2
40									x (Q4)	2

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

representing the variance estimation strata and replicates were created to capture explicit stratification and to identify clustering. For the 2005 through 2013 NSDUHs, variance estimation strata are defined at the SS region level, and each SS region appears in a different stratum every quarter. Because both census tracts and segments are nested within variance replicates, the variance contributions of both sampling units are covered by the nesting variables. Because one segment is selected per sampled census tract, the selection of census tracts at the first stage of selection may reduce variance by controlling the sample distribution and minimizing the chance of selecting neighboring and possibly similar segments within the same census tract.

To define the variance estimation strata for the 2005 through 2013 NSDUHs, the 900 SS regions were first placed in random order (States were randomly sorted, and regions were

randomly sorted within States). This list, numbered 1 to 900, defined the quarter 1 variance estimation strata (VESTRQ1). For quarter 2, the variance estimation strata, VESTRQ2, were defined as $VESTRQ1 - 150$ (or $VESTRQ1 - 150 + 900$ if $VESTRQ1 \leq 150$). Similarly, $VESTRQ3 = VESTRQ2 - 150$ (+ 900 if $VESTRQ2 \leq 150$), and $VESTRQ4 = VESTRQ3 - 150$ (+ 900 if $VESTRQ3 \leq 150$). As an example, an SS region that was assigned to stratum 151 in quarter 1 was assigned to stratum 1 (= 151 - 150) in quarter 2, stratum 751 (= 1 - 150 + 900) in quarter 3, and stratum 601 (= 751 - 150) in quarter 4. This method had the effect of assigning the regions to strata in a pseudo-random fashion while ensuring that each stratum consists of four SS regions from four different States.

The 2005 through 2013 definition of variance estimation strata has the effect of increasing the number of degrees of freedom for State-level estimates while preserving the number of degrees of freedom for national estimates (900). Each small sample State is in 48 different strata (12 SS regions \times 4 quarters); therefore, there are 48 degrees of freedom available for State estimates. Similarly, each large sample State is in 192 strata (48 SS regions \times 4 quarters) and therefore has 192 degrees of freedom for estimation.

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

In addition to variance estimation strata and replicates, a sample weight is computed for each final respondent (see Section 3.9.1). The use of sample weights in analyses of NSDUH data is necessary to properly represent the target population and to account for disproportionate sampling by age group. All weighted statistical analyses for which variance estimates are needed should use the stratum and replicate variables to identify nesting. Variance estimates can be computed using a clustered data analysis software package such as SUDAAN[®] (RTI International, 2008). The SUDAAN software package computes variance estimates for nonlinear statistics using such procedures as a first-order Taylor series approximation of the deviations of estimates from their expected values. The approximation is unbiased for sufficiently large samples. SUDAAN also recognizes positive covariance among estimates involving data from 2 or more years.²⁰ Using data from the 2007 and 2008 NSDUHs and examining multiple measures, the average relative change in the standard error (SE) after accounting for covariance was about 1 percent.

2.6 Other Sampling-Related Variables

Because area segments consist of one or more census blocks, a number of demographic and geographic variables are available for sampled areas. The demographic data include the following: population counts by age, race, and ethnicity; estimated civilian, noninstitutional population aged 12 or older; DU counts; estimated group quarters units; and group quarters population by type of

²⁰ Using the variance estimation strata and replicates, SUDAAN recognizes positive covariance among estimates from consecutive years. For nonconsecutive years, strata are treated as collapsing with zero covariance.

group quarter.²¹ For these variables, the block-level data were aggregated to form segment-level estimates.

The U.S. Census Bureau also makes available several geographic variables that can be associated with the 2005 through 2013 NSDUH sample segments. These are State, county and county name, census tract (equal to the most frequently occurring tract if the segment crosses multiple tracts), place name, census division and region, land area, CBSA/SES indicator (as defined in Section 2.3), county-level population density, and a rural or urban indicator.²² Each census block is assigned a rural or urban status based on population density and/or proximity to a census-designated urbanized area (UA) or urban cluster (UC). In the NSDUH sample, if one or more of the blocks within a segment is urban, the segment is defined as urban. If 100 percent of the blocks are rural, the segment is defined as rural. Defining rural or urban status in this way provides an aggregate variable that is needed for assigning minimum size requirements (see Section 2.3) and for some data analyses. However, the definition slightly overestimates the urban population.

The 2005 through 2013 NSDUH sample was designed to facilitate matching to external data at the census tract level. Because field enumeration of the sample segments occurs at the segment level rather than the block level (see Section 3.3.1), only the group of blocks in which a NSDUH respondent resides is known. Therefore, specific census blocks are not linked to respondents or listing units. Because there is no direct linking of listing units to specific census blocks, no mechanism currently is available for assigning block-level data to NSDUH respondents. For this reason, some variables (e.g., place name and rural or urban indicator) may have some error associated with them.

²¹ Data were obtained or derived from the Census 2000 Summary File 1 and adjusted using revised population counts from Claritas.

²² All variables were obtained or derived from the Census 2000 Summary File 1.

3. General Sample Allocation Procedures for the Main Study

In this chapter, the computational details of the procedural steps used to determine both person and dwelling unit (DU) sample sizes are discussed. The within-DU age group-specific selection probabilities for the design of the 2012 National Survey on Drug Use and Health (NSDUH) also are addressed. This optimization procedure was designed specifically to address the Substance Abuse and Mental Health Services Administration's (SAMHSA's) multiple precision and design requirements while simultaneously minimizing the cost of data collection. Costs were minimized by determining the smallest number of interviews and selected DUs necessary to achieve the various design requirements. In summary, this three-step optimization procedure proceeded as follows:

1. In the first step, the optimal number of interviews (i.e., responding persons) by domains of interest needed to satisfy the precision requirements was determined for several drug use outcome measures. In other words, 255 unknown m_{ha} values for each State h (51) and age group a (5) were initially sought to be determined. A solution to this multiple constraint optimization was achieved using Chromy's Algorithm (Chromy, 1987). This is described in further detail in Section 3.2.
2. Using the m_{ha} determined from Step 1, the next step was to determine the optimal number of selected dwelling (D_{hj}) units (i.e., third-stage sample) that were necessary. This step was achieved by applying parameter constraints (e.g., probabilities of selection and expected response rates) at the segment level j or the stage at which DUs would be selected, which was done on a quarterly basis using approximately 25 percent of the m_{ha} values. This step is described in further detail in Section 3.3.
3. The final step in this procedure entailed determining age group-specific probabilities of selection (S_{hja}) for each segment given the m_{ha} and D_{hj} from Steps 1 and 2. This was achieved using a modification of Brewer's Method of Selection (Cochran, 1977, pp. 261-263). The modification was designed to select 0, 1, or 2 persons from each DU.²³ A detailed discussion of the final step is given in Section 3.4. After calculating the required DUs and the selection probabilities, sample size constraints were applied to ensure adequate samples for supplemental studies and to reduce the field interviewer (FI) burden. Limits on the total number of expected interviews per segment also were applied. This process became iterative to reallocate the reduction in sample size to other segments not affected by such constraints. Details of this step in the optimization procedure are given in Section 3.5.

²³ Direct application of Brewer's method would require a fixed sample size.

3.1 Notation

h = 50 States plus the District of Columbia.

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

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

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

y_{ha} = Estimated number of persons in the target population in State h and age group a . The 2012 population is estimated using the 2000 census data adjusted to the 2007 Claritas population projections in the compound interest formula, $y = Ae^{Bx}$, where

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

First, B is computed as $[\ln(y/A)]/x$, where y = the population in 2007, A = the population in 2000, and $x = 7$. Then the 2012 population (y^*_{ha}) is computed using the original formula and this time allowing x to be 12. Finally, the 2012 population is adjusted by the ratio of estimated eligible listed DUs to the Claritas DU counts (U_{hj}). This adjustment factor considers the number of added DUs expected to be obtained through the half-open interval rule (1.01) and the probability of a DU being eligible (ϵ_h), both determined via historic data. The coefficient adjustment of 1.01 is estimated using historical data and is the proportion of all screened DUs (includes added DUs) over the original total of selected DUs (excluding added DUs). So, $y_{ha} = \{[1.01 * \epsilon_h * L_{hj} * (1/I_{hj}) / U_{hj}]\} * y^*_{ha}$, where ϵ_h , L_{hj} , and I_{hj} are defined further below. This adjustment is computed at the census block level and then aggregated to the State level.

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

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

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

I_{hj} = Subsegmentation inflation factor. For segments too large to count and to list efficiently in both time and cost, field listing personnel may request that a portion of the segment be

randomly sampled. First, they perform a quick count (best guess: $L_{1\ hj}^*$) of the entire segment. The sampling staff then subdivides the segment into roughly equal-sized subdivisions or subsegments (using a best guess estimate of the number of DUs in each subsegment: $B_{1\ hj}^*$) and selects one for regular counting and listing. Beginning in 2008, some large segments were subsegmented based on census information prior to being sent to the field for listing. In some of these segments, the selected subsegment was still too large for listing, and a second round of subsegmenting was required. The second-level subsegmenting was performed in a similar fashion as the first-level subsegmenting, in that the first-level subsegment was counted (best guess: $L_{2\ hj}^*$), and subdivided into roughly equal-sized subdivisions or subsegments (best guess: $B_{2\ hj}^*$). Then, one subsegment was selected for regular counting and listing by sampling staff. For the subsegment to represent the entire segment, the weights are adjusted up to reflect the unused portion of the segment.

= $(B_{1\ hj}^* / L_{1\ hj}^*)$, if one round of subsegmenting was done.

= $(B_{1\ hj}^* / L_{1\ hj}^*) * (B_{2\ hj}^* / L_{2\ hj}^*)$, if two rounds of subsegmenting were required.

= 1, if no subsegmenting was done.

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

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

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

ε_h = State-specific DU eligibility rate. This rate was derived from 2010 NSDUH quarters 3 and 4 and 2011 NSDUH quarters 1 and 2 data by taking the average eligibility rate within each State.

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

λ_{ha} = State- and age group-specific interview response rate. Using data from quarters 3 and 4 of the 2010 NSDUH and quarters 1 and 2 of the 2011 NSDUH, the additive effects of State and age group on interview response were determined by taking the average interview response rate within each State.

γ_{ha} = Expected number of persons within an age group per DU. This number was calculated using 2010 NSDUH quarters 3 and 4 and 2011 NSDUH quarters 1 and 2 data by dividing the weighted total number of rostered persons in an age group by the weighted total number of complete screened DUs by State.

δ_{ha} = State- and age group-specific maximum-of-two rule adjustment. The survey design restricts the number of interviews per DU to a total of two. This is achieved through a

modified Brewer's Method of Selection, which results in a loss of potential interviews in DUs where selection probabilities sum greater than 2. The adjustment is designed to inflate the number of required DUs to compensate for this loss. Using data from all four quarters of the 2010 NSDUH, the adjustment was computed by taking the average maximum-of-two rule adjustment within each State.

3.2 Determining Person Sample Sizes by State and Age Group

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

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

In preparation for the 2005 through 2009 NSDUHs and the 2010-2011 and 2012-2013 NSDUHs, several optimization models and other related analyses were conducted. Using historical 2001 survey data, estimates and RSEs for each of nine outcome measures of interest were computed. Estimates then were standardized to a prevalence of 10 percent. The outcome measures of interest were included to address not only the NSDUH recency-of-use estimates, but also such related generic substance abuse measures as treatment received for alcohol and illicit drug use and dependence on alcohol and illicit drugs.

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

Use of Legal (Licit) Substances

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

Use of Illicit Substances

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

Note that current use of illicit drugs provides a broad measure of illicit drug use; however, it is dominated by marijuana and cocaine use. Therefore, estimates of illicit drug use other

than marijuana use and cocaine use are included because these two measures reflect different types of drug abuse.

Drug or Alcohol Dependence

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

Treatment for Drugs and Alcohol Problems

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

These outcome measures, as well as the precision that is expected from this 2012 NSDUH design, are presented in [Table 3.1](#), which was updated using 2010 NSDUH data. RSEs were based on an average prevalence rate of 10 percent for each measure.

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.

As in the 1999 through 2011 surveys, racial groups were not oversampled for the 2012 NSDUH. Consistent with previous surveys, the 2012 NSDUH was designed to oversample the younger age groups.

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

Table 3.1 Expected Relative Standard Errors, by Age Group

Outcome Measure	Total			12-17		
	Estimate	RSE	SRSE	Estimate	RSE	SRSE
Past Month Cigarette Use	22.97	1.34	2.20	8.34	3.09	2.79
Past Month Alcohol Use	51.79	0.75	2.34	13.59	2.43	2.89
Past Month Use of Illicit Drugs	8.92	2.16	2.02	10.13	2.85	2.87
Past Month Use of Illicit Drugs Other Than Marijuana	3.56	3.32	1.91	4.50	4.31	2.81
Past Month Cocaine Use	0.58	7.34	1.68	0.25	18.52	2.77
Past Year, Dependent on Illicit Drugs	1.90	4.23	1.77	2.53	5.46	2.64
Past Year, Dependent on Alcohol but Not Illicit Drugs	2.81	4.07	2.08	1.10	9.25	2.92
Past Year, Received Treatment for Illicit Drug Use	0.86	6.37	1.78	0.84	9.91	2.74
Past Year, Received Treatment for Alcohol Use but Not for Illicit Drug Use	0.61	8.93	2.10	0.14	23.86	2.70
Average RSE	N/A	N/A	1.99	N/A	N/A	2.79
Target RSE	N/A	N/A	3.00	N/A	N/A	5.00
Outcome Measure	18-25			26+		
	Estimate	RSE	SRSE	Estimate	RSE	SRSE
Past Month Cigarette Use	34.24	1.37	2.97	22.83	1.68	2.74
Past Month Alcohol Use	61.53	0.81	3.07	54.85	0.88	2.92
Past Month Use of Illicit Drugs	21.51	1.85	2.90	6.57	3.44	2.74
Past Month Use of Illicit Drugs Other Than Marijuana	7.95	3.09	2.73	2.67	5.33	2.65
Past Month Cocaine Use	1.46	7.79	2.84	0.47	10.86	2.23
Past Year, Dependent on Illicit Drugs	5.37	4.13	2.95	1.22	7.70	2.56
Past Year, Dependent on Alcohol but Not Illicit Drugs	5.11	4.09	2.84	2.63	5.50	2.71
Past Year, Received Treatment for Illicit Drug Use	1.65	6.98	2.72	0.73	9.19	2.36
Past Year, Received Treatment for Alcohol Use but Not for Illicit Drug Use	0.87	10.02	2.82	0.62	11.10	2.64
Average RSE	N/A	N/A	2.87	N/A	N/A	2.62
Target RSE	N/A	N/A	5.00	N/A	N/A	5.00

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

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

Region/State	State FIPS	SS Regions	Total Segments	Total Respondents					
				12-17	18-25	26-34	35-49	50+	Total
Total Population		900	7,200	22,500	22,500	6,000	9,000	7,500	67,500
Northeast									
Connecticut	09	12	96	300	300	70	127	103	900
Maine	23	12	96	300	300	65	121	114	900
Massachusetts	25	12	96	300	300	77	124	99	900
New Hampshire	33	12	96	300	300	67	129	104	900
New Jersey	34	12	96	300	300	73	128	99	900
New York	36	48	384	1,200	1,200	312	491	397	3,600
Pennsylvania	42	48	384	1,200	1,200	283	479	438	3,600
Rhode Island	44	12	96	300	300	74	122	104	900
Vermont	50	12	96	300	300	68	120	112	900

(continued)

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

Region/State	State FIPS	SS Regions	Total Segments	Total Respondents					Total
				12-17	18-25	26-34	35-49	50+	
Midwest									
Illinois	17	48	384	1,200	1,200	332	489	379	3,600
Indiana	18	12	96	300	300	82	120	98	900
Iowa	19	12	96	300	300	76	117	107	900
Kansas	20	12	96	300	300	81	118	101	900
Michigan	26	48	384	1,200	1,200	297	487	416	3,600
Minnesota	27	12	96	300	300	81	122	97	900
Missouri	29	12	96	300	300	81	118	101	900
Nebraska	31	12	96	300	300	82	117	101	900
North Dakota	38	12	96	300	300	79	112	109	900
Ohio	39	48	384	1,200	1,200	307	478	415	3,600
South Dakota	46	12	96	300	300	79	113	108	900
Wisconsin	55	12	96	300	300	77	121	102	900
South									
Alabama	01	12	96	300	300	79	118	103	900
Arkansas	05	12	96	300	300	81	116	103	900
Delaware	10	12	96	300	300	75	121	104	900
District of Columbia	11	12	96	300	300	106	110	84	900
Florida	12	48	384	1,200	1,200	294	463	443	3,600
Georgia	13	12	96	300	300	85	128	87	900
Kentucky	21	12	96	300	300	81	119	100	900
Louisiana	22	12	96	300	300	84	117	99	900
Maryland	24	12	96	300	300	78	126	96	900
Mississippi	28	12	96	300	300	81	118	101	900
North Carolina	37	12	96	300	300	78	125	97	900
Oklahoma	40	12	96	300	300	84	114	102	900
South Carolina	45	12	96	300	300	76	120	104	900
Tennessee	47	12	96	300	300	80	120	100	900
Texas	48	48	384	1,200	1,200	359	496	345	3,600
Virginia	51	12	96	300	300	80	124	96	900
West Virginia	54	12	96	300	300	75	113	112	900
West									
Alaska	02	12	96	300	300	90	123	87	900
Arizona	04	12	96	300	300	87	117	96	900
California	06	48	384	1,200	1,200	343	500	357	3,600
Colorado	08	12	96	300	300	88	123	89	900
Hawaii	15	12	96	300	300	80	115	105	900
Idaho	16	12	96	300	300	87	115	98	900
Montana	30	12	96	300	300	76	111	113	900
Nevada	32	12	96	300	300	87	122	91	900
New Mexico	35	12	96	300	300	85	114	101	900
Oregon	41	12	96	300	300	83	114	103	900
Utah	49	12	96	300	300	108	113	79	900
Washington	53	12	96	300	300	83	120	97	900
Wyoming	56	12	96	300	300	84	112	104	900

FIPS = Federal information processing standards; SS = State sampling.

3.3 Third-Stage Sample Allocation for Each Segment

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

3.3.1 Dwelling Unit Frame Construction—Counting and Listing

The process by which the DU frame is constructed is called counting and listing. In summary, a certified lister visits the selected area and lists a detailed and accurate address (or description, if no address is available) for each DU within the segment boundaries. The lister is given a series of maps on which to mark the locations of these DUs. Map pages are formed so that the lister can easily navigate the segment and has sufficient space to denote the location of each DU. The number of map pages depends on the size and composition of the segment. In general, a sparsely populated rural segment has more map pages than a densely populated urban segment. Thus, segments in States like New York and Nevada have fewer map pages on average, while segments in States like South Dakota are much larger on average. The number of map pages per State and the average number of map pages per segment are summarized in [Table 3.3](#). The list of DUs constructed during counting and listing is entered into a database and serves as the frame from which the third-stage sample is drawn.

In some situations, the number of DUs within the segment boundaries was much larger than the specified maximum. To obtain a reasonable number of DUs for the frame, the lister first counted the DUs in such an area. The sampling staff then partitioned the segment into smaller pieces or subsegments and randomly selected one to be listed. Beginning in 2008, some large segments were partitioned into subsegments using census information prior to being sent to the field. Sampling staff then randomly selected one subsegment to send to the field for listing. In a few of these cases, additional subsegmenting was required for one of the following reasons: (1) the area experienced high growth and the census counts used in the initial subsegment were outdated, or (2) there was not enough information available during the first subsegment, and the initial subsegment was still too large to list. Thus, an additional level of subsegmenting was implemented to make listing feasible. The number of segments that were subsegmented in the 2012 NSDUH sample is summarized in [Table 3.4](#). For more information on the subsegmenting procedures, see Appendix B.

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	41,465	5.8
Alabama	96	591	6.2
Alaska	96	583	6.1
Arizona	96	550	5.7
Arkansas	96	656	6.8
California	384	1,363	3.5
Colorado	96	500	5.2
Connecticut	96	385	4.0
Delaware	96	458	4.8
District of Columbia	96	285	3.0
Florida	384	1,861	4.8
Georgia	96	544	5.7
Hawaii	96	370	3.9
Idaho	96	775	8.1
Illinois	384	2,180	5.7
Indiana	96	562	5.9
Iowa	96	836	8.7
Kansas	96	696	7.3
Kentucky	96	548	5.7
Louisiana	96	605	6.3
Maine	96	629	6.6
Maryland	96	374	3.9
Massachusetts	96	426	4.4
Michigan	384	1,985	5.2
Minnesota	96	586	6.1
Mississippi	96	687	7.2
Missouri	96	631	6.6
Montana	96	988	10.3
Nebraska	96	892	9.3
Nevada	96	416	4.3
New Hampshire	96	542	5.6
New Jersey	96	413	4.3
New Mexico	96	967	10.1
New York	384	1,551	4.0
North Carolina	96	503	5.2
North Dakota	96	1,092	11.4
Ohio	384	1,905	5.0
Oklahoma	96	675	7.0
Oregon	96	513	5.3
Pennsylvania	384	2,394	6.2
Rhode Island	96	458	4.8
South Carolina	96	601	6.3
South Dakota	96	878	9.1
Tennessee	96	561	5.8
Texas	384	2,088	5.4

(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	502	5.2
Vermont	96	634	6.6
Virginia	96	405	4.2
Washington	96	576	6.0
West Virginia	96	620	6.5
Wisconsin	96	581	6.1
Wyoming	96	1,044	10.9

Table 3.4 Segment and Dwelling Unit Summary

State	Total Segments	Total Subsegmented Segments	Second-Level Subsegmented Segments	Listed Dwelling Units	Sampled Dwelling Units	Added Dwelling Units
Total Population	7,200	644	2	1,543,451	212,746	1,528
Alabama	96	11	0	21,612	3,002	10
Alaska	96	13	0	21,565	2,388	36
Arizona	96	9	0	19,898	2,768	3
Arkansas	96	6	0	18,729	2,760	16
California	384	32	0	82,192	9,443	46
Colorado	96	8	0	20,921	3,059	12
Connecticut	96	8	0	21,042	2,832	23
Delaware	96	6	0	22,824	2,831	16
District of Columbia	96	20	1	29,800	4,974	81
Florida	384	66	0	87,798	12,734	34
Georgia	96	8	0	21,521	2,350	15
Hawaii	96	17	0	26,928	3,143	69
Idaho	96	6	0	18,235	2,297	3
Illinois	384	17	0	78,363	11,320	65
Indiana	96	4	0	18,432	2,482	9
Iowa	96	6	0	17,844	2,494	35
Kansas	96	6	0	18,700	2,577	21
Kentucky	96	5	0	21,453	2,828	24
Louisiana	96	7	0	21,122	2,731	10
Maine	96	8	0	20,528	3,807	59
Maryland	96	13	0	24,014	2,662	18
Massachusetts	96	11	0	20,890	3,014	50
Michigan	384	28	1	79,513	11,344	97
Minnesota	96	7	0	18,588	2,460	23
Mississippi	96	5	0	17,372	2,544	9
Missouri	96	10	0	19,043	2,855	24
Montana	96	15	0	17,084	3,281	14
Nebraska	96	6	0	17,758	2,545	11

(continued)

Table 3.4 Segment and Dwelling Unit Summary (continued)

State	Total Segments	Total Subsegmented Segments	Second-Level Subsegmented Segments	Listed Dwelling Units	Sampled Dwelling Units	Added Dwelling Units
Nevada	96	13	0	20,864	2,348	6
New Hampshire	96	4	0	19,775	2,905	85
New Jersey	96	7	0	21,390	2,608	14
New Mexico	96	13	0	18,605	2,768	3
New York	384	51	0	92,887	14,429	118
North Carolina	96	9	0	22,294	2,831	17
North Dakota	96	9	0	18,132	3,357	17
Ohio	384	23	0	80,173	11,664	58
Oklahoma	96	8	0	19,621	2,952	8
Oregon	96	7	0	19,902	2,535	12
Pennsylvania	384	23	0	78,670	11,843	64
Rhode Island	96	2	0	20,144	2,593	27
South Carolina	96	4	0	20,454	3,282	24
South Dakota	96	5	0	19,393	2,617	19
Tennessee	96	10	0	20,136	2,521	11
Texas	384	43	0	85,899	9,003	45
Utah	96	6	0	20,250	1,783	10
Vermont	96	5	0	19,440	3,250	42
Virginia	96	15	0	21,945	2,546	30
Washington	96	12	0	21,144	2,678	22
West Virginia	96	3	0	20,670	3,188	34
Wisconsin	96	5	0	17,308	2,431	9
Wyoming	96	9	0	20,586	3,089	20

During counting and listing, the lister moves about the segment in a prescribed fashion called the "continuous path of travel." Beginning from a starting point noted on the map,²⁴ the lister attempts to move in a clockwise fashion, makes each possible right turn, makes U-turns at segment boundaries, and does not break street sections. Within apartment buildings and group quarters, the lister attempts to apply the same rules; that is, the lister moves in a clockwise fashion and enumerates building floors from bottom to top. Following these defined rules and always looking for DUs on the right-hand side of the street (or hall), 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. A detailed description of the counting and listing procedures is provided in the 2012 counting and listing general manual (RTI International, 2011).

²⁴ Sampling staff review each map and determine the most logical starting point. They choose an intersection of two boundaries of the segment that seems most appropriate considering the segment's composition.

3.3.2 Determining Dwelling Unit Sample Size

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

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

At this point in the procedure, only two components in the formula are unknown: D_{hj} and S_{hja} . Selection probabilities are segment- and age group-specific, and to maximize the number of selected persons within a 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 (see Section 1.5), is set to the largest allowable selection probability (S_{hja}) of 0.99. D_{hj} then is computed as

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

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

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

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

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

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

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

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

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

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

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

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

Similarly, the interview adjustment is computed as

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

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

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

3.6 Dwelling Unit Selection and Release Partitioning

After derivation of the required DU sample size within each State and segment (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. Systematic sampling creates a heterogeneous sample of DUs by dispersing the sample throughout the segment. In addition, it minimizes social contagion from neighboring selected DUs that could have an impact on response rates and prevalence estimates. The listing order was used to approximate geographic location because a standard address is not available for all listed DUs.

To compensate for quarterly variations in response rates and yields, a sample partitioning procedure was implemented in all quarters. The entire sample (D_{hj}) still would be selected, but only certain percentages of the total would be released into the field. An initial percentage would be released in all segments at the beginning of the quarter. Based on interquarter work projections, additional percentages would be released 1 month into the quarter as needed and if field staff could handle the added workload. Each partitioning of the sample is a valid sample and helps manage the sample sizes by State without jeopardizing the validity of the study. Incidentally, a reserve sample of 20 percent also was selected, over and above the required D_{hj} sample, to allow for supplemental releases based on State experiences within each quarter. Thus, the 96 percent quarter 1 sample (see Section 3.2) was increased to the 115.2 percent level (i.e., $0.96 * 1.20 = 1.152$). In quarter 1, the D_{hj} sample was allocated out to States in the following release percentages:

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

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

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

As described in Section 3.9, a weight adjustment is applied to all DUs within a segment to account for the partial release of sample. The DU release adjustment (weight component #7) is equal to the inverse of the percentage of the sample that is released into the field. For example, if only DUs in release 1 were made available to the field, the DU release adjustment would equal $120/80$ or 1.5. If releases 1, 4, and 5 were fielded, the adjustment would equal $120/100$ or 1.2 because $80/120 + 10/120 + 10/120 = 100/120$. 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, a procedure called the half-open interval rule was implemented. This procedure required that the interviewer look both on the property of each selected DU and between that DU and the next listed DU for any unlisted units. When found in these specific locations, the unlisted units became part of the sample (added DUs). If the number of added DUs linked to any particular sample DU did not exceed 5, or if the number for the entire segment was less than or equal to 10, the FI was instructed to consider these DUs as part of his or her assignment. If either of these limits was exceeded, special subsampling procedures were implemented, as described in Appendix C. The number of added DUs in the 2012 NSDUH sample is summarized in the last column of [Table 3.4](#).

Table 3.5 Quarterly Sample Sizes and Percentages Released

Region/State	Quarter 1			Quarter 2		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
Total Population	59,483	49,564	83	65,082	56,482	87
Northeast						
Connecticut	784	654	83	845	601	71
Maine	1,057	885	84	1,204	953	79
Massachusetts	972	805	83	800	762	95
New Hampshire	842	704	84	880	770	88
New Jersey	692	574	83	778	778	100
New York	4,100	3,413	83	4,731	3,942	83
Pennsylvania	2,952	2,456	83	3,886	3,556	92
Rhode Island	742	621	84	853	602	71
Vermont	972	811	83	991	863	87
Midwest						
Illinois	3,174	2,651	84	3,518	3,082	88
Indiana	717	597	83	759	659	87
Iowa	760	634	83	723	660	91
Kansas	739	616	83	832	624	75
Michigan	3,252	2,716	84	3,515	3,365	96
Minnesota	765	633	83	800	565	71
Missouri	726	610	84	804	803	100
Nebraska	698	580	83	783	655	84
North Dakota	902	752	83	1,019	934	92
Ohio	3,129	2,603	83	3,547	3,094	87
South Dakota	677	559	83	729	699	96
Wisconsin	770	642	83	825	651	79

(continued)

Table 3.5 Quarterly Sample Sizes and Percentages Released (continued)

Region/State	Quarter 1			Quarter 2		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
South						
Alabama	867	719	83	890	890	100
Arkansas	753	628	83	835	766	92
Delaware	838	695	83	908	683	75
District of Columbia	1,477	1,228	83	1,605	1,137	71
Florida	3,737	3,117	83	3,787	3,470	92
Georgia	641	535	83	692	606	88
Kentucky	707	592	84	835	835	100
Louisiana	732	614	84	823	749	91
Maryland	777	644	83	742	619	83
Mississippi	709	597	84	810	612	76
North Carolina	780	646	83	838	803	96
Oklahoma	685	567	83	960	801	83
South Carolina	815	680	83	901	862	96
Tennessee	696	583	84	827	721	87
Texas	2,545	2,124	83	2,675	2,338	87
Virginia	781	655	84	824	687	83
West Virginia	878	731	83	966	965	100
West						
Alaska	753	623	83	667	532	80
Arizona	895	745	83	857	573	67
California	2,690	2,237	83	2,971	2,347	79
Colorado	806	671	83	880	762	87
Hawaii	779	650	83	902	900	100
Idaho	629	527	84	603	603	100
Montana	902	755	84	967	844	87
Nevada	589	487	83	596	472	79
New Mexico	710	593	84	809	706	87
Oregon	769	644	84	726	667	92
Utah	476	394	83	508	485	95
Washington	834	695	83	939	628	67
Wyoming	811	672	83	917	801	87

(continued)

Table 3.5 Quarterly Sample Sizes and Percentages Released (continued)

Region/State	Quarter 3			Quarter 4		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
Total Population	64,283	55,816	87	59,644	50,884	85
Northeast						
Connecticut	824	823	100	754	754	100
Maine	1,154	963	83	1,006	1,006	100
Massachusetts	770	769	100	712	678	95
New Hampshire	889	782	88	860	649	75
New Jersey	898	671	75	741	585	79
New York	4,379	3,650	83	4,103	3,424	83
Pennsylvania	3,758	3,124	83	3,423	2,707	79
Rhode Island	866	758	88	868	612	71
Vermont	1,093	911	83	889	665	75
Midwest						
Illinois	3,562	3,110	87	3,134	2,477	79
Indiana	755	628	83	716	598	84
Iowa	775	651	84	692	549	79
Kansas	767	731	95	694	606	87
Michigan	3,463	2,892	84	3,167	2,371	75
Minnesota	706	647	92	615	615	100
Missouri	873	685	78	790	757	96
Nebraska	844	703	83	811	607	75
North Dakota	953	913	96	911	758	83
Ohio	3,432	3,004	88	3,232	2,963	92
South Dakota	728	637	88	722	722	100
Wisconsin	754	596	79	689	542	79

(continued)

Table 3.5 Quarterly Sample Sizes and Percentages Released (continued)

Region/State	Quarter 3			Quarter 4		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
South						
Alabama	807	771	96	783	622	79
Arkansas	839	668	80	765	698	91
Delaware	947	829	88	792	624	79
District of Columbia	1,544	1,285	83	1,670	1,324	79
Florida	3,692	3,082	83	3,501	3,065	88
Georgia	670	640	96	717	569	79
Kentucky	814	711	87	821	690	84
Louisiana	809	702	87	799	666	83
Maryland	742	708	95	691	691	100
Mississippi	831	758	91	817	577	71
North Carolina	815	712	87	806	670	83
Oklahoma	905	828	91	823	756	92
South Carolina	942	942	100	833	798	96
Tennessee	775	584	75	760	633	83
Texas	2,628	2,197	84	2,452	2,344	96
Virginia	737	581	79	652	623	96
West Virginia	992	663	67	829	829	100
West						
Alaska	672	644	96	589	589	100
Arizona	774	646	83	804	804	100
California	3,060	2,681	88	2,611	2,178	83
Colorado	911	911	100	865	715	83
Hawaii	993	869	88	913	724	79
Idaho	642	642	100	604	525	87
Montana	882	882	100	800	800	100
Nevada	838	734	88	824	655	79
New Mexico	830	796	96	734	673	92
Oregon	709	652	92	684	572	84
Utah	461	459	100	483	445	92
Washington	863	749	87	809	606	75
Wyoming	916	842	92	884	774	88

3.8 Quarter-by-Quarter Deviations

This section describes corrections and/or modifications that were implemented in the process of design optimization. "Design" refers to deviations from the original proposed plan of design. "Procedural" refers to changes made in the calculation methodologies. Finally, "Dwelling Unit Selection" addresses changes that occurred after sample size derivations, specifically corrections implemented during fielding of the sample (i.e., sample partitioning as described in Section 3.6). Quarter 1 deviations are not included because the methods and procedures described above were all implemented in quarter 1. Subsequently, any changes would have been made after quarter 1.

Quarter 2

- Design: An additional 20 percent reserve sample was added to the 104 percent quarterly sample to allow for supplemental releases where needed. Thus, the total quarter 2 sample was increased to the 124.8 percent level.
- Procedural: To predict State response rates more accurately, the most current four quarters of data were used in the computation of State-specific yield and response rates. Thus, data from quarters 1 through 4 of the 2011 NSDUH were used to compute average yields, DU eligibility, screening response, and interviewer response rates.
- Dwelling Unit Selection: The quarter 2 D_{hj} sample was partitioned into the following release percentages:
- Release 1:* 67 percent of entire sample (80/120, main sample + 20 percent reserve);
 - Release 2:* 4 percent of entire sample (5/120, main sample + 20 percent reserve);
 - Release 3:* 4 percent of entire sample (5/120, main sample + 20 percent reserve);
 - Release 4:* 8 percent of entire sample (10/120, main sample + 20 percent reserve);
 - Release 5:* 8 percent of entire sample (10/120, main sample + 20 percent reserve); and
 - Release 6:* 8 percent of entire sample (10/120, main sample + 20 percent reserve).

Quarter 3

- Design:** Using the completed cases from quarter 1 and the projected number of completes from quarter 2, each State's midyear surplus/shortfall was computed. The quarter 3 104 percent sample then was adjusted by this amount. An additional 20 percent sample also was included, bringing the total quarter 3 adjusted sample to the 124.8 percent level.
- Procedural:** Data from quarters 2 through 4 of the 2011 NSDUH and quarter 1 of the 2012 NSDUH were used to compute State-specific average yields, DU eligibility, screening response, and interviewer response rates.
- Dwelling Unit Selection:** The quarter 3 D_{hj} sample was partitioned into the following release percentages:
- Release 1:* 67 percent of entire sample (80/120, main sample + 20 percent reserve);
 - Release 2:* 4 percent of entire sample (5/120, main sample + 20 percent reserve);
 - Release 3:* 4 percent of entire sample (5/120, main sample + 20 percent reserve);
 - Release 4:* 8 percent of entire sample (10/120, main sample + 20 percent reserve);
 - Release 5:* 8 percent of entire sample (10/120, main sample + 20 percent reserve); and
 - Release 6:* 8 percent of entire sample (10/120, main sample + 20 percent reserve).

Quarter 4

- Design:** The State and age 96 percent quarterly sample sizes were adjusted to meet the yearly targets based on completed cases from quarters 1 and 2 and the projected number of completes from quarter 3. An additional 20 percent sample also was included, bringing the total quarter 4 adjusted sample to the 115.2 percent level.
- Procedural:** Data from quarters 3 and 4 of the 2011 NSDUH and quarters 1 and 2 of the 2012 NSDUH were used to compute State-specific average yields, DU eligibility, screening response, and interviewer response rates.
- Dwelling Unit Selection:** The quarter 4 D_{hj} sample was partitioned into the following release percentages:

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

3.9 Sample Weighting Procedures

At the conclusion of data collection for the last quarter, design weights are constructed for each quarter of the State-level study, reflecting the various stages of sampling. At the time this report was published, the weights for the 2012 NSDUH had not yet been computed. However, the planned procedures are described in this section.

3.9.1 Main Study Sampling Weights

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

Stage 1: Selection of census tract.

Stage 2: Selection of segment.

Stage 3: Selection of DU.

Three possible adjustments exist with this stage of selection:

- (1) subsegmentation inflation: by-product of counting and listing (includes up to two levels of subsegmenting);
- (2) added DU: results from the half-open interval rule when subsampling is needed; and
- (3) release adjustment.

Stage 4: Selection of person within a DU.

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

1. *Nonresponse Adjustment at the Dwelling Unit Level.* This adjustment is to account for the failure to complete the within-DU roster. The potential list of variables for the 51-State main study DU nonresponse modeling is presented in [Table 3.6](#).
2. *Dwelling Unit-Level Poststratification.* This adjustment involves using screener data of demographic information (e.g., age, race, gender). DU weights will be adjusted to the intercensal population estimates derived from the 2000 census for various demographic domains. In short, explanatory variables used during modeling will consist of counts of eligible persons within each DU that fall into the various demographic categories. Consequently, these counts, multiplied by the newly adjusted DU weight and summed across all DUs for various domains, will add to the census population estimates. This adjustment is useful for providing more stable control totals for subsequent adjustments and pair weights. Potential explanatory variables are listed in [Table 3.7](#).
3. *Extreme Weight Treatment at the Dwelling Unit Level.* If it is determined that design-based weights (stages 1 and 2) along with any of their respective adjustments result in an unsatisfactory unequal weighting effect (i.e., variance of the DU-level weights is too high, with high frequency of extreme weights), then extreme weights will be further adjusted. This adjustment will be implemented by doing another weight calibration. The control totals are the DU-level poststratified weights, and the same explanatory variables as in DU-level poststratification will be used so that the extreme weights are controlled and all the distributions in various demographic groups are preserved.
4. *Selected Person Weight Adjustment for Poststratification to Roster Data.* This step utilizes control totals derived from the DU roster that are already poststratified to the census population estimates. This adjustment assists in bias reduction and improved precision by taking advantage of the properties of a two-phase design. Selected person sample weights (i.e., those that have been adjusted at the DU level and account for fourth-stage sampling) are adjusted to the DU weight sums of all eligible rostered persons. Any demographic information used in modeling is based solely on screener information because this is the only information available for all rostered persons. Potential explanatory variables for this adjustment are a combination of the variables presented in [Table 3.8](#).
5. *Person-Level Nonresponse Adjustment.* This adjustment allows for the correction of weights resulting from the failure of selected sample persons to complete the interview. Respondent sample weights will be adjusted to the weight of all selected persons. Again, demographic information used in modeling is based solely on screener information. Potential explanatory variables for this adjustment are a combination of the variables presented in [Table 3.8](#).
6. *Person-Level Poststratification.* This step is to adjust the final person sample weights to the census population estimates derived from the 2000 census. These are the same outside control totals used in the second adjustment. However, demographic variables for this adjustment are based on questionnaire data, not screener data as in adjustments 2, 4, and 5. Potential explanatory variables used in modeling are presented in [Table 3.7](#).
7. *Extreme Weight Treatment at the Person Level.* This adjustment will be implemented in the same manner as described in adjustment 3, except that the weights reflect the fourth stage of selection.

All weight adjustments for the 2012 main study's final analysis weights will be derived from a GEM technique. To help reduce computational burden at all adjustment steps, separate models will be fit for clusters of States, based on census division definitions as shown in [Table 3.9](#). Furthermore, model variable selection at each adjustment will be done using a combination method of forward and backward selection processes. The forward selection will be used for the model enlargement. Within each enlargement, backward selection will be used. The final adjusted weight, which is the product of weight components 1 through 15, is the analysis weight used in estimation. [Exhibit 1](#) presents a flowchart of steps used in the weighting process, and [Exhibit 2](#) displays all individual weight components.

Full details of the finalized modeling procedures, as well as final variables used in each adjustment step, will be described in the forthcoming report on the person-level sampling weight calibration for the 2012 NSDUH (Chen et al., in press).

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

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

CBSA = core-based statistical area.

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

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

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

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

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

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

Group Quarters Indicator	1: College Dorm 2: Other Group Quarters 3: Nongroup Quarters
Percentage of Owner-Occupied Dwelling Units in Segment (% Owner)	1: 0 - <10% 2: 10% - <50% 3: 50% - 100%
Percentage of Blacks in Segment (% Black)	1: 0 - <10% 2: 10% - <50% 3: 50% - 100%
Percentage of Hispanics in Segment (% Hispanic)	1: 0 - <10% 2: 10% - <50% 3: 50% - 100%
Population Density	1: CBSA > 1,000,000 2: CBSA < 1,000,000 3: Non-CBSA Urban 4: Non-CBSA Rural
Quarter	1: Quarter 1 2: Quarter 2 3: Quarter 3 4: Quarter 4
Segment Combined Median Rent and Housing Value (Rent/Housing)	1: First Quintile 2: Second Quintile 3: Third Quintile 4: Fourth Quintile 5: Fifth Quintile
State	
Age	1: 12-17 2: 18-25 3: 26-34 4: 35-49 5: 50+
Gender	1: Male 2: Female

(continued)

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

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

CBSA = core-based statistical area.

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

Table 3.9 Model Group Definitions (Census Divisions)

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

Exhibit 1 Flowchart of Sample Weighting Steps

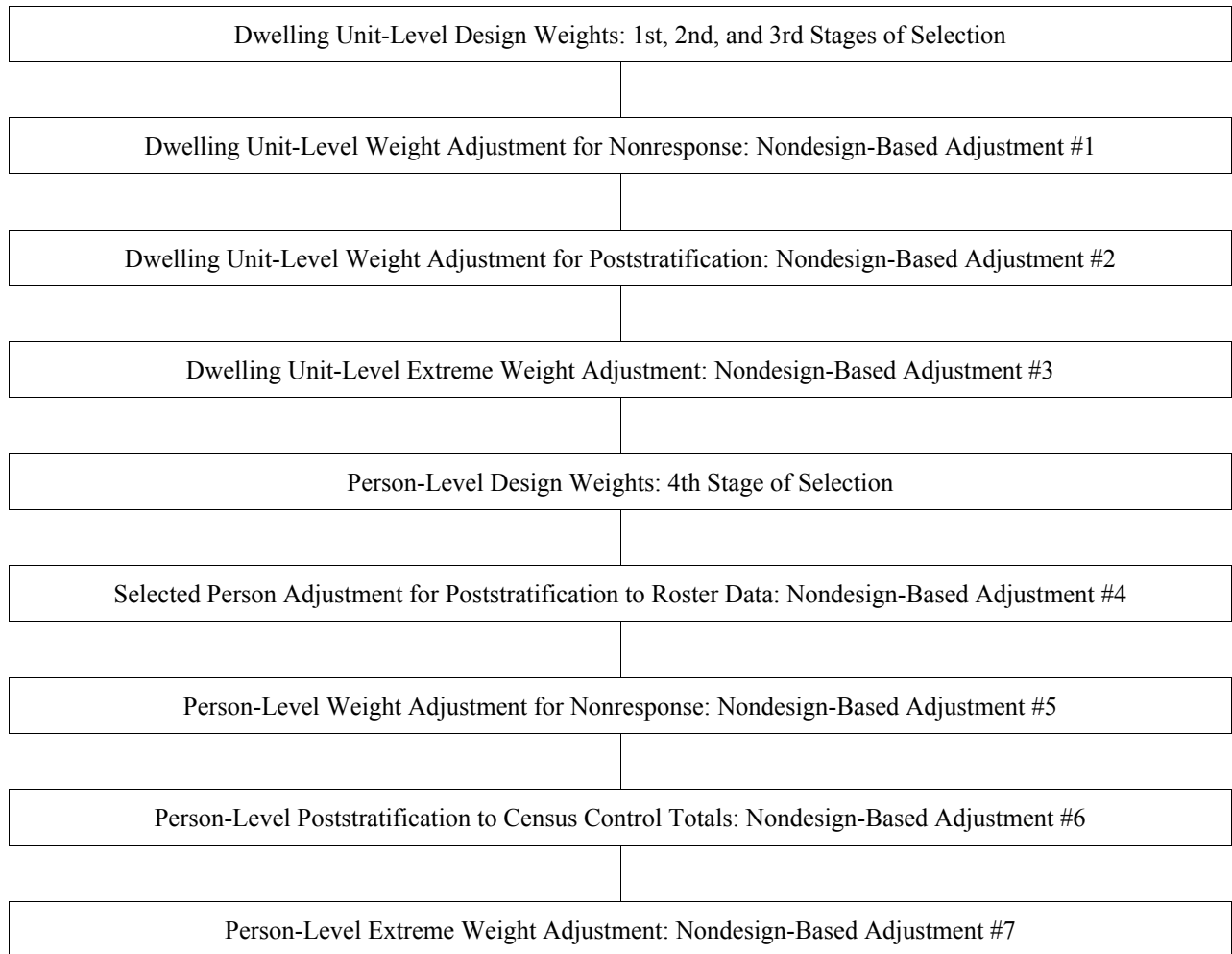


Exhibit 2 Sample Weight Components

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

3.9.2 Quality Control Measures in Design-Based Weighting Procedures

Quality control (QC) measures are applied to every component of the DU-level and person-level design weights. In addition to the QC measures outlined below, SAS programs are examined for errors, warnings, and uninitialization in the log by a sampling team member and reviewed by a different sampling team member. The following QC measures are employed to ensure the accuracy of design-based weight calculations:

- For segments that are subsegmented, check that the subsegmenting adjustment factor is greater than 1 (i.e., the count for the entire segment is greater than the count for the subsegment). This check is also performed for segments that are subsegmented twice.
- Compare the DU eligibility indicator with the completed screener indicator. Make sure all screener-complete DUs are eligible.
- Compare the final screening code with the DU eligibility and completed screener indicators to ensure these variables are defined correctly.
- Check the subsampling rate for added DUs that are subsampled. Review the frequency distribution of the DU subsampling rates to check values and ensure that the correct number of DUs are adjusted.
- Check that the minimum and maximum values of the DU release weight factor are within the expected range and that there are no missing values.
- Check the household-level weight to ensure that there are no missing values and the sum is close to the expected value.
- Compare the person-level indicators for eligible, selected, and complete. Make sure that all completed cases are selected and that all selected cases are eligible.
- Compare the final interview code with the person-level eligibility indicator to make sure this variable is defined correctly.
- Make sure that the probability of selection is nonmissing for all selected persons.
- Check the maximum-of-two selected persons' adjustment to make sure that the maximum value is two.
- Check the person-level weight to ensure that there are no missing values and the sum is close to the expected value.

4. General Sample Allocation Procedures for the Mental Health Surveillance Study and Expanded Mental Health Surveillance Study

As was done in the 2008 through 2011 National Surveys on Drug Use and Health (NSDUHs), a Mental Health Surveillance Study (MHSS) was embedded in the 2012 NSDUH. In 2008, a split-sample MHSS was conducted to calibrate the Kessler-6 (K6) nonspecific psychological distress scale and two competing functional impairment scales with a "gold-standard" clinical psychiatric interview in order to generate prevalence estimates of serious mental illness (SMI) among adults aged 18 or older. Based on the results from the 2008 MHSS, an abbreviated version of the World Health Organization-Disability Assessment Scale (WHODAS) (Novak, Colpe, Barker, & Gfroerer, 2010; Rehm et al., 1999) was adopted for the 2009 through 2012 surveys. The overarching goal of the MHSS was to fit a prediction model for SMI among adults aged 18 or older that can be used to create accurate model-based estimates of SMI prevalence at the national and domain levels (e.g., by age group and race/ethnicity). In addition, the MHSS data have the potential for a variety of important analyses beyond this primary purpose, including the model-free estimation of SMI prevalence and model-based and model-free estimation of moderate mental illness (MMI), mild mental illness (LMI), any mental illness (AMI), and specific mental disorders. The procedures for conducting the study remained the same. That is, a sample of respondents was contacted by phone for clinical follow-up.

4.1 Respondent Universe and Sampling Methods for the Mental Health Surveillance Study and Expanded Mental Health Surveillance Study

The 2012 MHSS was designed to yield 500 clinical follow-up interviews during 2012. In addition to the MHSS, a random sample of 1,000 clinical interviews was conducted to further refine and adjust the calibration of the NSDUH questions to measure SMI. This study is called the Expanded Mental Health Surveillance Study (EMHSS). The total sample size for the 2012 MHSS and EMHSS is 1,500 completed interviews. The sample was embedded in the main study sample; therefore, the initial interview for the validation cases was included in the target of 45,000 main study adult interviews. The target population for the MHSS excluded persons whose main study interview was conducted in Spanish.

A new selection algorithm was developed for the 2010 MHSS to mitigate the problem of extreme weights seen in the 2008 MHSS and was continued in 2011 and 2012. A subsample of eligible respondents aged 18 or older was selected for clinical follow-up with probabilities based on their K6 nonspecific psychological distress scale score (Kessler et al., 2003) and WHODAS score, and an age group adjustment factor was applied. A probability sampling algorithm was programmed in the computer-assisted interviewing (CAI) instrument so that field interviewers (FIs) could recruit selected respondents for the subsequent clinical psychiatric interview conducted by telephone.

4.2 Person Sample Allocation Procedures for the Mental Health Surveillance Study and Expanded Mental Health Surveillance Study

Table 4.1 shows some of the age-related factors used to compute sampling rates. Based on 2010 population estimates and the 2012 planned sample, the average weighting for persons aged 50 or older is 8.5 times as large as the average weighting for persons aged 18 to 25. (Smaller differences occur for intermediate age groups.) To compensate for this initial disparity in weights and to focus on persons aged 18 or older as a whole, sampling rates were set for persons aged 18 to 25, then adjusted for the other three age groups by applying the equalization factor, F , shown in Table 4.1.²⁵ The eligibility of questionnaire respondents for the clinical follow-up was based on the language used to complete the questionnaire. Persons completing the Spanish-language questionnaire were not eligible to be selected for the clinical follow-up. Response rates shown are the product of the percentage agreeing to the follow-up survey and the proportion of those who actually participate.²⁶

Table 4.1 MHSS/EMHSS Age-Related Factors

Age	2010 Population	Planned Sample	Average Weight	Weight Equalization Factor	Eligibility Factor	Response Rate Factor	Overall Age-Related Factor ¹
18 to 25	34,072,349	22,500	1,514	1.0000	97.08%	67.72%	1.0000
26 to 34	36,523,574	6,000	6,087	4.0198	93.03%	71.89%	3.9517
35 to 49	62,042,733	9,000	6,894	4.0198	94.64%	67.97%	4.1088
50 or Older	96,633,922	7,500	12,885	4.0198	96.76%	60.00%	4.5525

¹ The overall age-related factor is the weight equalization factor divided by the eligibility and response rate factors and then normalized.

Predictive models were developed to estimate the probability of SMI based on the K6 and WHODAS scores. In addition, the sample distribution by K6 and WHODAS scores was computed from the 2010 NSDUH data by the four sample allocation age groups represented in the 18 or older population.

The general sample allocation strategy was to find an allocation that provided the most precise estimate of SMI so that appropriate cut points could be established that produced the correct overall prevalence measure. A total of 225 strata were defined based on the combination of 25 possible K6 scores (0 to 24)²⁷ and 9 possible WHODAS scores (0 to 8). Predicted probabilities of SMI were used to obtain proportionality factors, $r_{h,age}$, for setting sampling rates by stratum (denoted h) and age group:

²⁵ Use of the derived weight equalization factors in Table 4.1 would have greatly increased the sampling rate for persons aged 50 or older. An adjusted set of factors that partially reduced the unequal weighting effects across age groups was specified instead. The adjusted equalization factors for the 35 to 49 and 50 or older age groups were set equal to the factor for the 26 to 34 age group.

²⁶ Eligibility and response rate factors were computed using 2010 MHSS data.

²⁷ In the prediction model, a recoded form of K6 score was used: scores 0 to 7 were recoded as 0, and all other scores had 7 subtracted from them to give a recoded total ranging from 0 to 17. These scores were reverse recoded to get back to the original K6 scores that were used in the two-way matrix. Thus, the predicted probabilities of mental illness are all identical for K6 scores of 0 to 7.

$$r_{h,age} \propto \frac{\sqrt{P_h(1-P_h)}}{E_{age} RR_{age}} * F_{age},$$

where P_h refers to the predicted probability of SMI in stratum h , and F_{age} , E_{age} , and RR_{age} refer to the age-specific weight equalization factors, eligibility factors, and response rate factors, respectively. These proportionality factors then were multiplied by the projected sample counts and scaled to achieve an overall respondent sample of 1,500 persons aged 18 or older to obtain the stratum and age-specific sampling rates. As an example, the predicted probability of SMI for a person with a K6 score of 10 and a WHODAS score of 6 was 0.1398. For the 18 to 25 age group, the proportionality factor then would be

$$r_{h,18-25} = \frac{\sqrt{0.1398(1-0.1398)}}{0.9708 * 0.6772} * 1.000 = 0.5275$$

An adjustment factor of 0.0916 was applied to each proportionality factor in order to achieve an overall sample of 1,500 persons. Thus, the sampling rate for this stratum and age group was $0.5275 * 0.0916 = 0.0483$.

Projected yields of positive cases based on the predicted probability of SMI broken out by age group are provided in [Table 4.2](#). In addition, [Tables 4.3](#) and [4.4](#) provide the 2012 MHSS/EMHSS sample allocation by K6 group and WHODAS score, respectively.

Table 4.2 Projected Yields of Predicted Positive Cases

	Age Group				
	18 to 25	26 to 34	35 to 49	50 or Older	18 or Older
SMI	76	69	92	45	281
AMI	195	174	224	126	720
Total Sample	338	336	476	351	1,500

AMI = any mental illness, SMI = serious mental illness.

The probability sample of 1,500 clinical follow-up interviews was distributed across four calendar quarters with approximately 375 follow-up interviews per quarter. Throughout the 2012 survey, the MHSS/EMHSS sample was monitored and the sampling parameters were modified on an as-needed basis.

The 2012 MHSS/EMHSS resulted in 1,622 completed clinical interviews. An estimated 84 percent of selected persons agreed to participate in the MHSS/EMHSS, and 79 percent of those persons completed the clinical interview.

The analysis weight for the MHSS/EMHSS sample will be separate from the main study analysis weight. To compute the MHSS/EMHSS analysis weight, the main study analysis weight will be multiplied by the inverse of the probability of selecting the person for clinical follow-up (the inverse of the sampling parameter used to select the person) and nonresponse and poststratification adjustments ([Exhibit 3](#)).

Table 4.3 2012 MHSS/EMHSS Sample Allocation, by K6 Group

K6 Group	Percent of Population¹	Assumed SMI Rate (Percent)^{2,3}	Expected Sample Size	Expected SMI Count	Overall Sampling Rate
0 to 3	53.54	0.92	497	0	0.02370
4 to 5	13.67	1.21	145	0	0.02386
6 to 7	9.08	1.58	114	0	0.02534
8 to 9	6.26	2.86	98	0	0.03126
10 to 11	4.41	5.51	99	4	0.04192
12 to 15	6.71	12.28	225	52	0.05731
16 or Higher	6.33	41.85	322	225	0.07980
Total	100.00	4.70	1,500	281	

K6 = Kessler-6, a 6-item psychological distress scale; SMI = serious mental illness.

¹ Source: 2010 National Survey on Drug Use and Health.

² Source: 2009 Mental Health Surveillance Study.

³ To compute assumed SMI rates, SMI estimates by K6 and WHODAS score were averaged (weighted) across K6 scores. These rates are not the actual SMI rates that were used in the sample allocation.

Table 4.4 2012 MHSS/EMHSS Sample Allocation, by WHODAS Score

WHODAS Score	Percent of Population¹	Assumed SMI Rate (Percent)^{2,3}	Expected Sample Size	Expected SMI Count	Overall Sampling Rate
0	74.26	1.06	738	0	0.02210
1	7.41	2.61	116	3	0.03473
2	4.95	4.67	104	4	0.04695
3	3.37	7.68	93	8	0.06132
4	2.61	13.45	89	22	0.07585
5	2.05	21.26	89	36	0.09634
6	1.83	32.44	87	57	0.10606
7	1.44	42.75	75	56	0.11673
8	2.10	58.81	108	95	0.11413
Total	100.00	4.70	1,500	281	

K6 = Kessler-6, a 6-item psychological distress scale; SMI = serious mental illness; WHODAS = World Health Organization-Disability Assessment Scale.

¹ Source: 2010 National Survey on Drug Use and Health.

² Source: 2009 Mental Health Surveillance Study.

³ To compute assumed SMI rates, SMI estimates by K6 and WHODAS score were averaged (weighted) across K6 scores. These rates are not the actual SMI rates that were used in the sample allocation.

In addition to the analysis weight, MHSS/EMHSS-specific variance estimation strata and replicates will be created to appropriately account for the design when computing variances of estimates. To create the MHSS/EMHSS variance strata, groups of 18 adjacent main study variance strata will be collapsed. Thus, a total of 50 MHSS/EMHSS variance strata will be formed. The variance replicates will be retained from the main study.

Exhibit 3 MHSS/EMHSS Sample Weight Components

Dwelling Unit-Level Weight Components	
#1.	Inverse Probability of Selecting Census Tract
#2.	Inverse Probability of Selecting Segment
#3.	Quarter Segment Weight Adjustment
#4.	Subsegmentation Inflation Adjustment
#5.	Inverse Probability of Selecting Dwelling Unit
#6.	Inverse Probability of Added/Subsampled Dwelling Unit
#7.	Dwelling Unit Release Adjustment
#8.	Dwelling Unit Nonresponse Adjustment
#9.	Dwelling Unit Poststratification Adjustment
#10.	Dwelling Unit Extreme Weight Adjustment

Person-Level Weight Components	
#11.	Inverse Probability of Selecting a Person within a Dwelling Unit
#12.	Selected Person Poststratification to Roster Adjustment
#13.	Person-Level Nonresponse Adjustment
#14.	Person-Level Poststratification Adjustment
#15.	Person-Level Extreme Weight Adjustment

MHSS/EMHSS Weight Components	
#16.	Inverse Probability of Selecting a Person for the MHSS/EMHSS
#17.	Person-Level Nonresponse Adjustment
#18.	Person-Level Poststratification Adjustment

5. General Sample Procedures for the Questionnaire Field Test

A Questionnaire Field Test (QFT) was conducted during the 2012 National Survey on Drug Use and Health (NSDUH). The field test overlapped with quarters 3 and 4 of the main study with data collection beginning on September 1, 2012, and ending on November 3, 2012. Approximately 2,000 persons were interviewed in order to test a redesigned questionnaire protocol for implementation in the 2015 NSDUH. The QFT was designed to address the impact that the redesign protocol had on the reporting of core substance use or the reporting on noncore survey items, on screening and interview timing, and on data quality. In addition, the QFT attempted to measure the feasibility of the redesign protocol based on field interviewer (FI) debriefings or respondent feedback to interviews. Data from the QFT were compared with data from the 2011 NSDUH and with data from quarters 3 and 4 of the 2012 main study. This chapter provides the sampling procedures for the QFT.

5.1 Target Population

Similar to NSDUH, the respondent universe for the QFT was the civilian, noninstitutionalized population aged 12 or older. In order to control costs, persons residing in Alaska and Hawaii as well as persons who do not speak English were excluded from the QFT. Therefore, the sample was representative of the English-speaking, noninstitutional population aged 12 or older in the contiguous United States.

5.2 Selection of State Sampling Regions and Segments

An estimated 213 segments were needed to yield approximately 2,000 completed interviews. To achieve national representation of the English-speaking population, a probability proportional to size sample of 213 (of 876) State sampling (SS) regions was selected. Alaska and Hawaii were excluded from the QFT sample to control costs. This design maximized the efficiency (i.e., increased precision) of the QFT estimates by reducing variation in the weights. In addition, this design had the benefit of placing the sample in heavily populated areas where a sufficient mix of FIs with various experience levels to meet staffing goals were expected. As shown in [Table 5.1](#), a large portion of the sample was selected from the eight largest States.

Within each selected SS region, a sample of dwelling units (DUs) was drawn from the segment that was retired from use in quarter 1 of the 2012 NSDUH. If an insufficient number of eligible DUs remained in a segment, or if significant access problems were expected in a segment, the segment was replaced with the 2011 quarter 4 retired segment in the same SS region. A total of 6 segments were replaced because they had fewer than 10 eligible DUs remaining, and a total of 7 segments were replaced because of anticipated access problems in the segments.

Table 5.1 Number of 2012 Questionnaire Field Test State Sampling Regions and Achieved Sample Sizes, by State

State	Population Rank (12 or Older)	Current Design	NSDUH SS Regions	Number of QFT SS Regions/ Segments (PPS)	2012 QFT Respondents
CA	1	3,600	48	23	170
TX	2	3,600	48	14	146
NY	3	3,600	48	11	105
FL	4	3,600	48	18	169
IL	5	3,600	48	10	72
PA	6	3,600	48	10	121
OH	7	3,600	48	7	103
MI	8	3,600	48	9	86
GA	9	900	12	6	60
NC	10	900	12	5	50
NJ	11	900	12	6	52
VA	12	900	12	6	83
MA	13	900	12	4	33
WA	14	900	12	5	46
IN	15	900	12	6	63
AZ	16	900	12	4	14
TN	17	900	12	4	51
MO	18	900	12	2	16
WI	19	900	12	4	38
MD	20	900	12	3	32
MN	21	900	12	4	36
CO	22	900	12	6	33
AL	23	900	12	4	45
SC	24	900	12	3	31
KY	25	900	12	3	28
LA	26	900	12	5	66
OR	27	900	12	1	8
OK	28	900	12	5	40
CT	29	900	12	5	41
IA	30	900	12	0	0
MS	31	900	12	0	0
AR	32	900	12	0	0
KS	33	900	12	2	19
NV	34	900	12	2	33
UT	35	900	12	6	63

(continued)

Table 5.1 Number of 2012 Questionnaire Field Test State Sampling Regions and Achieved Sample Sizes, by State (continued)

State	Population Rank (12 or Older)	Current Design	NSDUH SS Regions	Number of QFT SS Regions/ Segments (PPS)	2012 QFT Respondents
NM	36	900	12	1	4
WV	37	900	12	2	23
NE	38	900	12	3	25
ID	39	900	12	0	0
ME	40	900	12	2	12
NH	41	900	12	1	11
HI	42	900	12	0	0
RI	43	900	12	0	0
MT	44	900	12	1	16
DE	45	900	12	0	0
SD	46	900	12	0	0
AK	47	900	12	0	0
VT	48	900	12	0	0
ND	49	900	12	0	0
DC	50	900	12	0	0
WY	51	900	12	0	0
Total		67,500	900	213	2,044

SS = State sampling; PPS = probability proportional to size; QFT = Questionnaire Field Test.

5.3 Selection of Dwelling Units

The procedures described in Sections 3.3 and 3.4 were used to determine DU sample sizes and person probabilities of selection for the QFT. A sufficient number of DUs was drawn to account for the lower sample yield resulting from conducting interviews in English only. Data from the 2011 NSDUH were used to predict eligibility and nonresponse rates by State, and these estimates were used in the QFT sample allocation. The number of sample DUs was limited to 50 per segment in order to ensure a reasonable spread across QFT segments. DUs that were not selected for the 2010, 2011, and 2012 main studies were eligible for selection in the QFT.

Similar to the main study, a small reserve sample (20 percent) of DUs from each segment was selected, and the total sample was partitioned into four equal probability subsamples: 105/120, 5/120, 5/120, and 5/120. Although the majority of the sample (105/120) was released at the beginning of the QFT data collection period, having the additional sample partitions allowed for greater flexibility in controlling the sample size and provided the ability to ensure data collection goals were attained within the field period. Two additional 5 percent partitions (5/120) were released in all but six States²⁸ after 4 weeks of data collection.

²⁸ Additional sample was not released in the following States: Connecticut, New Mexico, Oregon, South Carolina, Utah, and Virginia.

A total of 5,358 DUs were sampled and yielded 2,044 completed interviews, as shown in [Table 5.2](#). The half-open interval procedure for missed DUs (see Section 3.7) was implemented during the QFT, but it is not scheduled to be implemented in the 2014 or 2015 NSDUHs.

Table 5.2 Summary of the 2012 Questionnaire Field Test Sample Results

Statistic	Total	Rate
State Sampling (SS) Regions	213	
Segments	213	
Selected Dwelling Units	5,358	
Eligible Dwelling Units	4,623	0.86
Completed Screening Interviews	3,837	0.83
Selected Persons	2,823	
Eligible Persons (English Speaking)	2,760	0.98
Completed Interviews	2,044	0.74

5.4 Age Group Allocations

The respondent sample was allocated to the three major age groups in the following proportions: 25 percent aged 12 to 17, 25 percent aged 18 to 25, and 50 percent aged 26 or older. Within the 26 or older age group, 15 percent of the sample was allocated to persons aged 26 to 34, 20 percent of the sample was allocated to persons aged 35 to 49, and 15 percent was allocated to persons aged 50 or older. This sample allocation matched the planned allocation for the 2015 NSDUH partial redesign. One implication of the respondent sample allocation by age groups is a potential impact on QFT response rates. Retaining more of the 26 or older adults identified in households to complete interviews may have had a negative effect on unweighted interview response rates.²⁹ Specifically, respondents aged 50 or older have shown lower propensities to participate than other age groups in the NSDUH sample. The unweighted interview response rate for the QFT sample was 72.41 percent compared with 76.52 percent for the 2011 main study comparison sample and 79.31 percent for the 2012 quarters 3 and 4 main study comparison sample.³⁰ On the other hand, although a smaller proportion of 12 to 17 year olds were selected, this age group continued to drive the number of DUs needed (i.e., the age group continued to be sampled at the highest rate). Thus, fewer DUs were needed to yield the desired sample than would be needed under the current sample design.

5.5 Selection of Persons

After DUs were selected within each QFT segment, an interviewer visited each selected DU to obtain a roster of all persons residing in the DU. This roster information was used to select 0, 1, or 2 persons for the survey. Sampling rates were preset by segment and age group. Roster information was entered directly into the electronic screening instrument that automatically implemented this stage of selection based on the segment and age group sampling parameters. As illustrated in [Table 5.2](#), a total of 2,823 people were selected from within 4,623 screened and eligible DUs, which yielded 2,044 completed interviews.

²⁹ Weighted interview response rates were not expected to change as a result of the age group allocation.

³⁰ Other factors influencing the interview response rates were the shorter data collection period and a limited number of FIs available for use in converting QFT refusals.

5.6 Creation of Variance Estimation Strata and Replicates

The nature of the stratified, clustered sampling design of the NSDUH main study and QFT samples requires that the design structure be taken into consideration when computing variances of survey estimates. Key nesting variables were created for the QFT and main study comparison samples to capture explicit stratification and to identify clustering.

To allow for comparisons between the QFT and main study samples, a common set of stratification and clustering variables were defined. Because SS regions serve as strata for the main study samples and as primary sampling units (PSUs) for the QFT sample, there was no direct way of capturing the covariance between the samples and using the entire main study sample. Instead, the approach used for the 1999 paper-and-pencil interviewing (PAPI) and computer-assisted interviewing (CAI) mode analysis was followed in developing a design structure that could be used to simultaneously analyze all three samples (Chromy, Davis, Packer, & Gfroerer, 2002, pp. 141-142). Steps in the process were as follows:

- Within QFT sampling strata (census regions), variance strata were generally formed by assigning two sequential QFT-selected SS regions to the same variance strata on the sorted sampling frame. Each sampled SS region was then assigned to a replicate (1 or 2). However, there were three QFT SS regions per variance strata for three randomly selected strata. This was necessary because there were an odd number of QFT SS regions selected in three of the census regions. Within these three strata, the third SS region was randomly assigned to either replicate 1 or replicate 2. This led to a total of 105 QFT variance strata, with two replicates per strata.
- Using the sorted QFT sampling frame of SS regions, the main study SS regions not selected for the QFT were assigned to QFT sampling strata sequentially, in accordance with the assignments of selected QFT SS regions. These assignments kept the number of SS regions per strata as equal as possible given the distribution of QFT sampled SS regions within the sorted SS region frame. For SS regions not selected for the QFT sample, the original replicate assignments of either replicate 1 or replicate 2 were maintained.

Although this approach to design structure variables does not fit the main study perfectly, it does capture the total variance and allows for taking advantage of any covariance induced by the overlapping SS regions between the samples.

References

- American Psychiatric Publishing, Inc. (2000). *Diagnostic and statistical manual of mental disorders, 4th ed., text revision* (DSM-IV-TR; doi: 10.1176/appi.books.9780890423349). Retrieved January 29, 2013, from <http://www.psychiatryonline.com/resourceTOC.aspx?resourceID=1>
- Chen, P., Dai, L., Cribb, D., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (in press). Person-level sampling weight calibration. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283201000003C, Phase I, Deliverable No. 39, RTI/ 0212800.001.107.004). Research Triangle Park, NC: RTI International.
- Chromy, J. R. (1979). Sequential sample selection methods. In *Proceedings of the 1979 American Statistical Association, Survey Research Methods Section, Washington, DC* (pp. 401-406). Washington, DC: American Statistical Association.
- Chromy, J. R. (1981). Variance estimators for a sequential sample selection procedure. In D. Krewski, R. Platek, & J. N. K. Rao (Eds.), *Current topics in survey sampling: Proceedings of the symposium held May 7-9, 1980, sponsored by the Ottawa Chapter and the Survey Research Methods Section of the American Statistical Association* (pp. 329-347). New York, NY: Academic Press.
- Chromy, J. R. (1987). Design optimization with multiple objectives. In *Proceedings of the 1987 American Statistical Association, Survey Research Methods Section, San Francisco, CA* (pp. 194-199). Alexandria, VA: American Statistical Association.
- Chromy, J., Davis, T., Packer, L., & Gfroerer, J. (2002). Mode effects on substance use measures: Comparison of 1999 CAI and PAPI data. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 135-159). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.
- Chromy, J. R., & Penne, M. A. (2002). Pair sampling in household surveys. In *Proceedings of the 2002 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, New York, NY* (pp. 552-554). Alexandria, VA: American Statistical Association.
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York, NY: John Wiley & Sons.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002, November). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition. (SCID-I/NP)*. New York, NY: New York State Psychiatric Institute, Biometrics Research.
- Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S. L., Manderscheid, R. W., Walters, E. E., & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*, *60*, 184-189.

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

Novak, S. P., Colpe, L. J., Barker, P. R., & Gfroerer, J. C. (2010). Development of a brief mental health impairment scale using a nationally representative sample in the USA. *International Journal of Methods in Psychiatric Research*, 19(S1), 49-60. doi: 10.1002/mpr.313.

Office of Management and Budget. (2003, June 6). *OMB Bulletin No. 03-04: Revised definitions of metropolitan statistical areas, new definitions of micropolitan statistical areas and combined statistical areas, and guidance on uses of the statistical definitions of these areas*. Retrieved January 29, 2013, from http://www.whitehouse.gov/omb/bulletins_b03-04

Rehm, J., Üstün, T. B., Saxena, S., Nelson, C. B., Chatterji, S., Ivis, F., & Adlaf, E. (1999). On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. *International Journal of Methods in Psychiatric Research*, 8, 110-123. doi:10.1002/mpr.61.

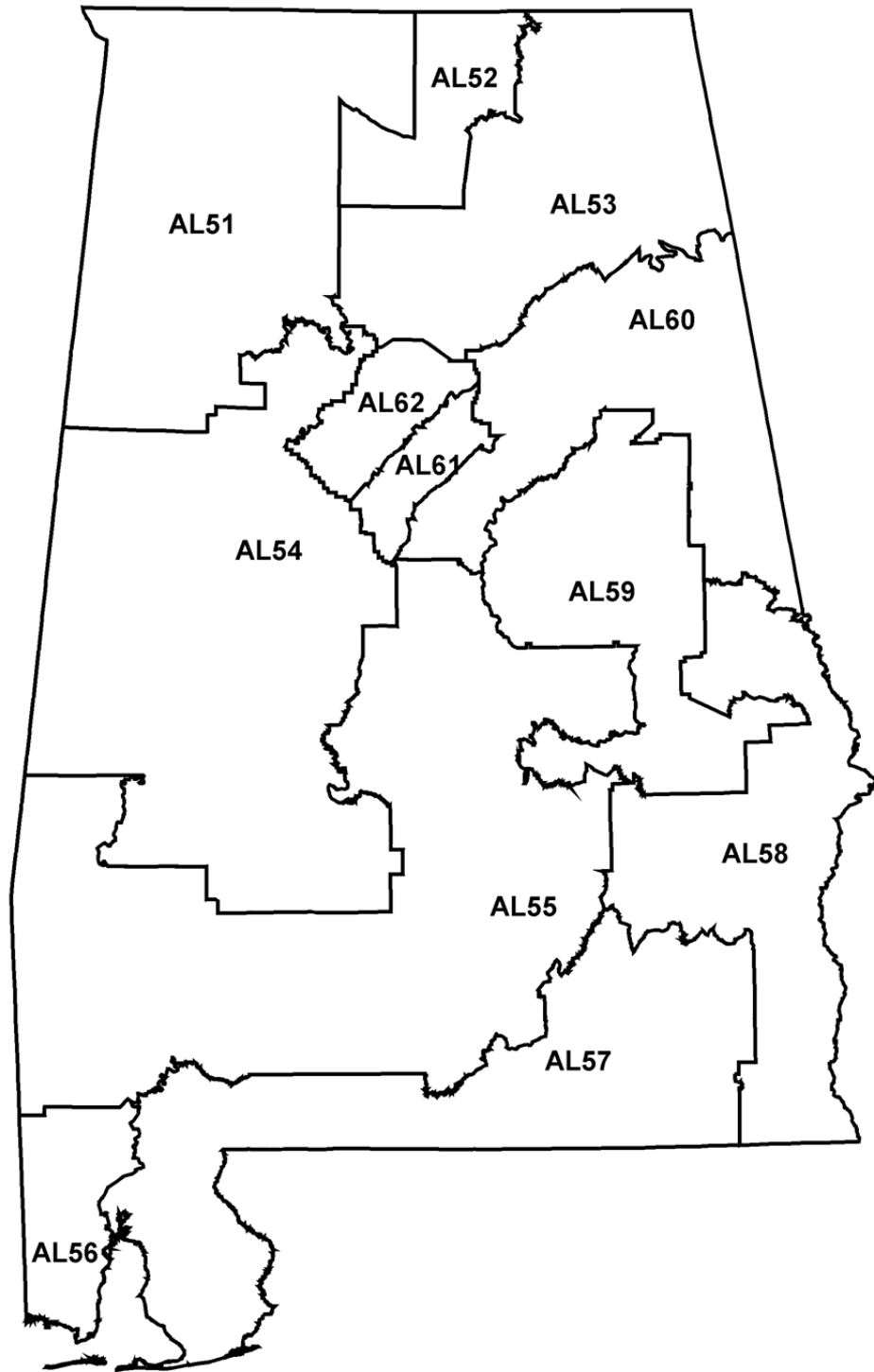
RTI International. (2008). *SUDAAN®*, Release 10.0 [computer software]. Research Triangle Park, NC: Author.

RTI International. (2011). *2012 National Survey on Drug Use and Health: Counting and listing general manual* (RTI/0212800, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283201000003C). Research Triangle Park, NC: Author.

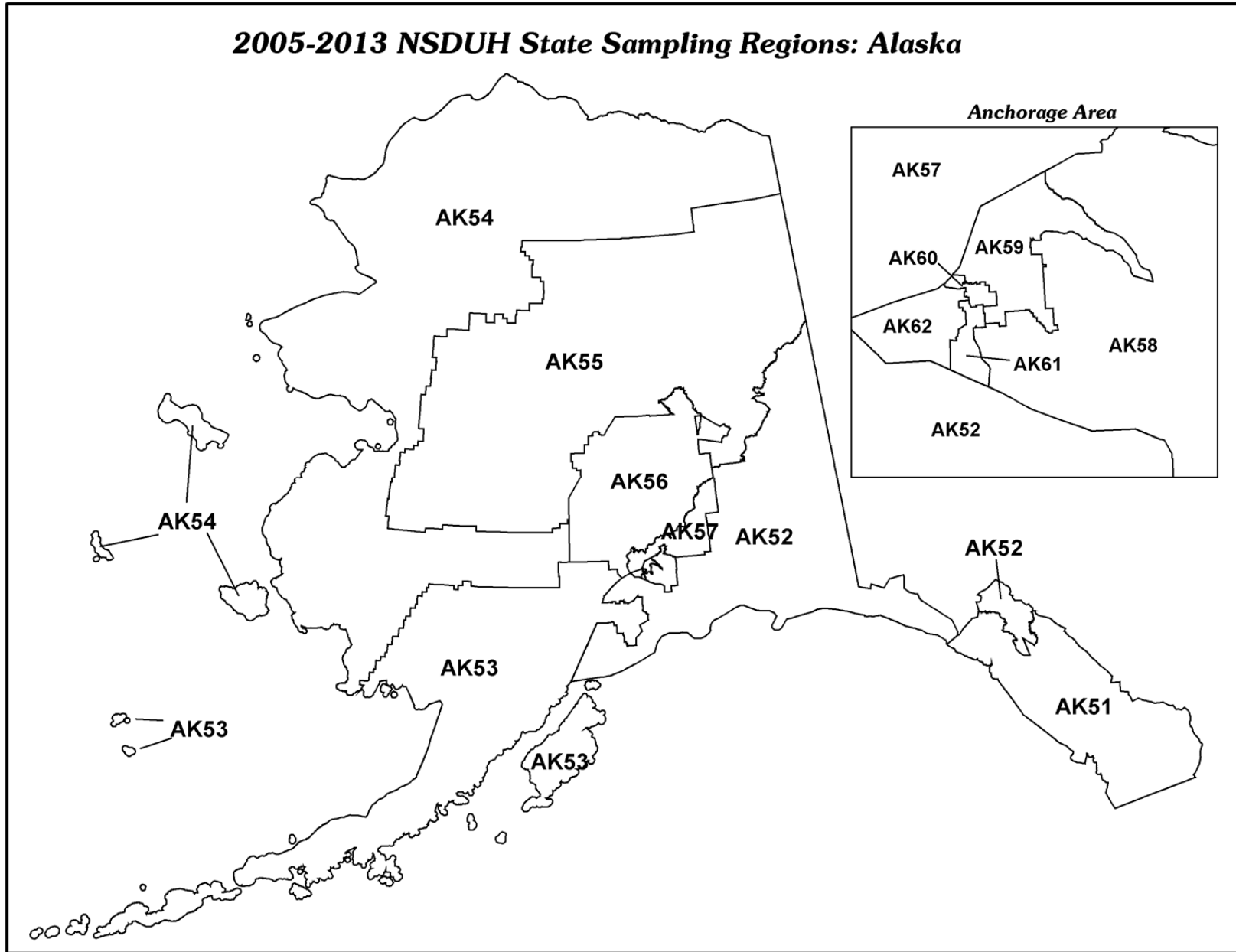
U.S. Census Bureau, Redistricting Data Office. (2009). Appendix A: Geographic terms and concepts. In *2008 Redistricting data prototype (Public Law 94-171) summary file*. Washington, DC: Author. [Available as a PDF at <http://www.census.gov/rdo/>; see <http://www.census.gov/rdo/pdf/pl94-171.pdf>]

**Appendix A:
2005 through 2013 NSDUH State Sampling Regions**

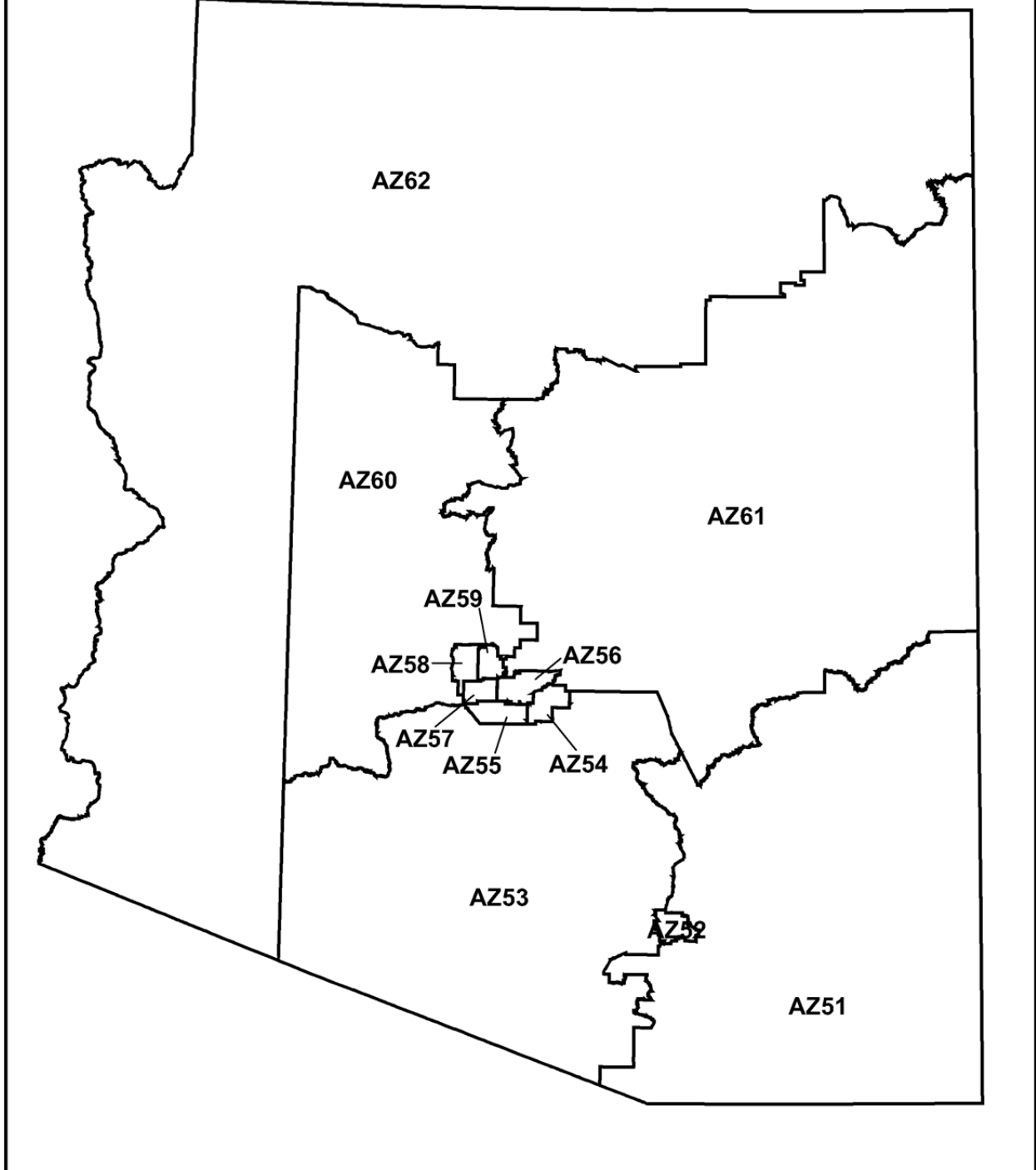
2005-2013 NSDUH State Sampling Regions: Alabama



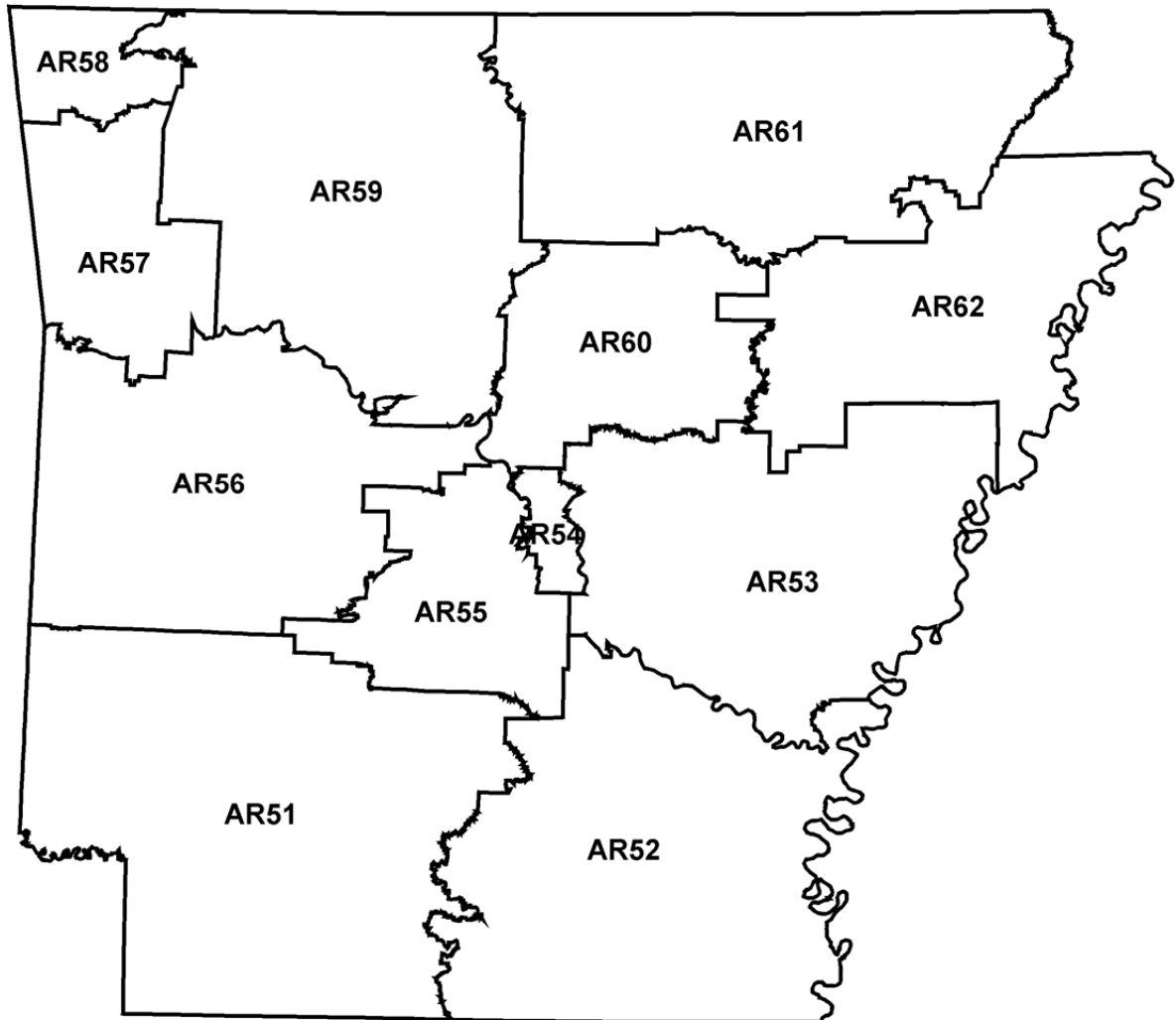
2005-2013 NSDUH State Sampling Regions: Alaska



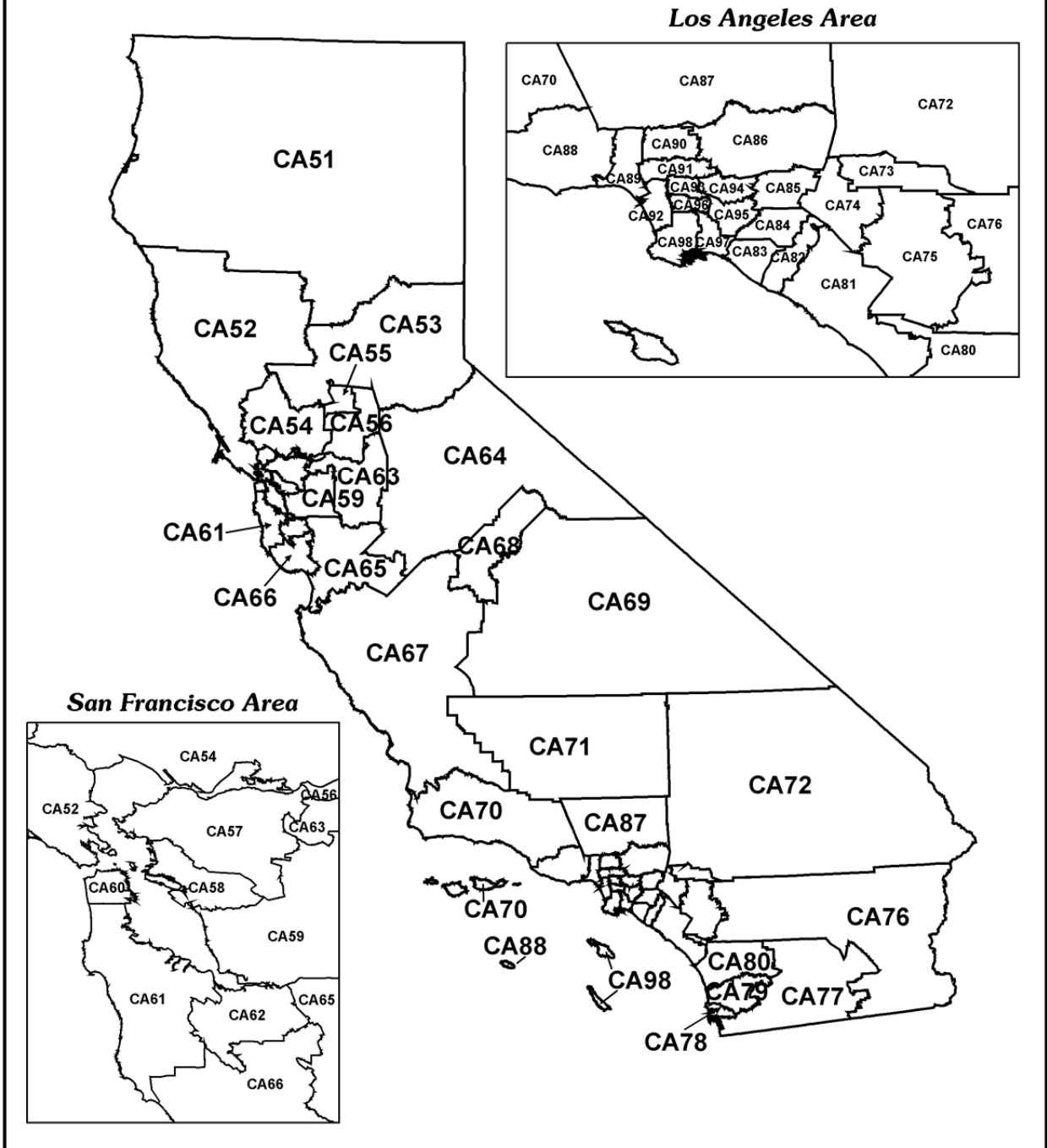
2005-2013 NSDUH State Sampling Regions: Arizona



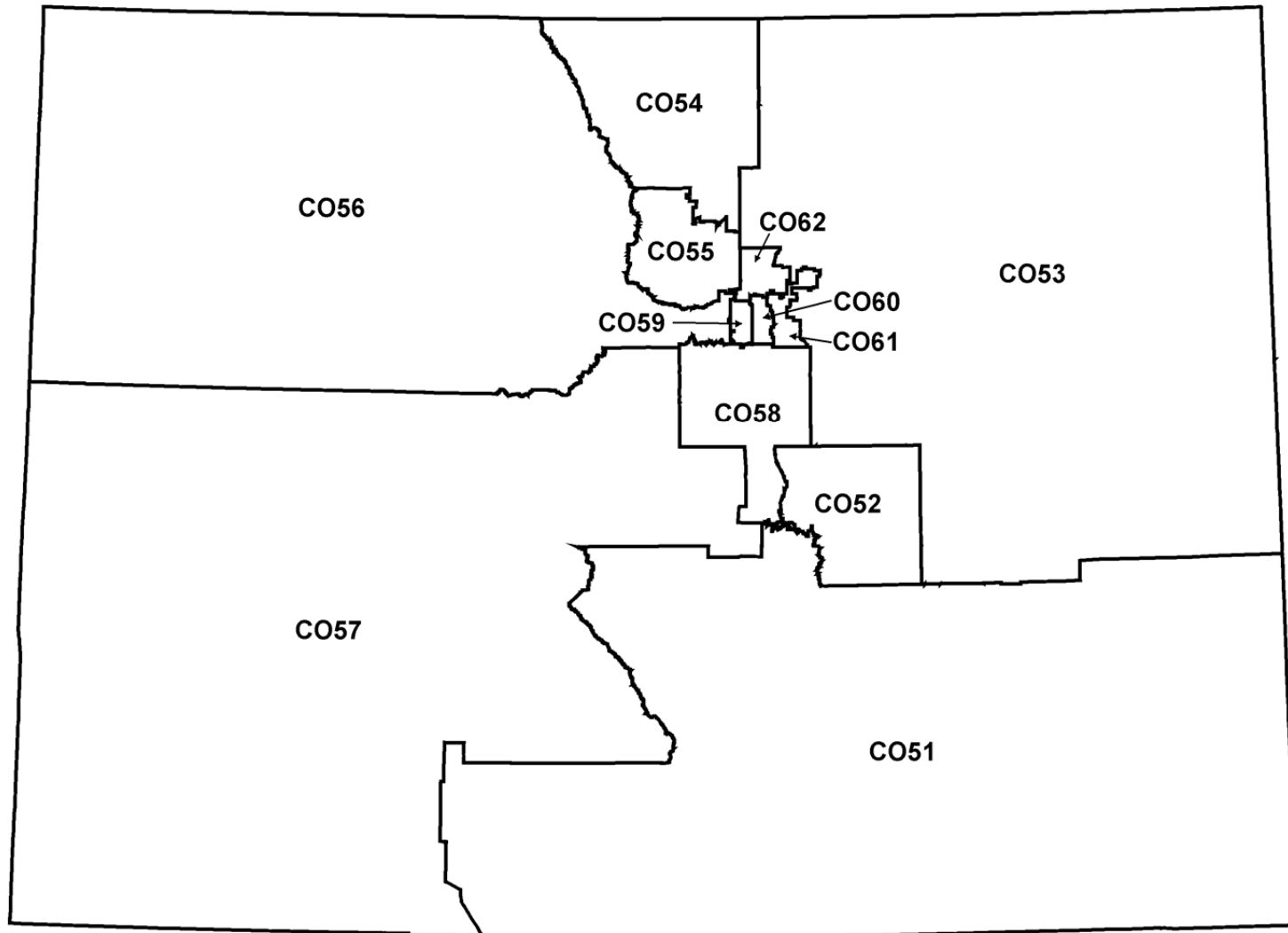
2005-2013 NSDUH State Sampling Regions: Arkansas



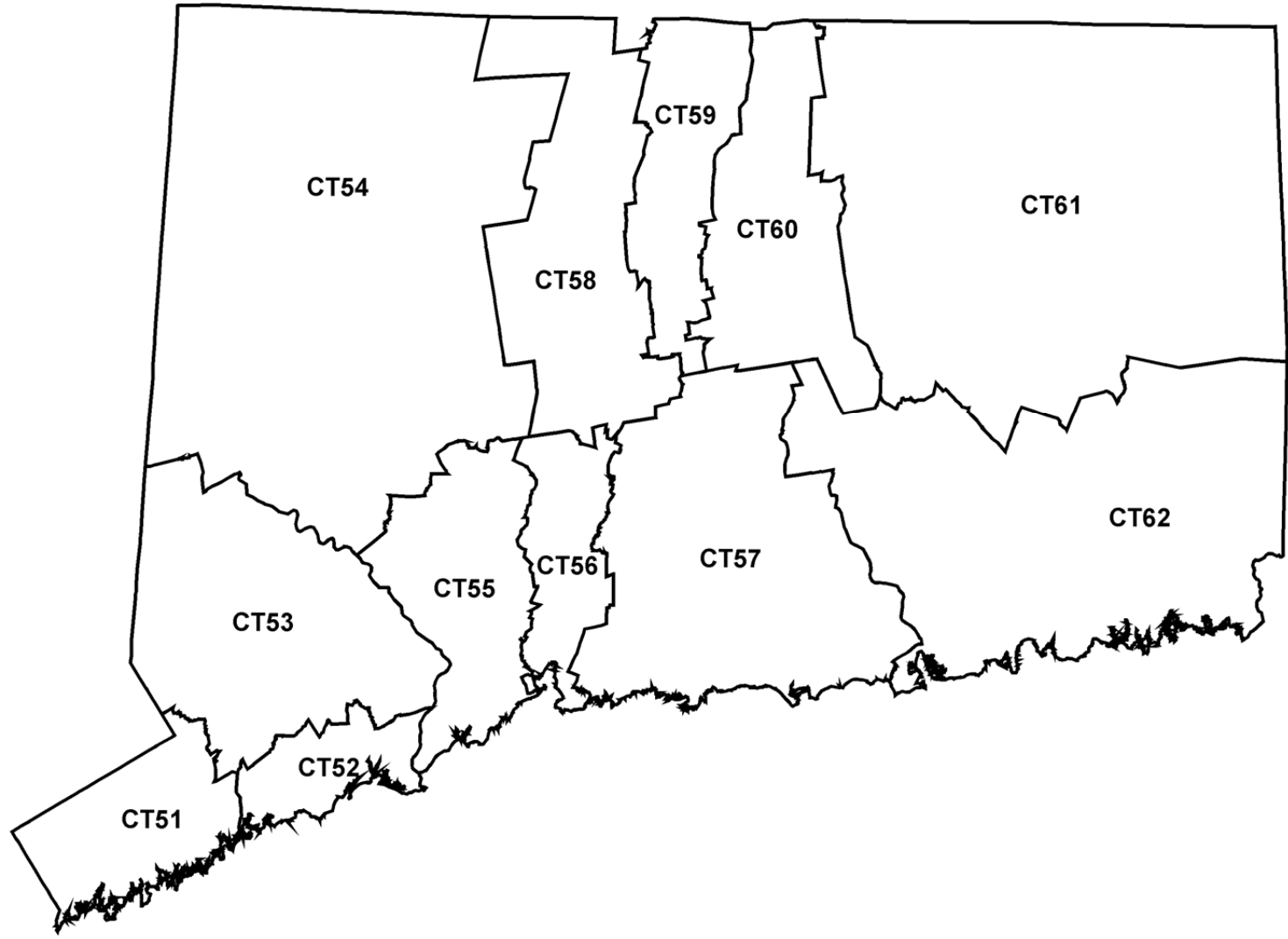
2005-2013 NSDUH State Sampling Regions: California



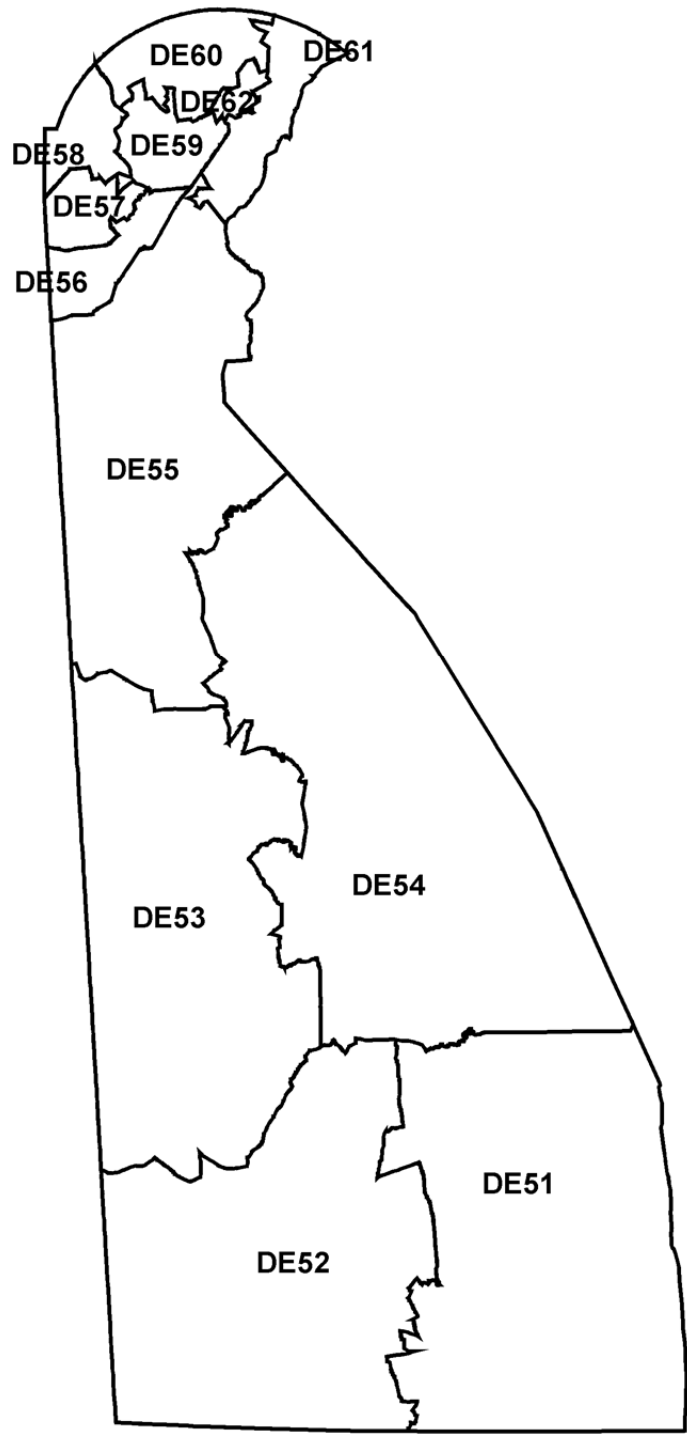
2005-2013 NSDUH State Sampling Regions: Colorado



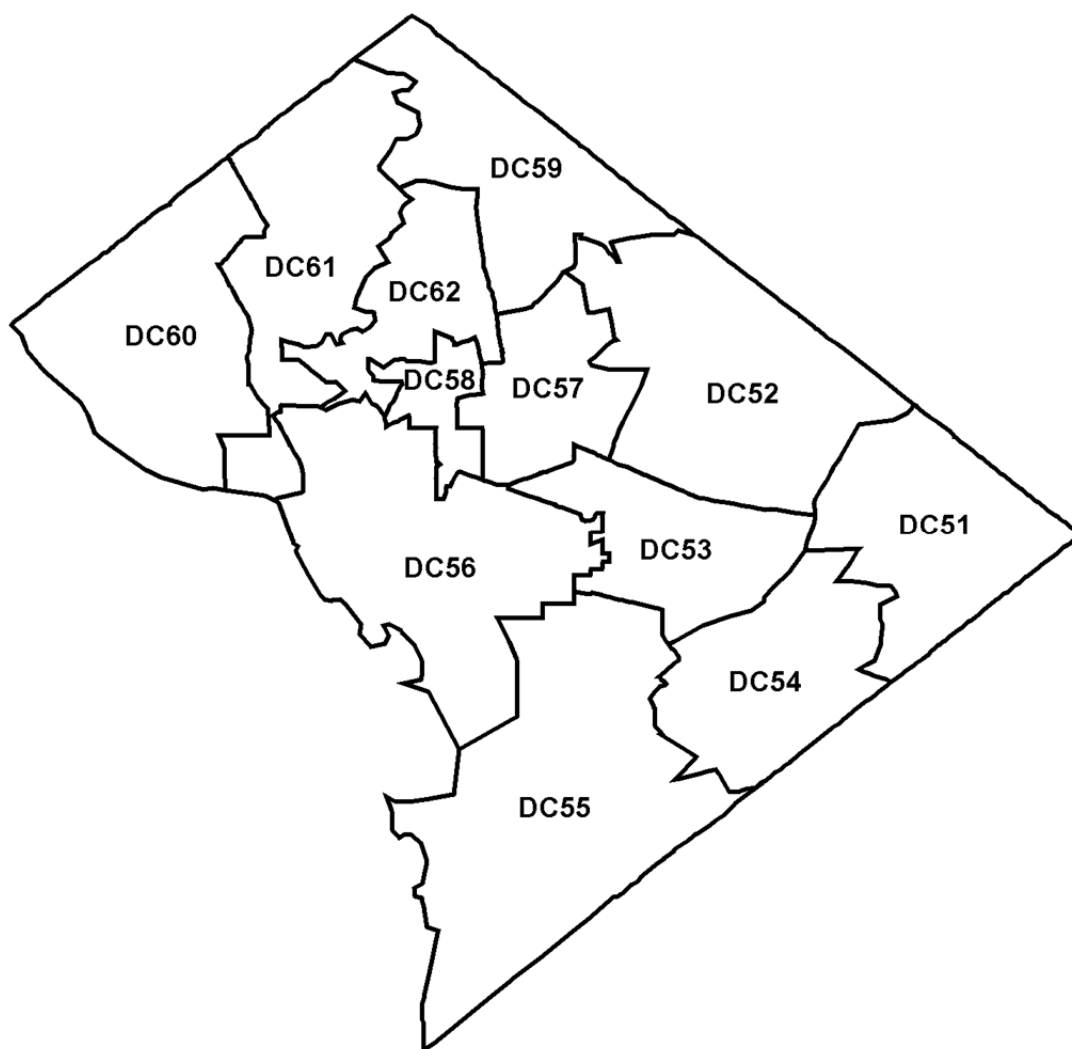
2005-2013 NSDUH State Sampling Regions: Connecticut



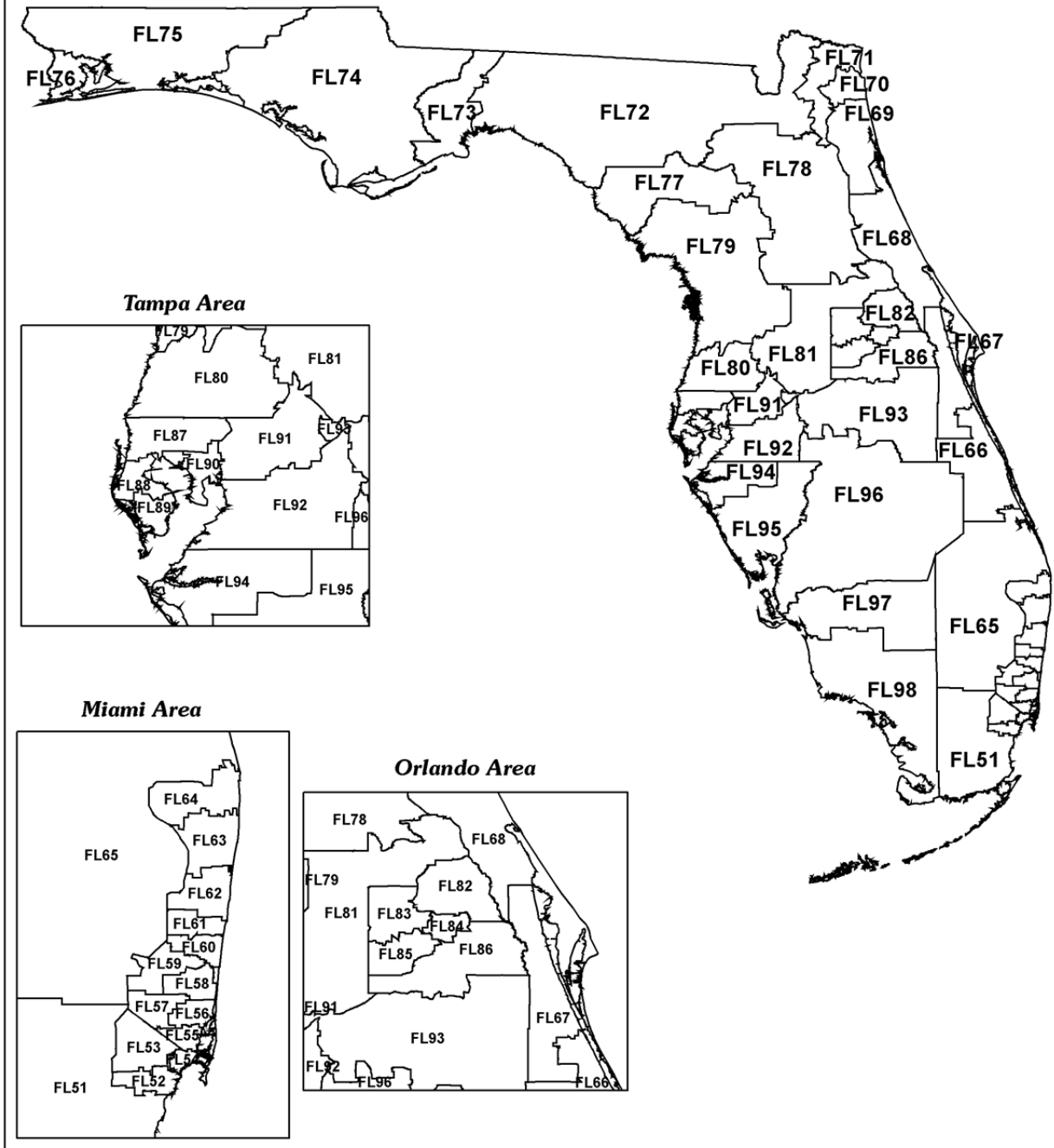
2005-2013 NSDUH State Sampling Regions: Delaware



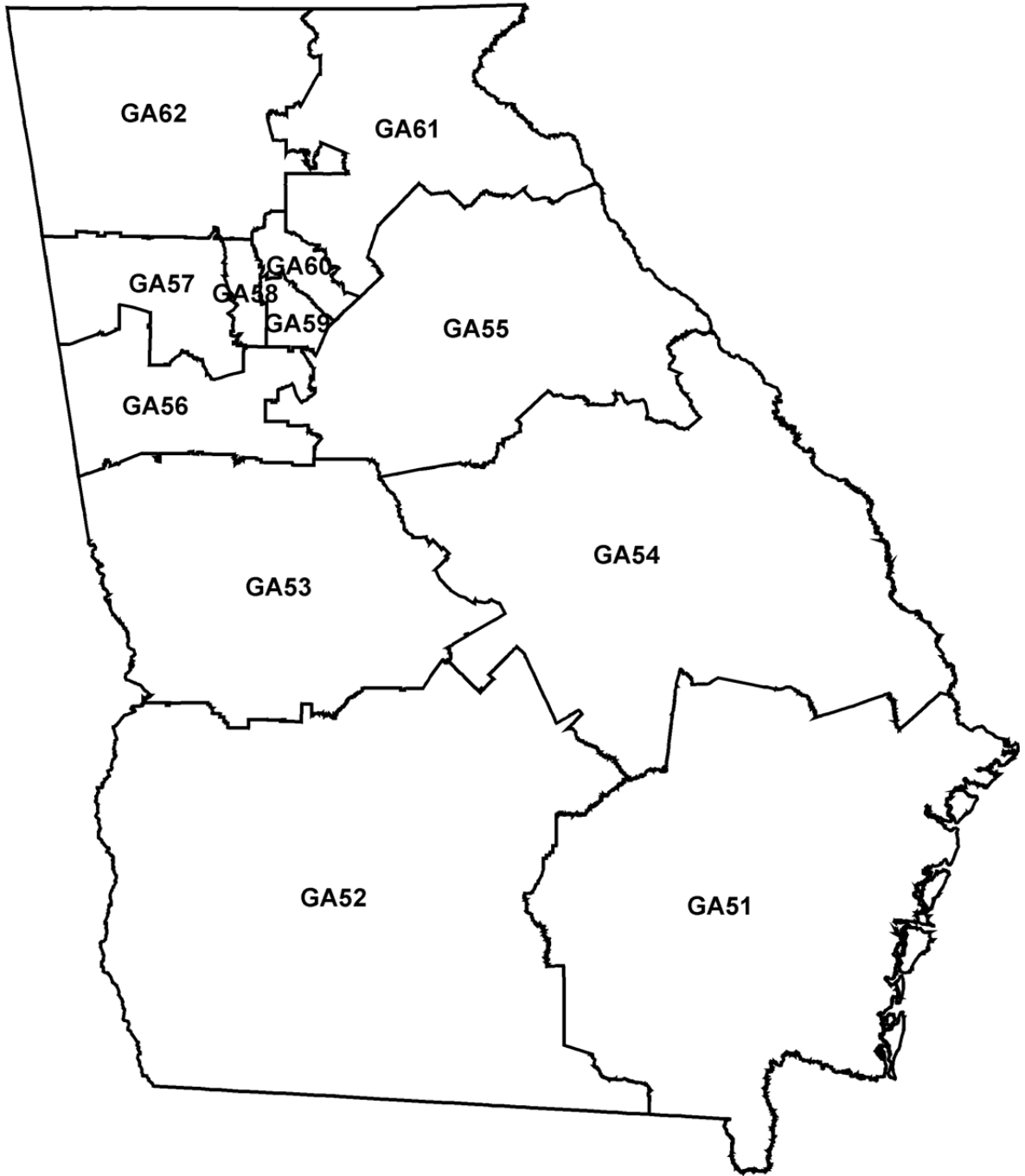
2005-2013 NSDUH State Sampling Regions: District of Columbia



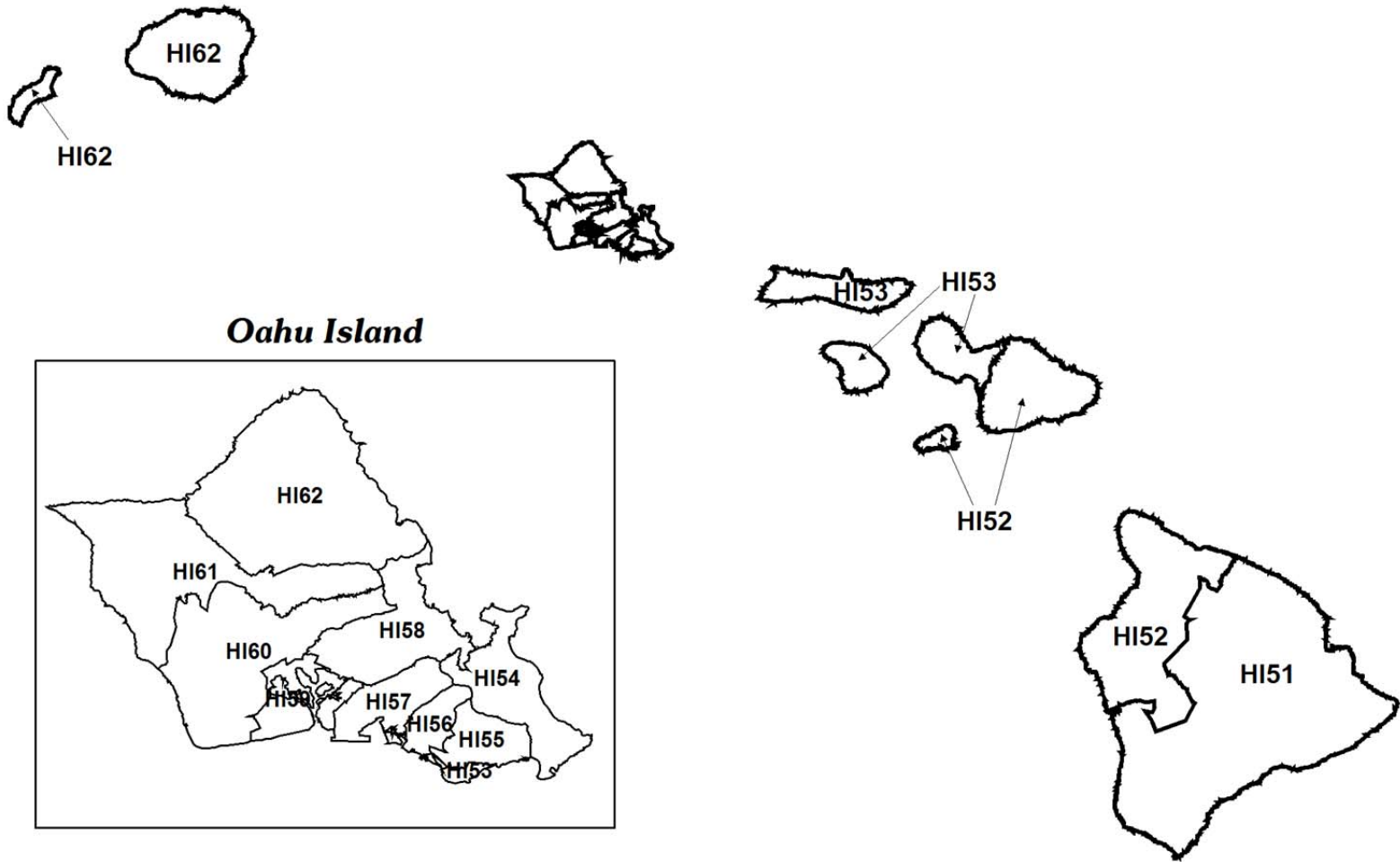
2005-2013 NSDUH State Sampling Regions: Florida



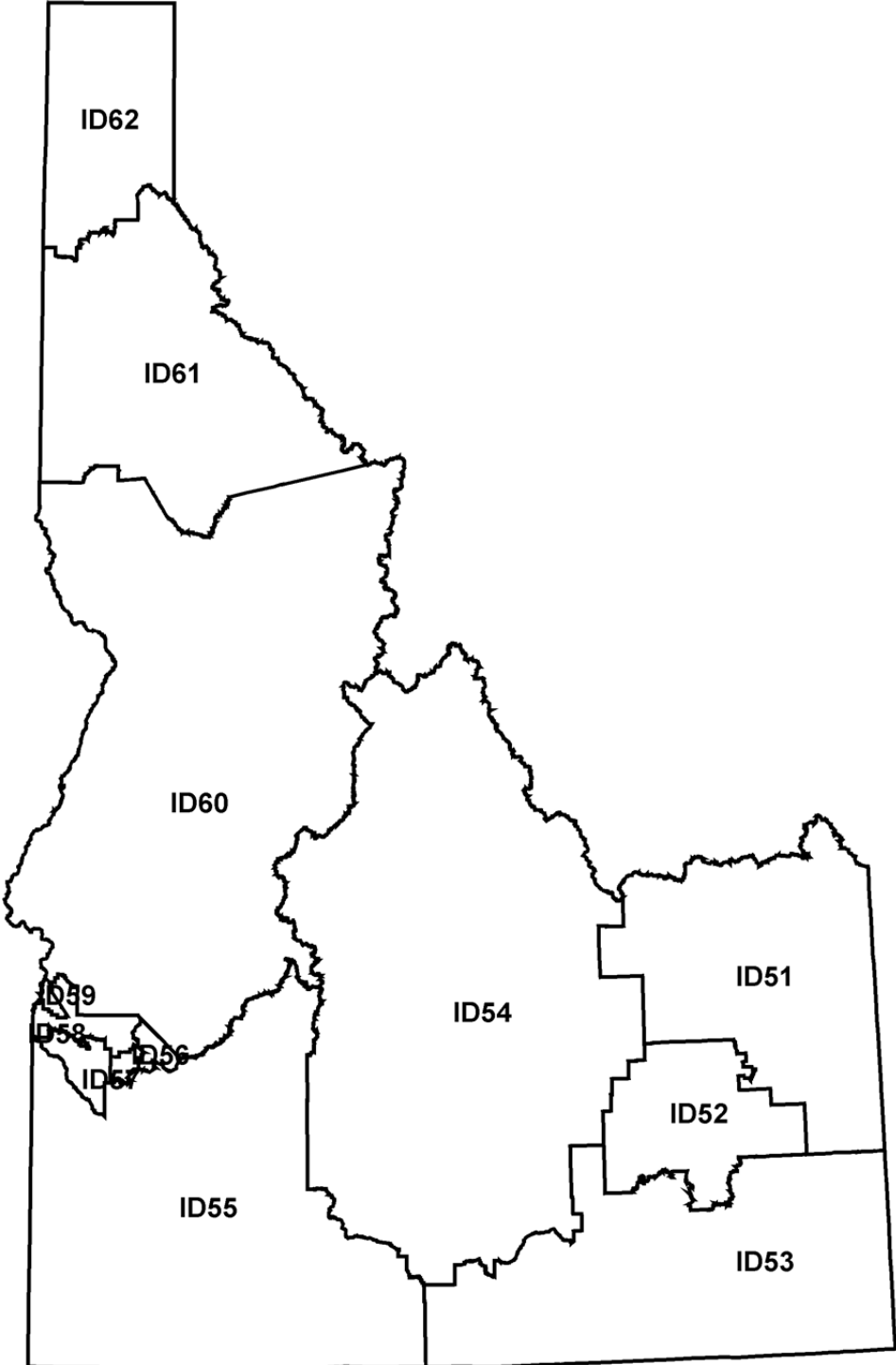
2005-2013 NSDUH State Sampling Regions: Georgia



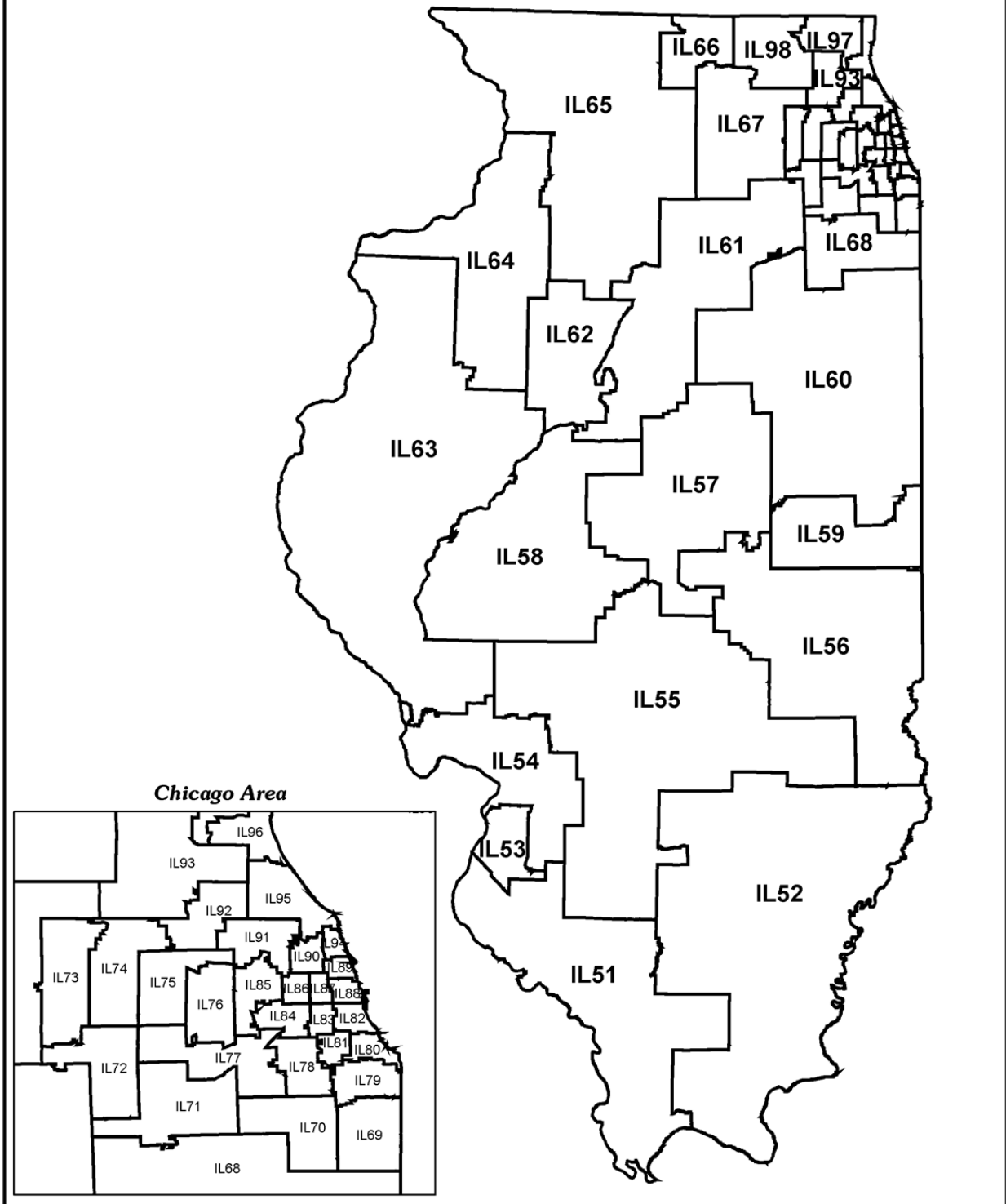
2005-2013 NSDUH State Sampling Regions: Hawaii



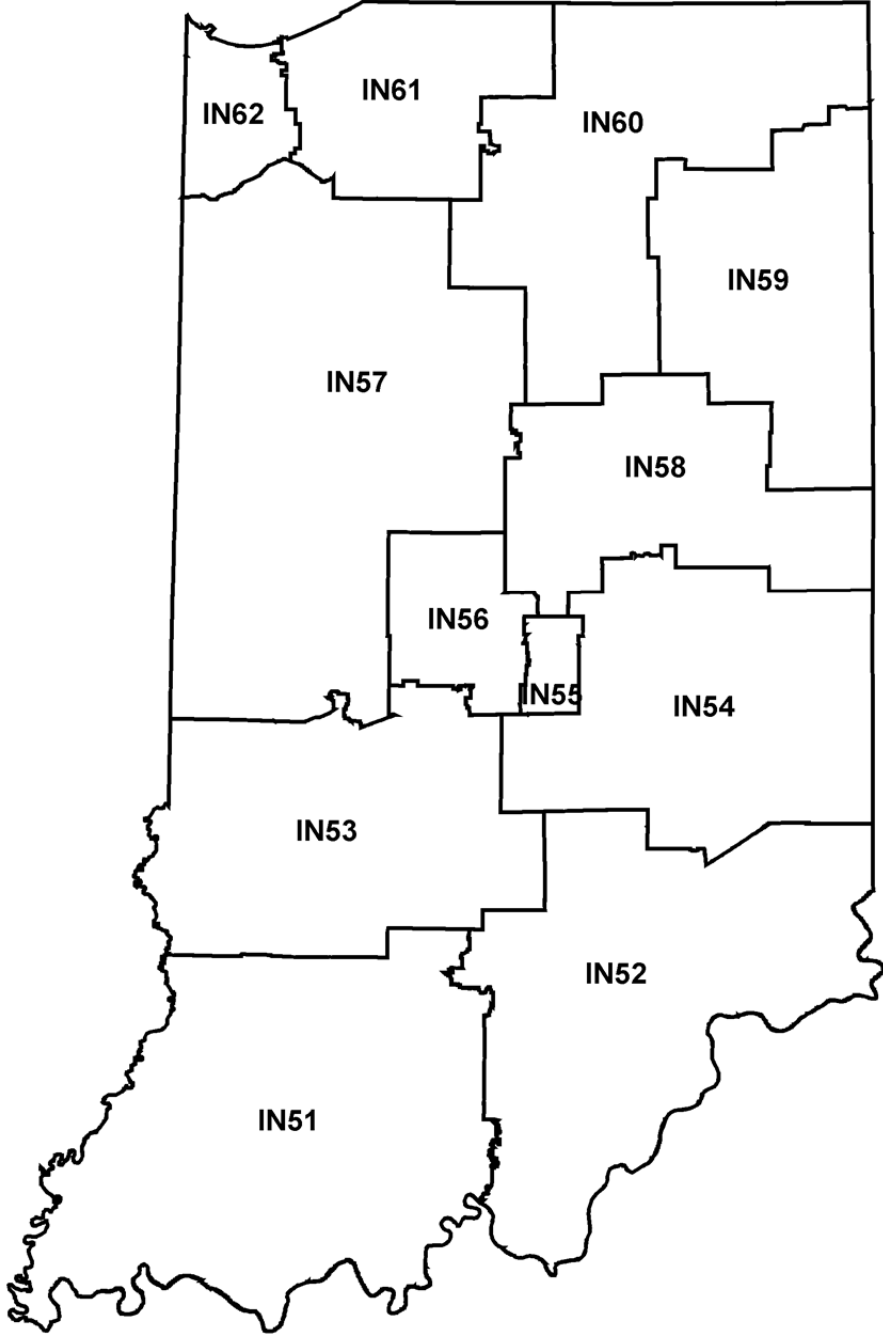
2005-2013 NSDUH State Sampling Regions: Idaho



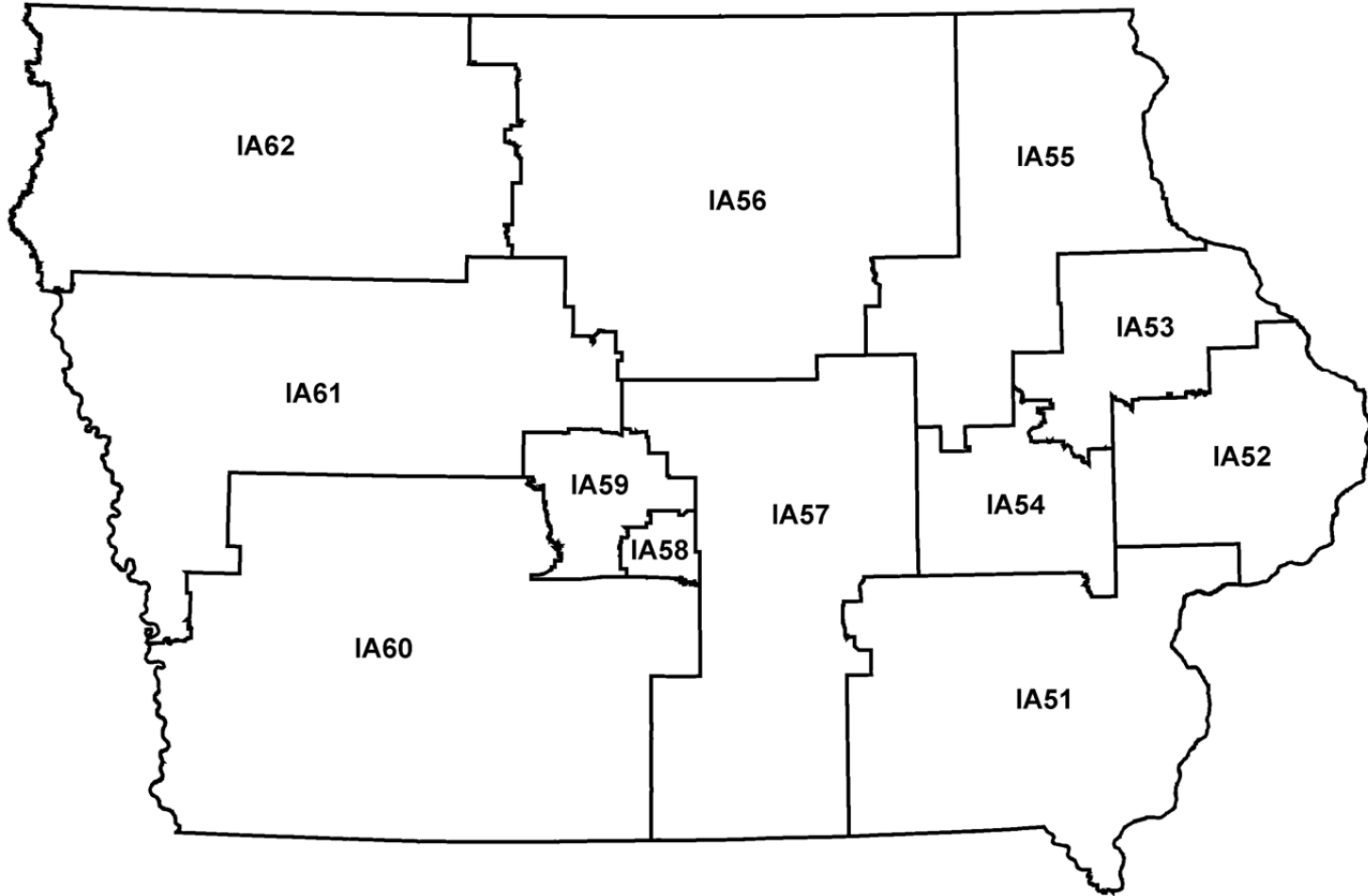
2005-2013 NSDUH State Sampling Regions: Illinois



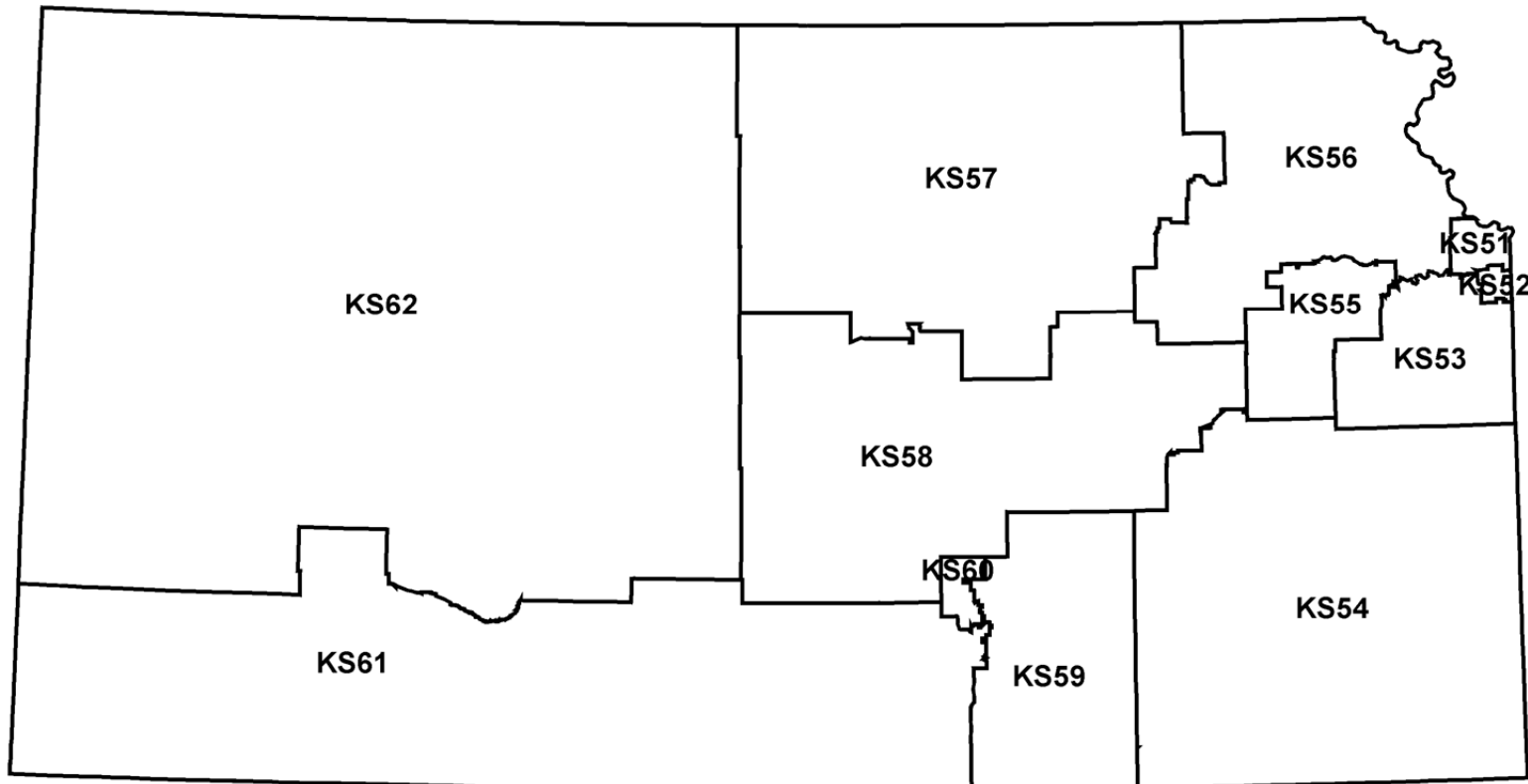
2005-2013 NSDUH State Sampling Regions: Indiana



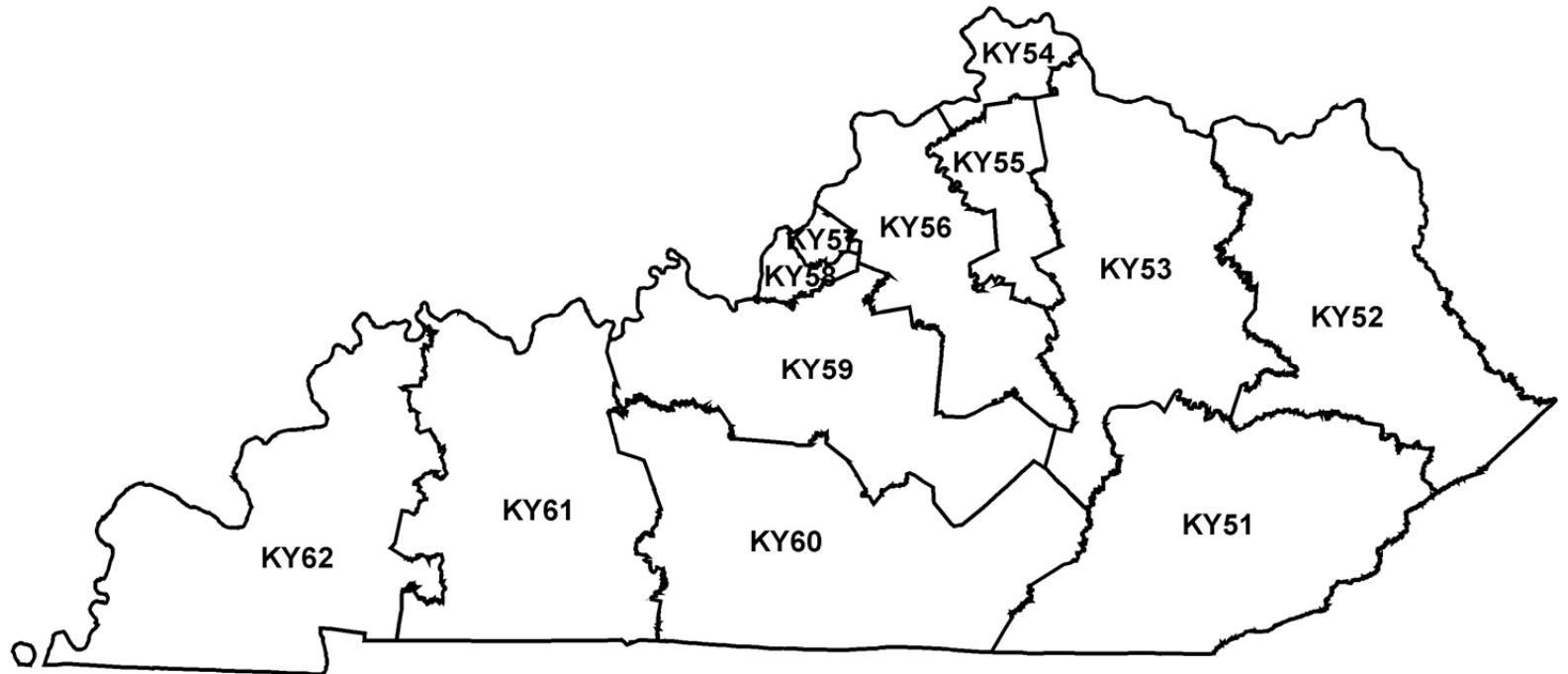
2005-2013 NSDUH State Sampling Regions: Iowa



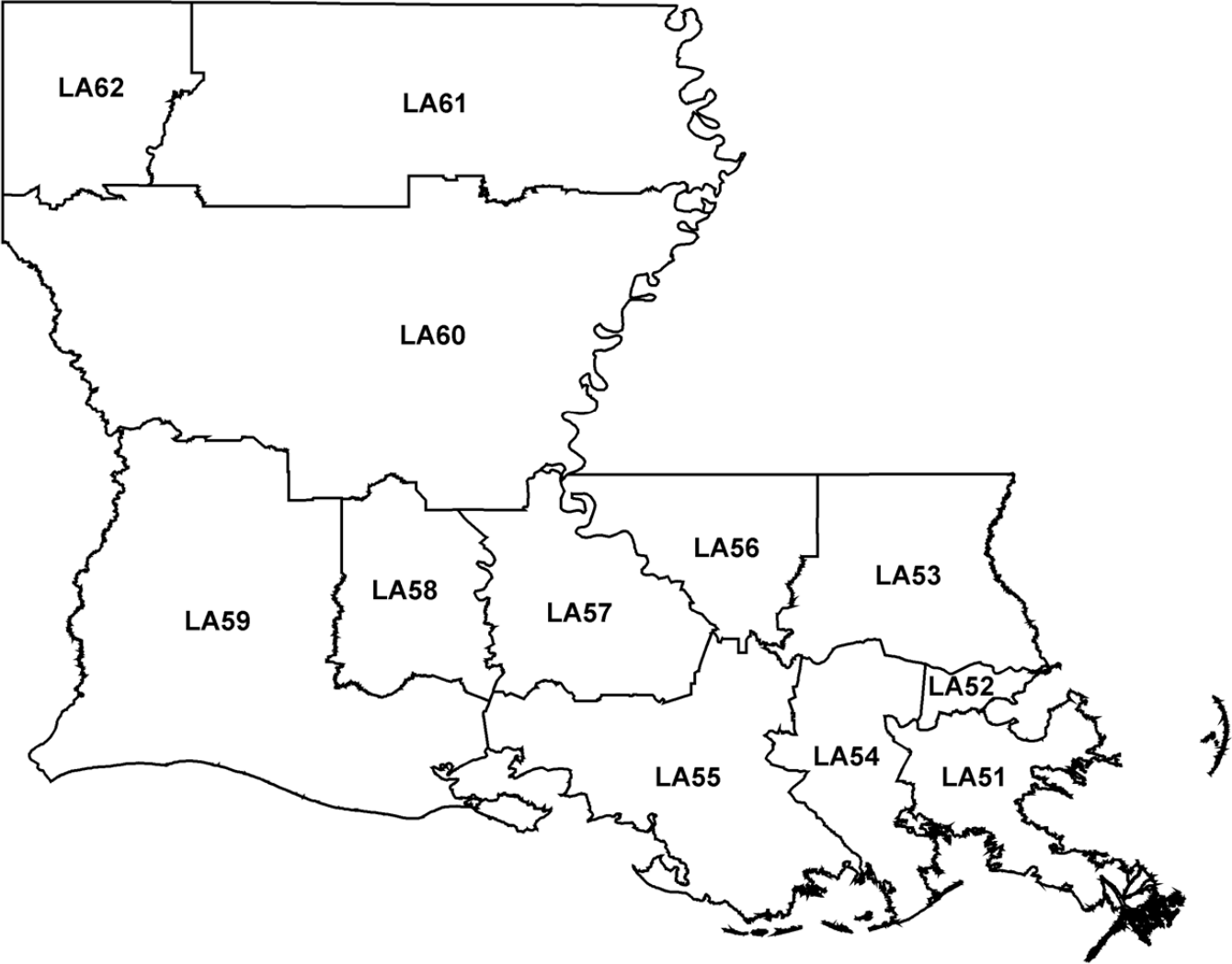
2005-2013 NSDUH State Sampling Regions: Kansas



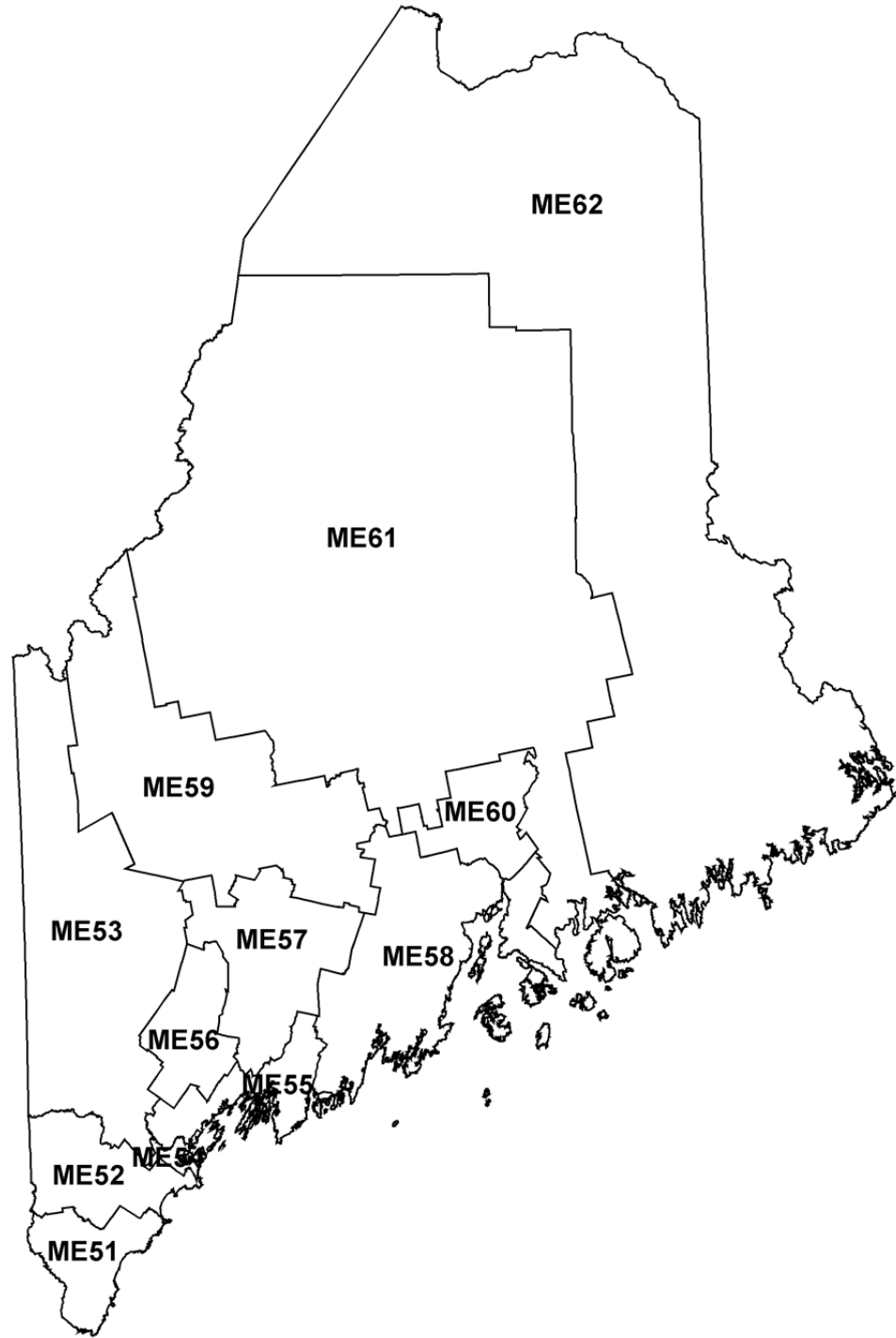
2005-2013 NSDUH State Sampling Regions: Kentucky



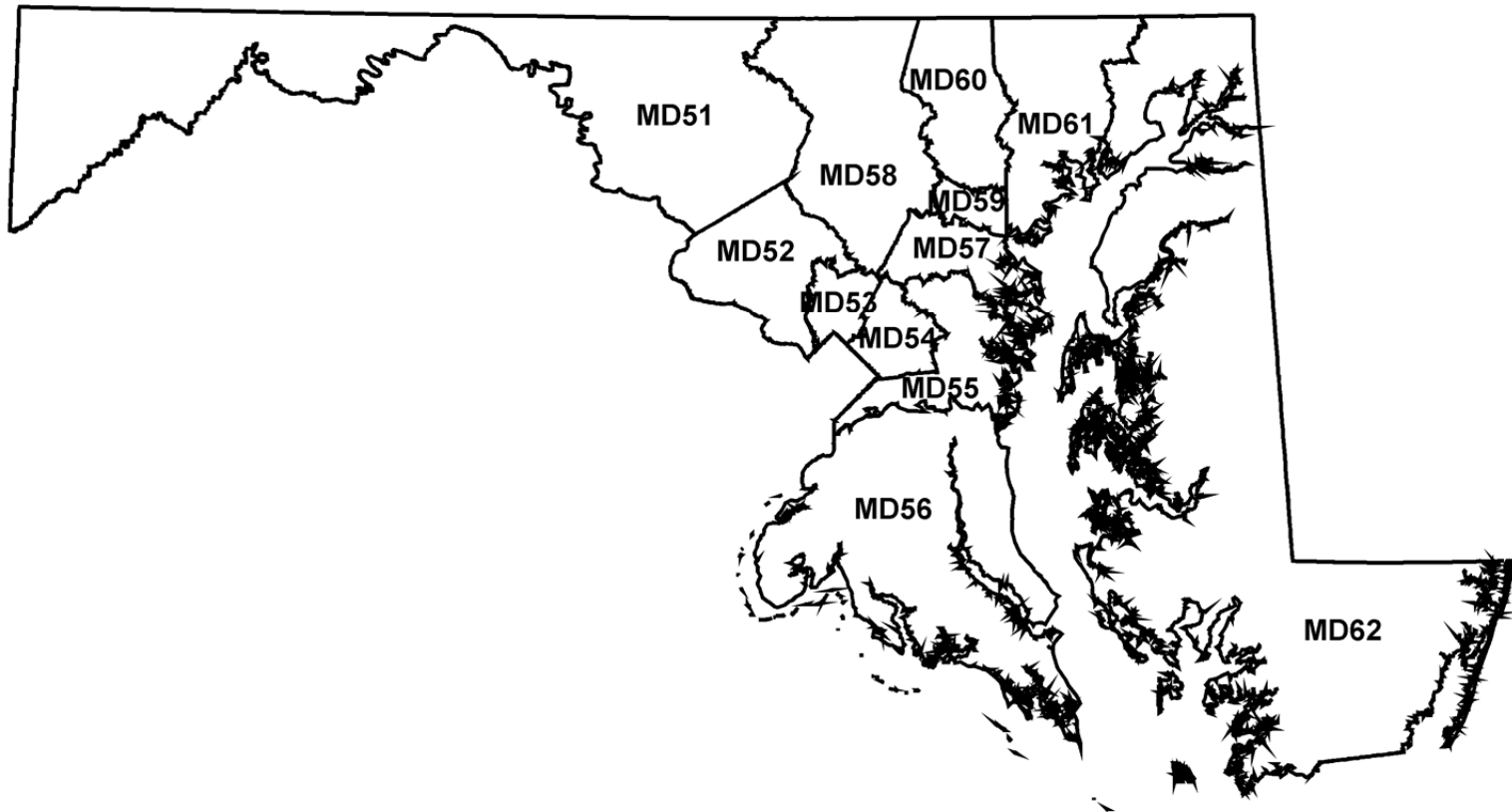
2005-2013 NSDUH State Sampling Regions: Louisiana



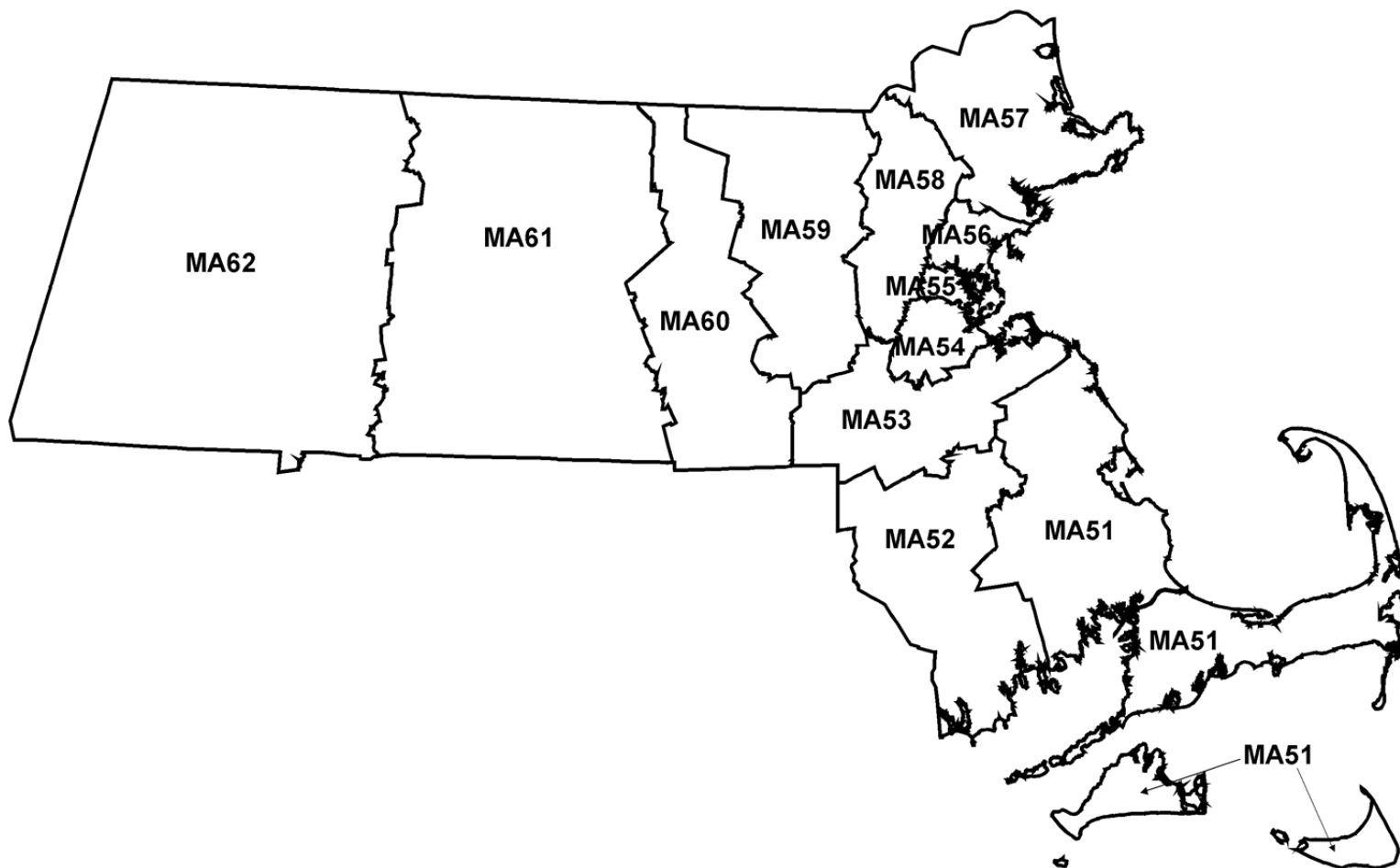
2005-2013 NSDUH State Sampling Regions: Maine



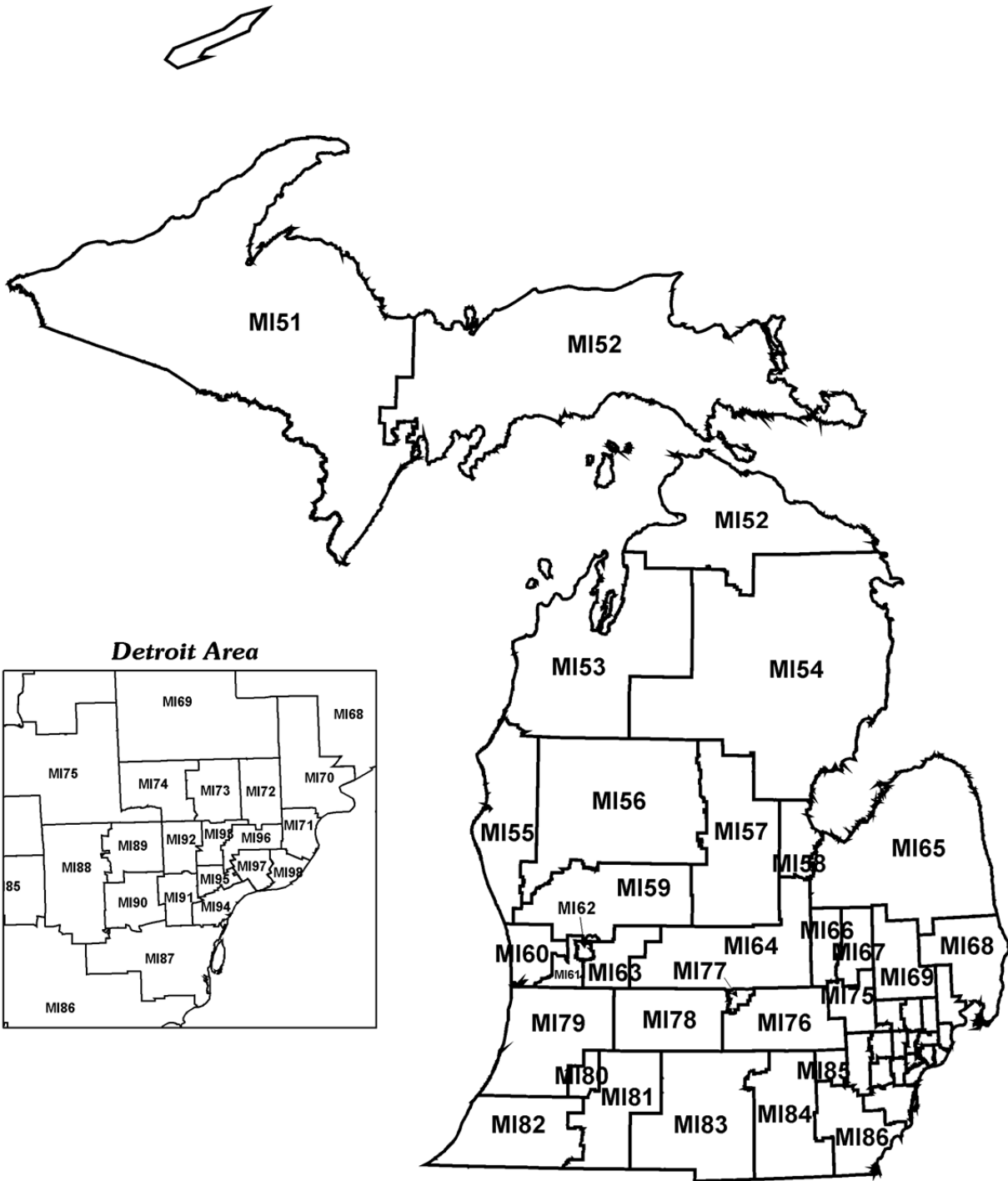
2005-2013 NSDUH State Sampling Regions: Maryland



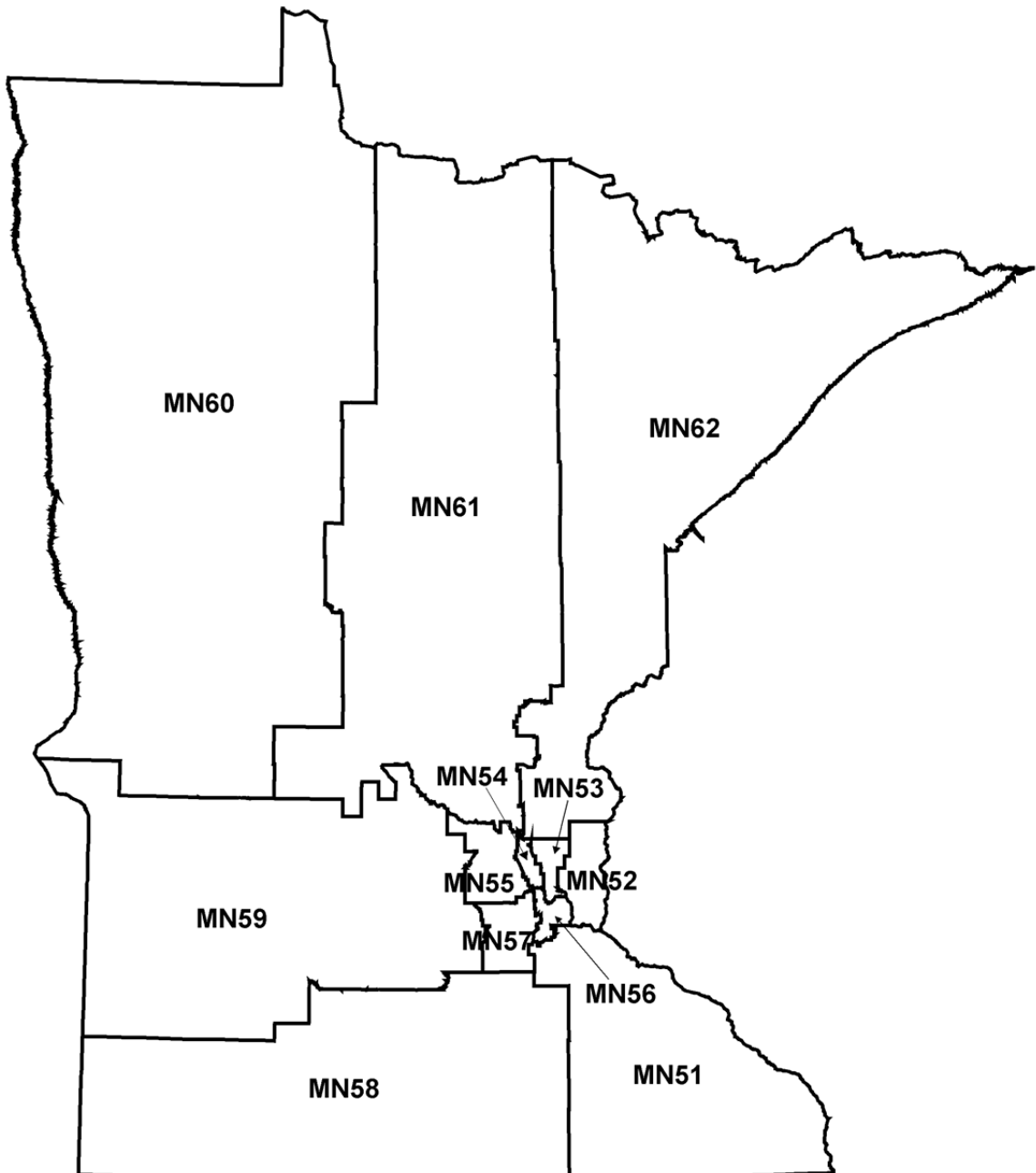
2005-2013 NSDUH State Sampling Regions: Massachusetts



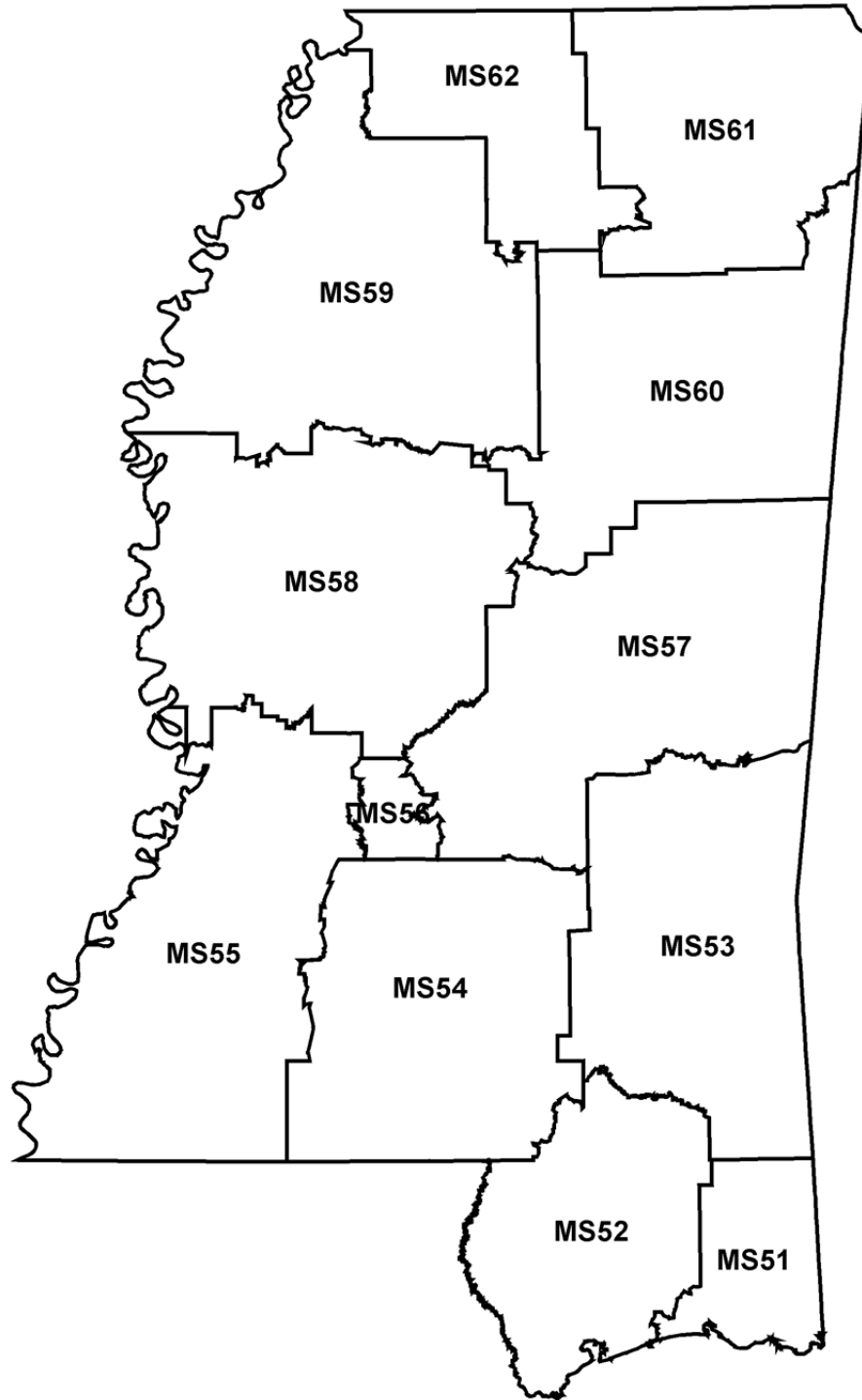
2005-2013 NSDUH State Sampling Regions: Michigan



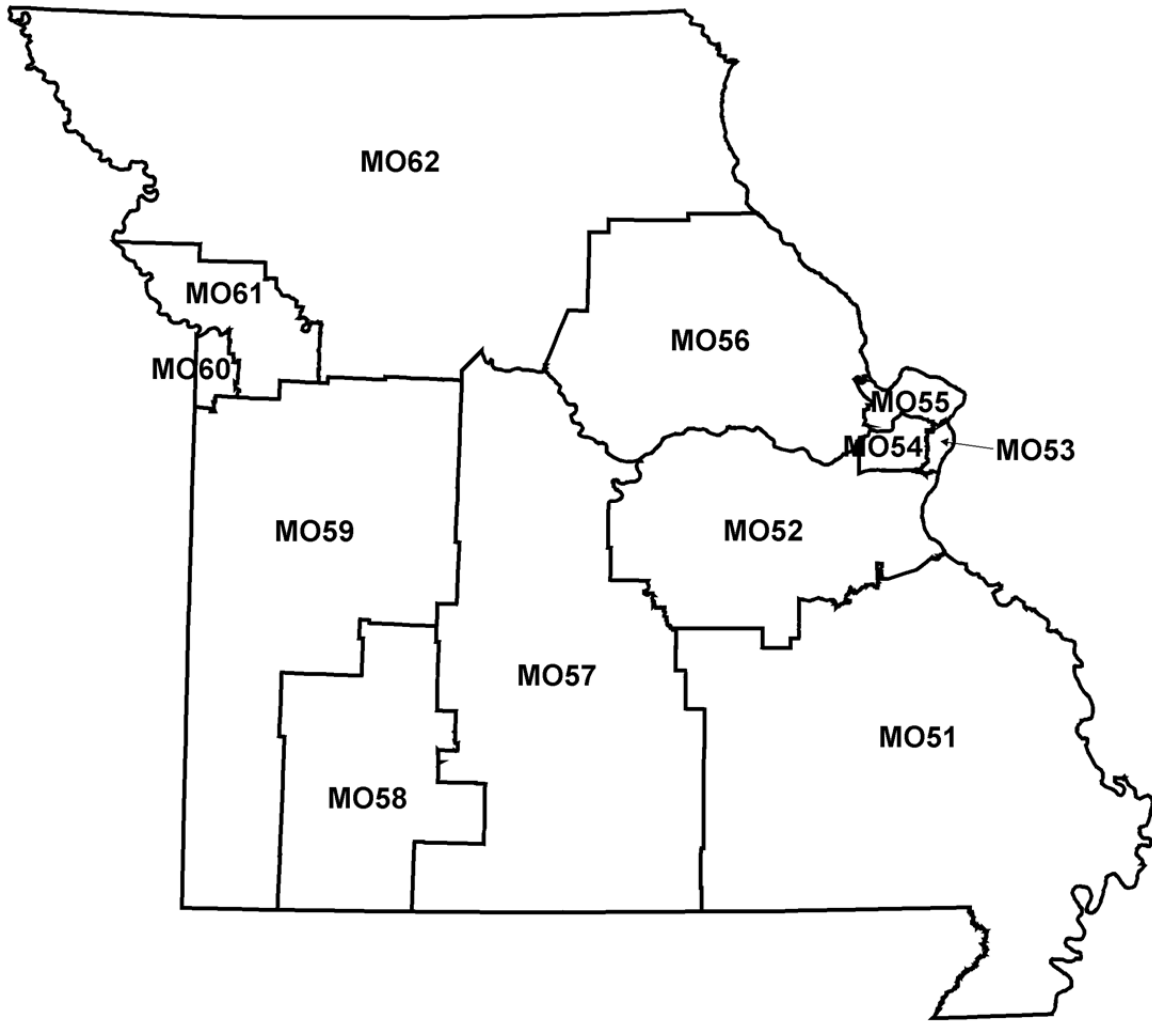
2005-2013 NSDUH State Sampling Regions: Minnesota



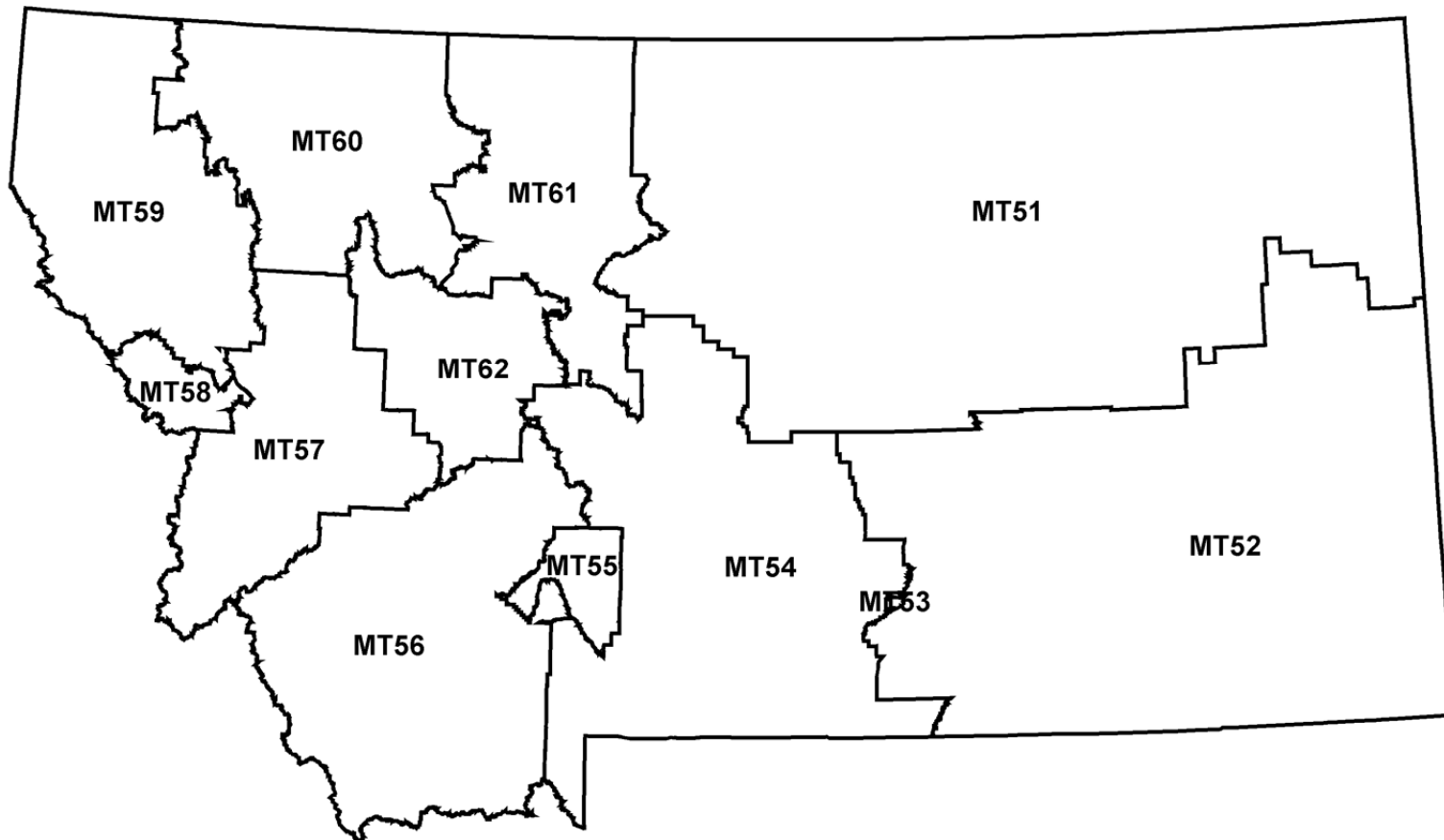
2005-2013 NSDUH State Sampling Regions: Mississippi



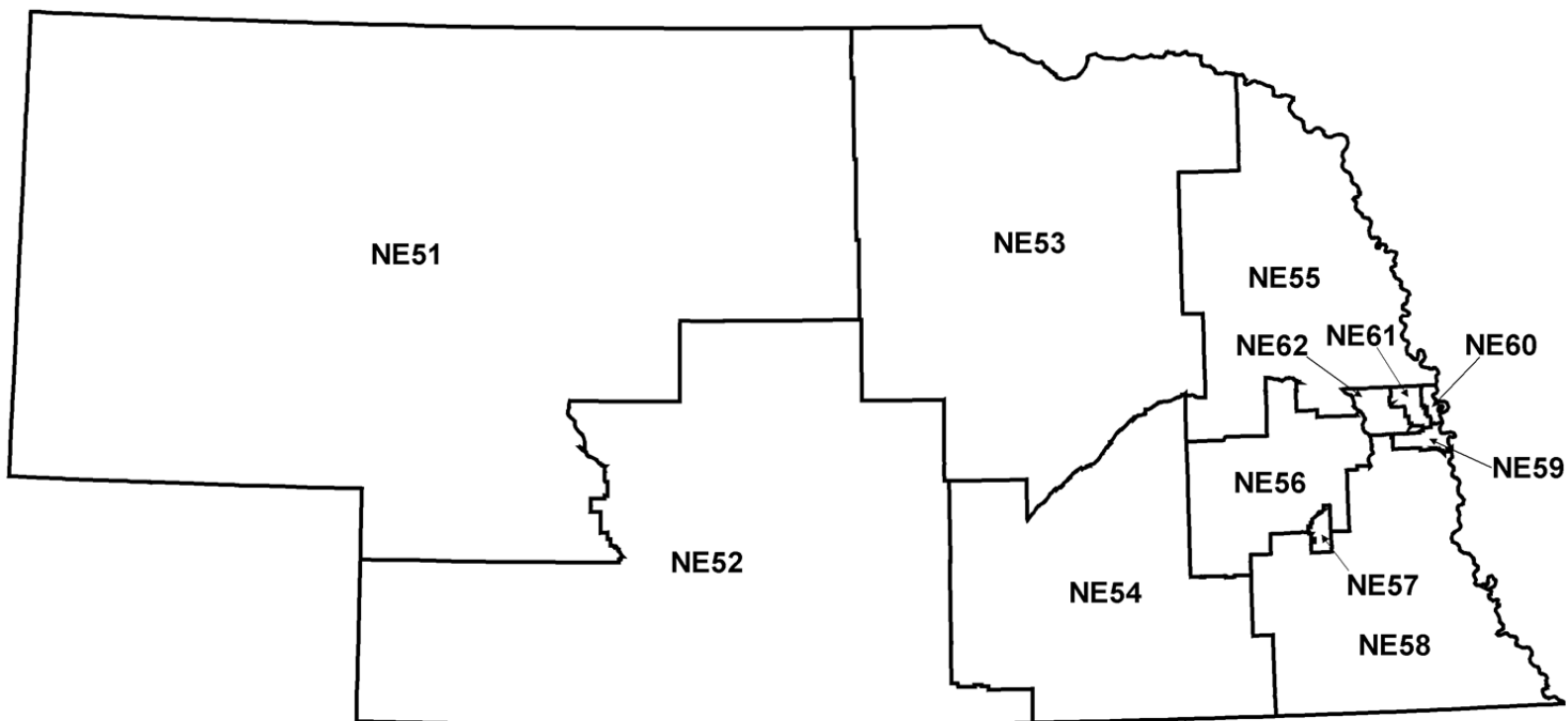
2005-2013 NSDUH State Sampling Regions: Missouri



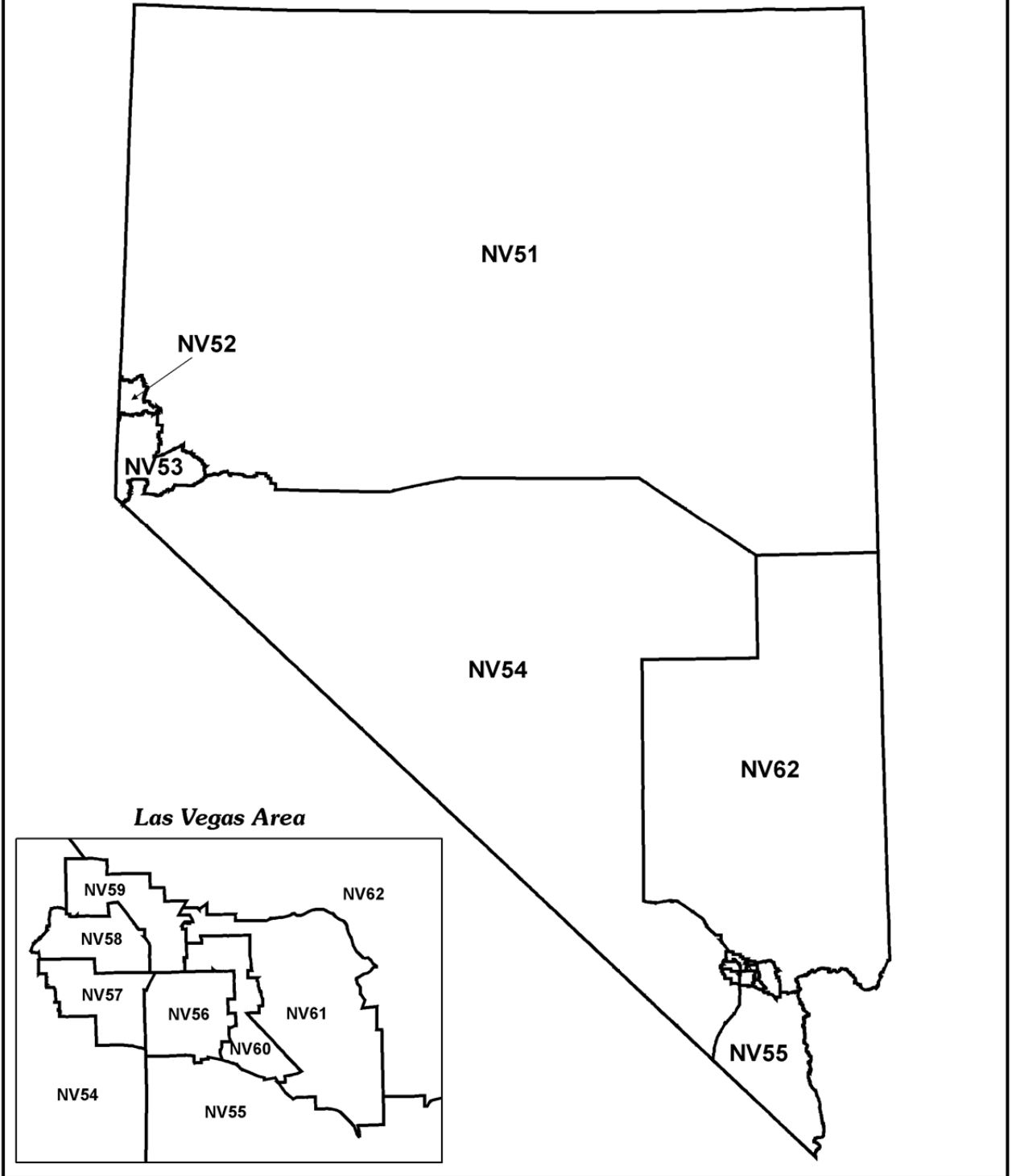
2005-2013 NSDUH State Sampling Regions: Montana



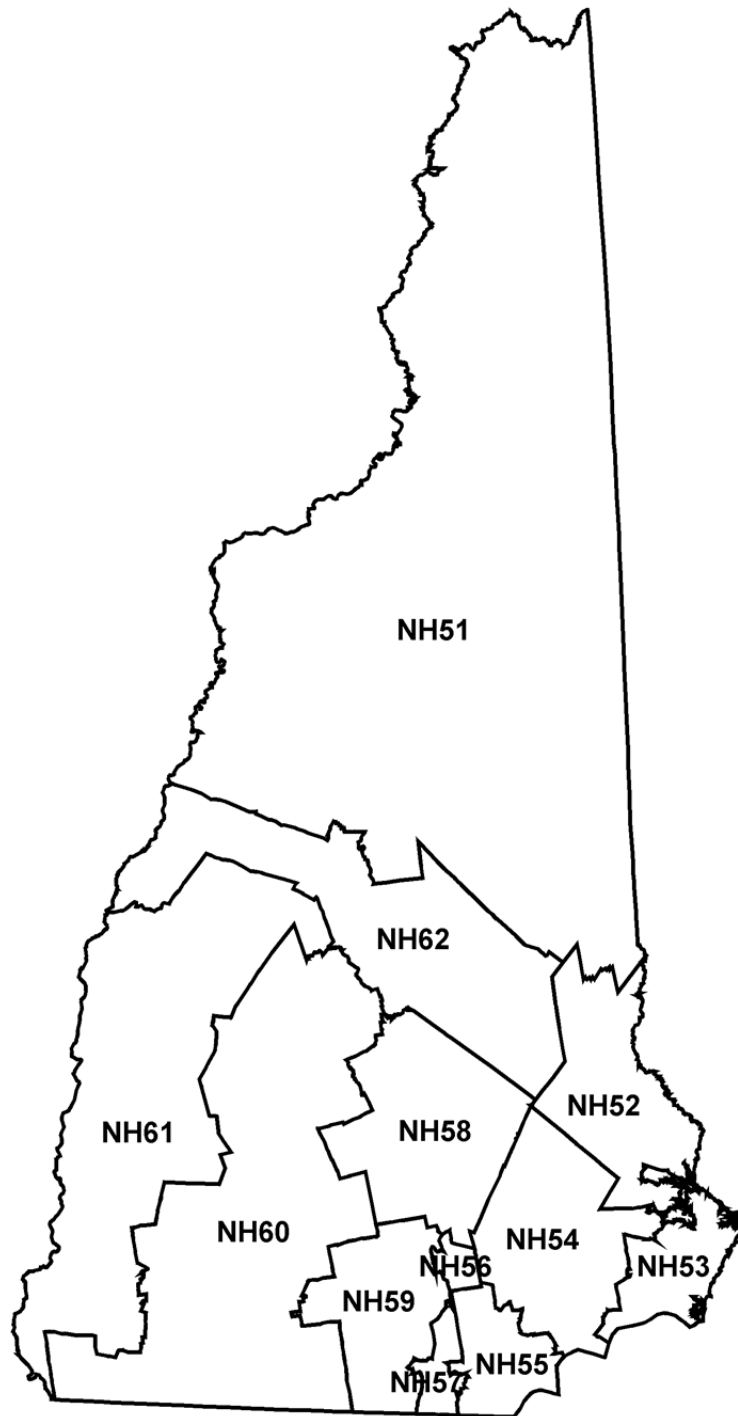
2005-2013 NSDUH State Sampling Regions: Nebraska



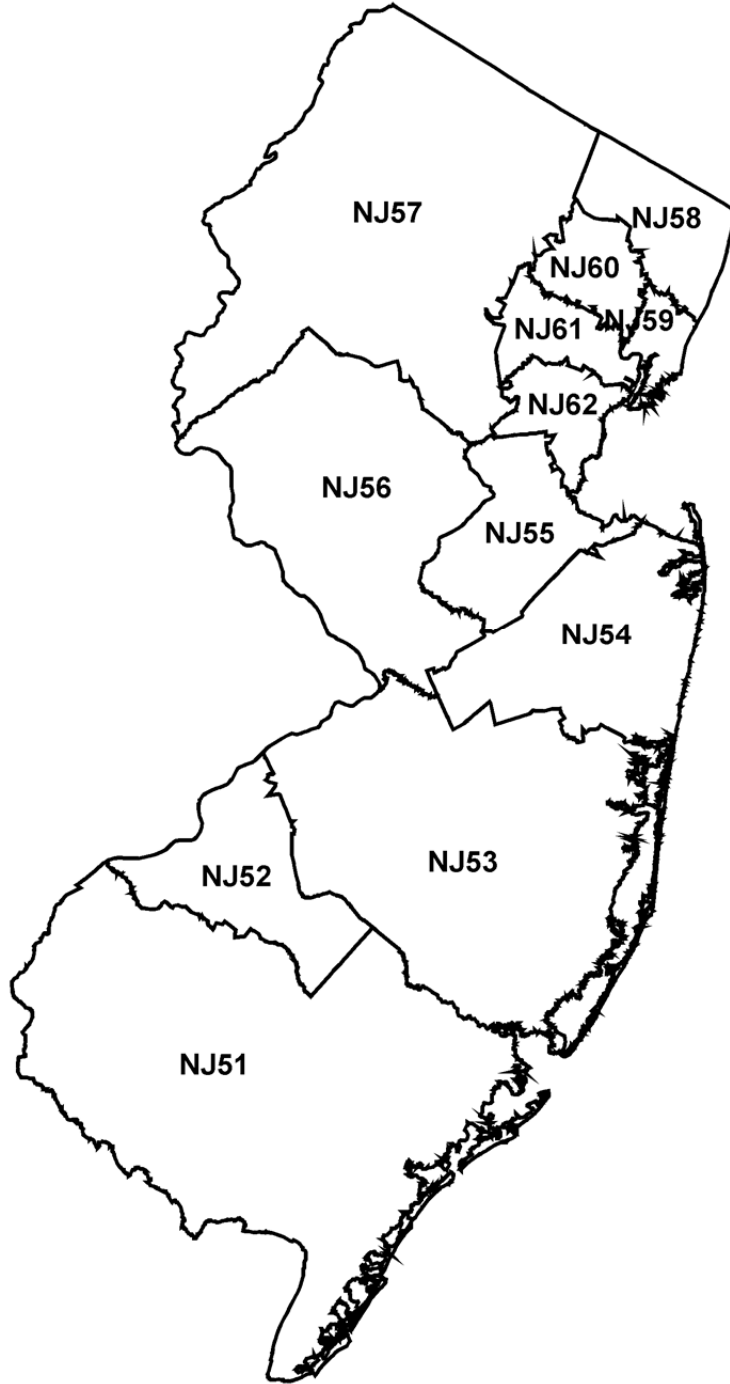
2005-2013 NSDUH State Sampling Regions: Nevada



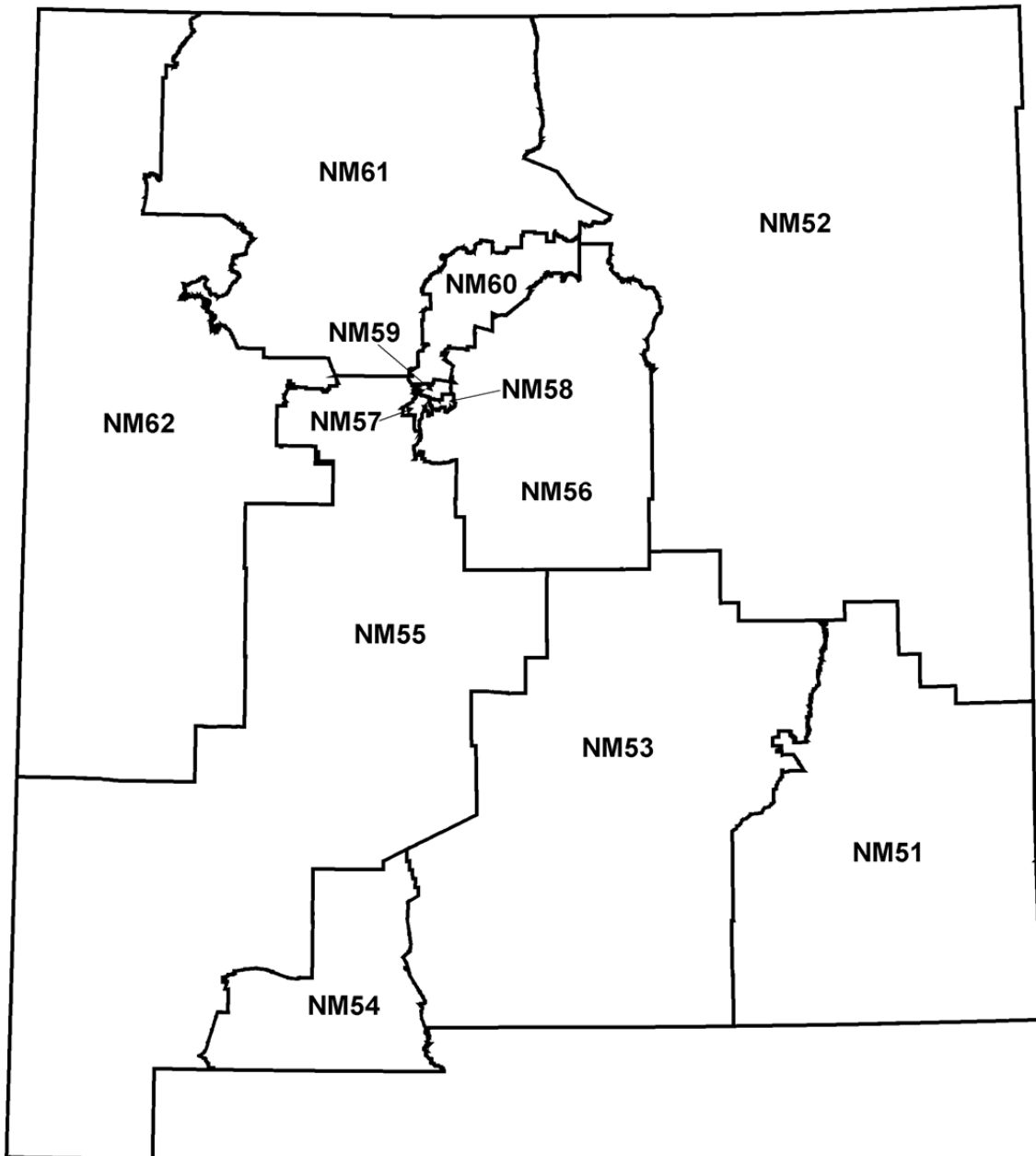
2005-2013 NSDUH State Sampling Regions: New Hampshire



2005-2013 NSDUH State Sampling Regions: New Jersey

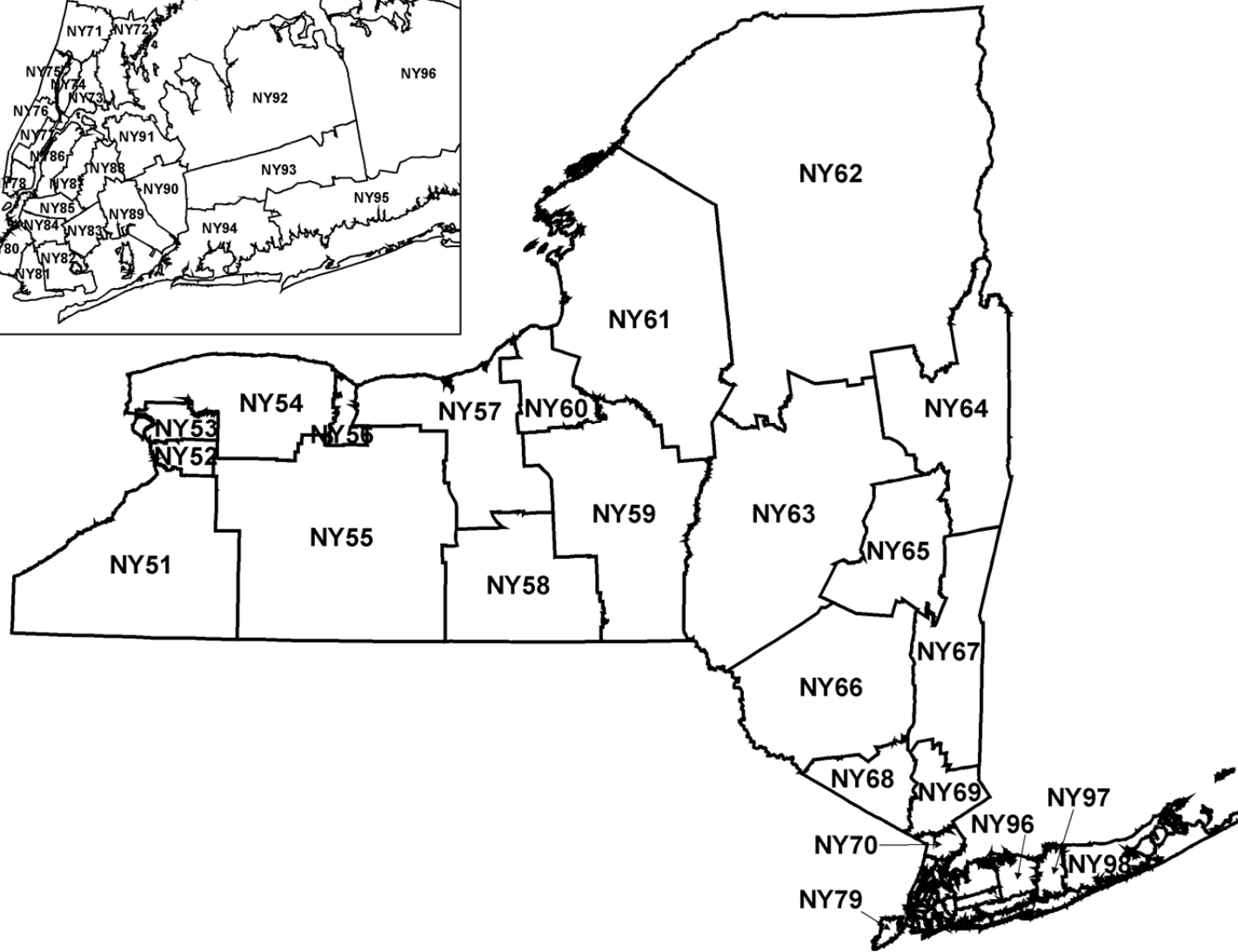
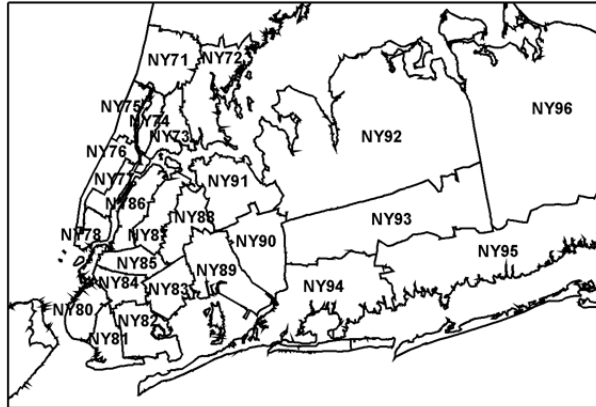


2005-2013 NSDUH State Sampling Regions: New Mexico

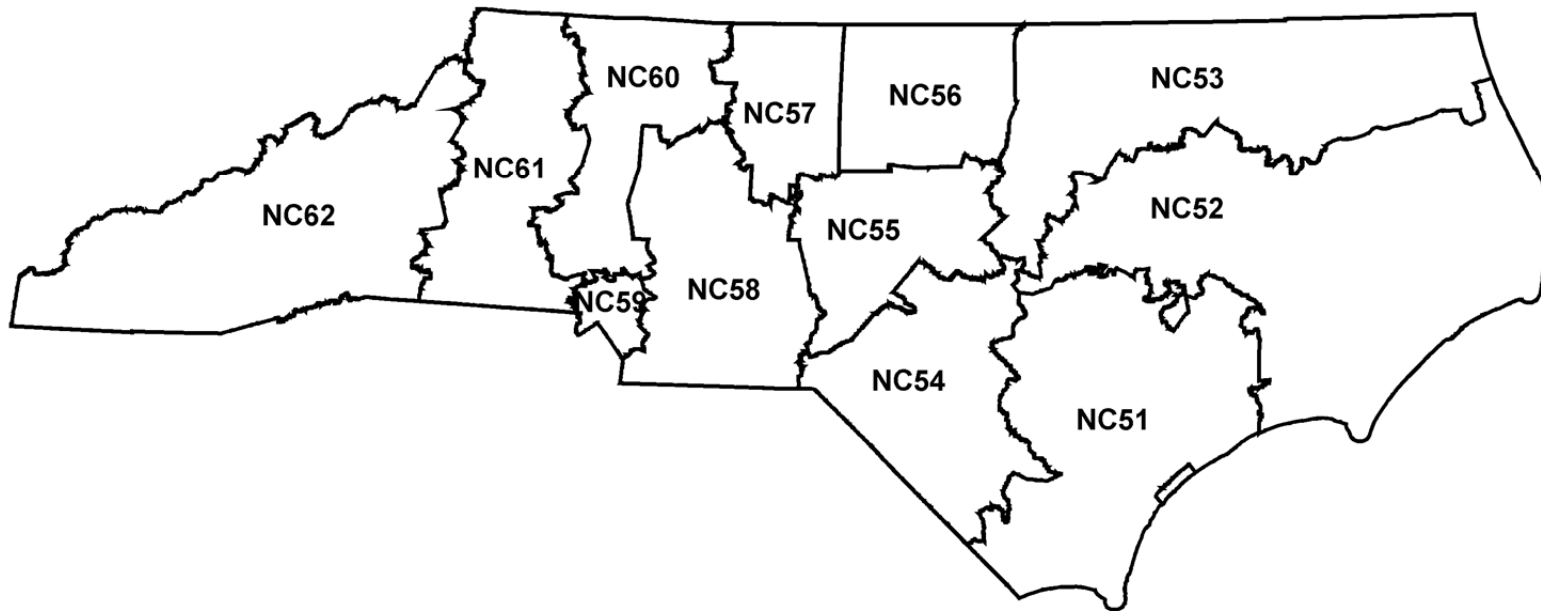


2005-2013 NSDUH State Sampling Regions: New York

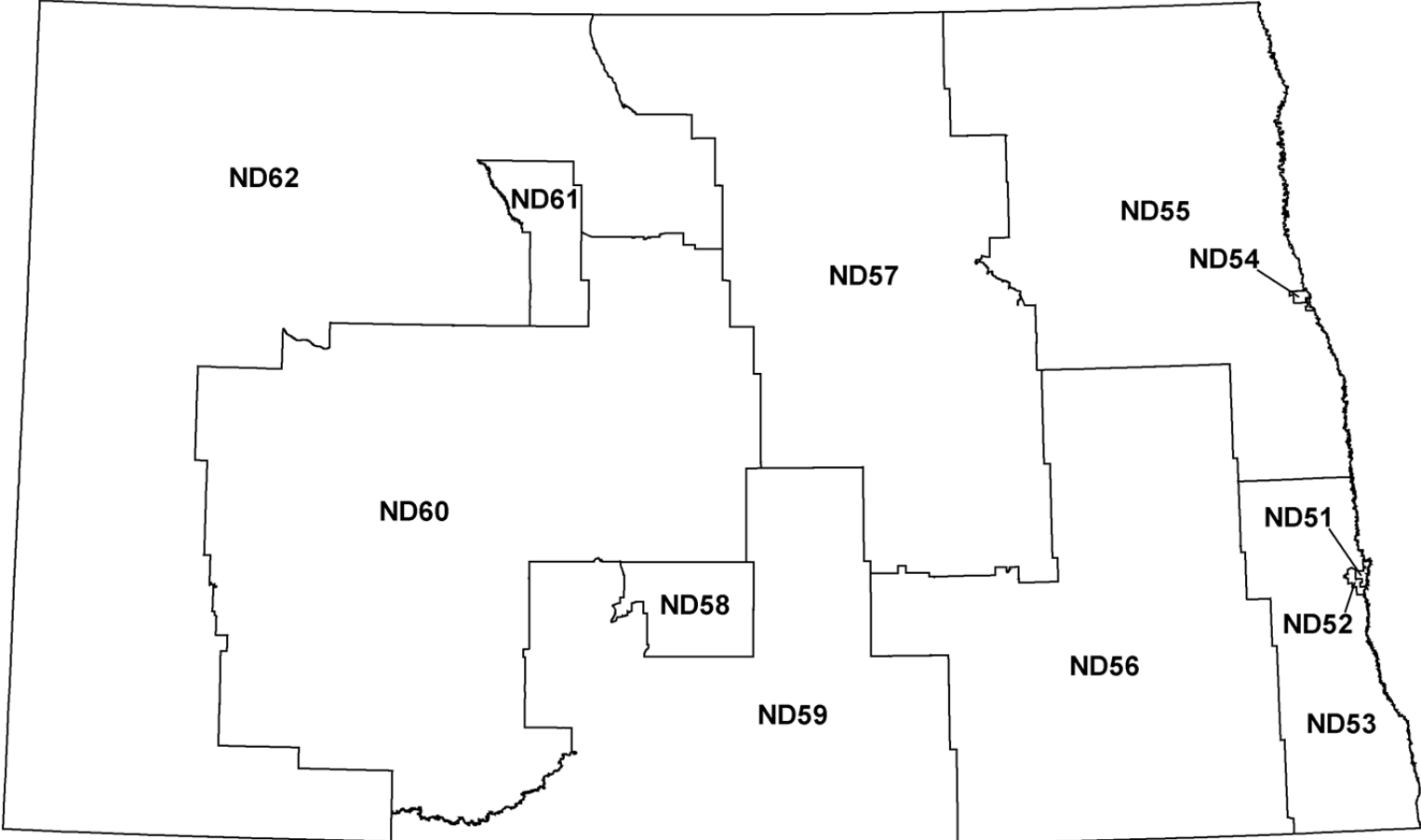
New York City Area



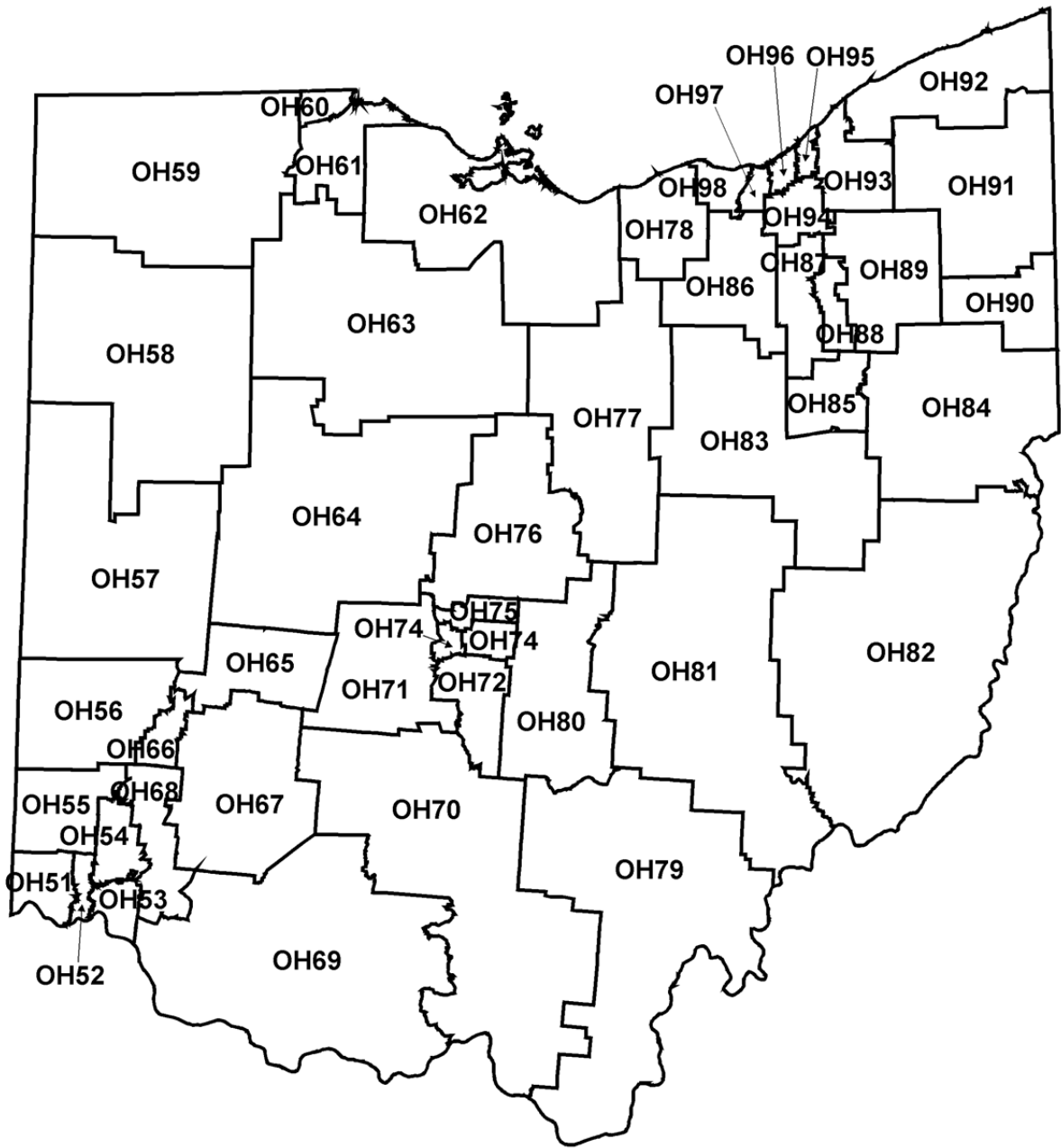
2005-2013 NSDUH State Sampling Regions: North Carolina



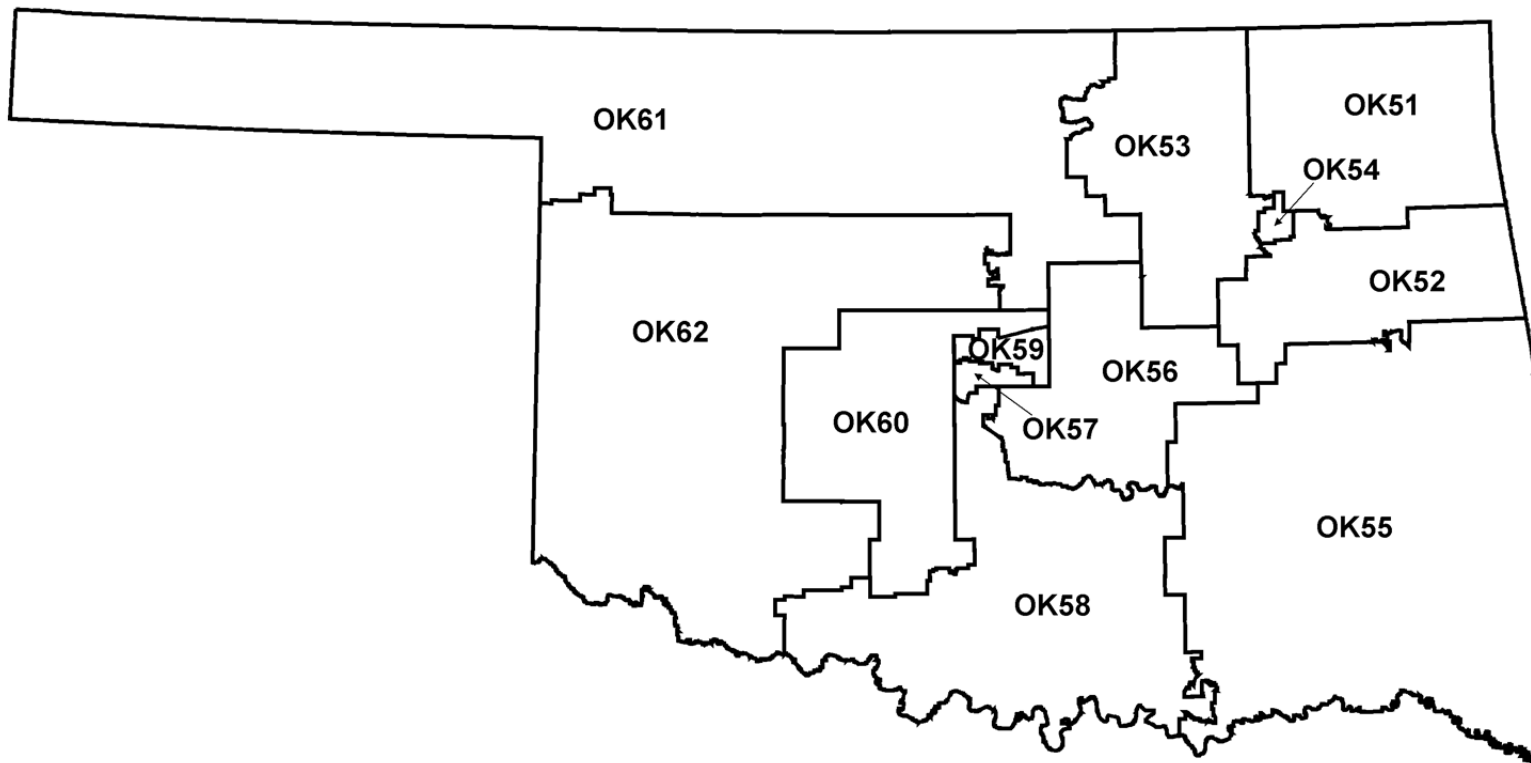
2005-2013 NSDUH State Sampling Regions: North Dakota



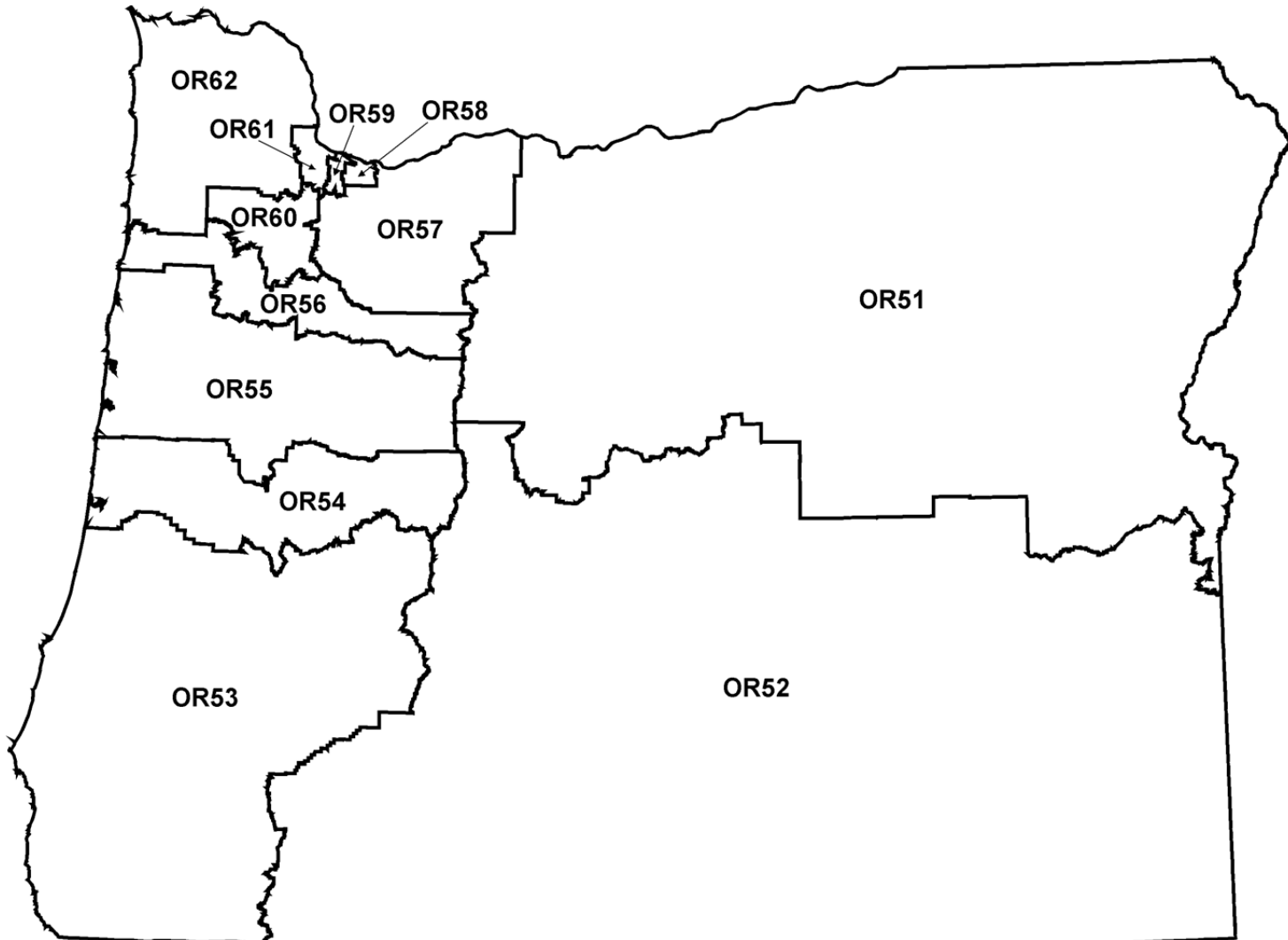
2005-2013 NSDUH State Sampling Regions: Ohio



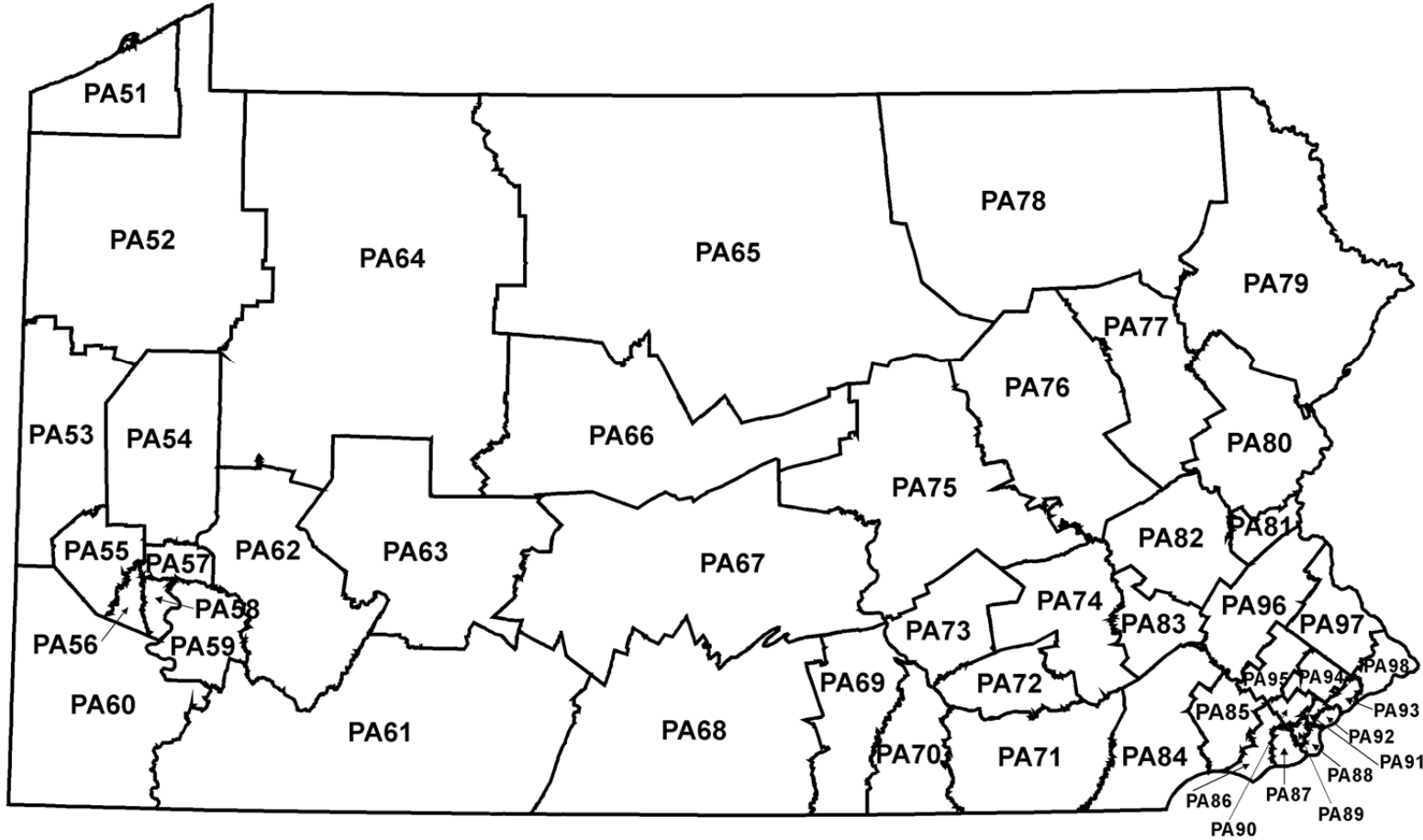
2005-2013 NSDUH State Sampling Regions: Oklahoma



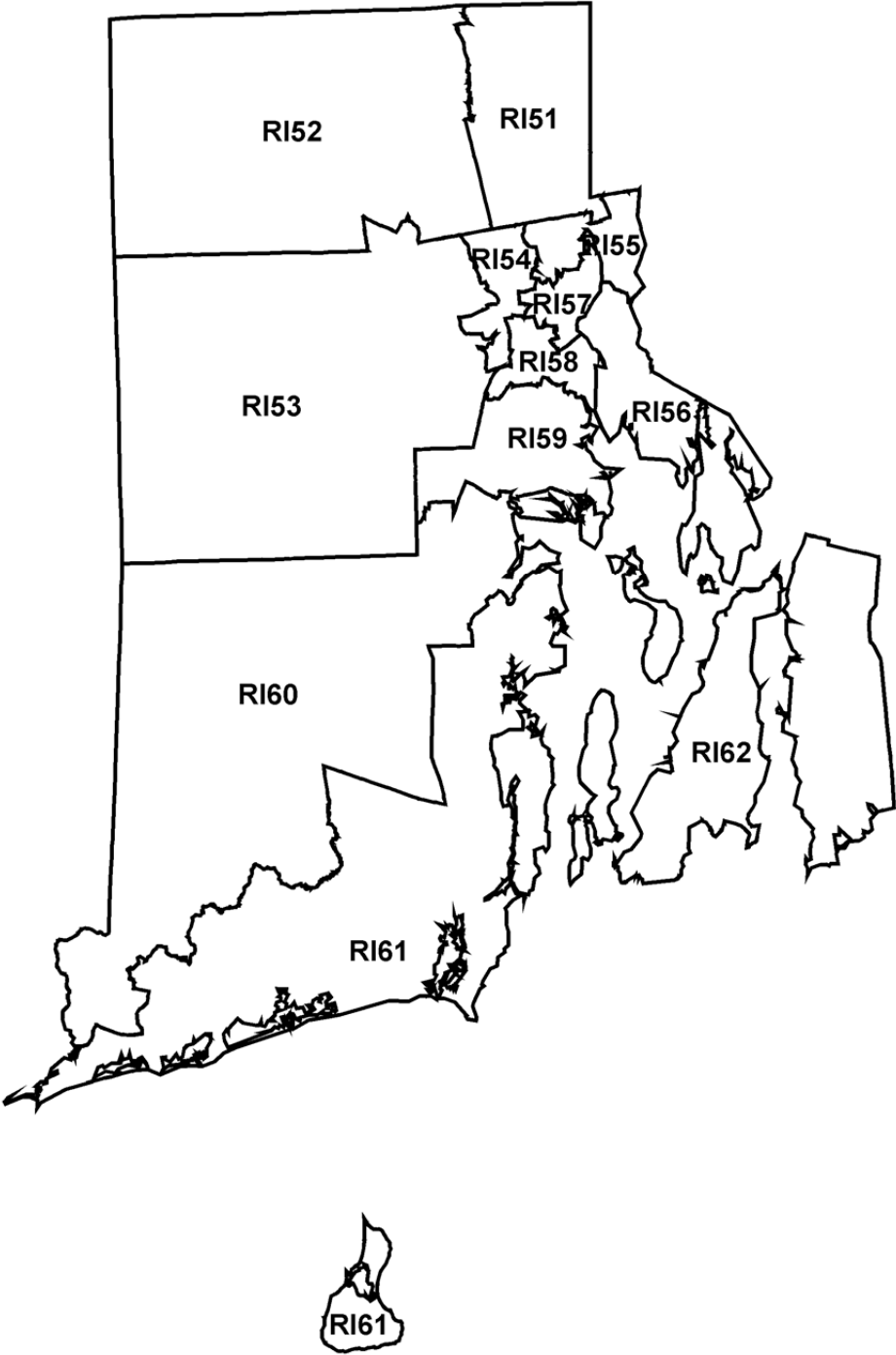
2005-2013 NSDUH State Sampling Regions: Oregon



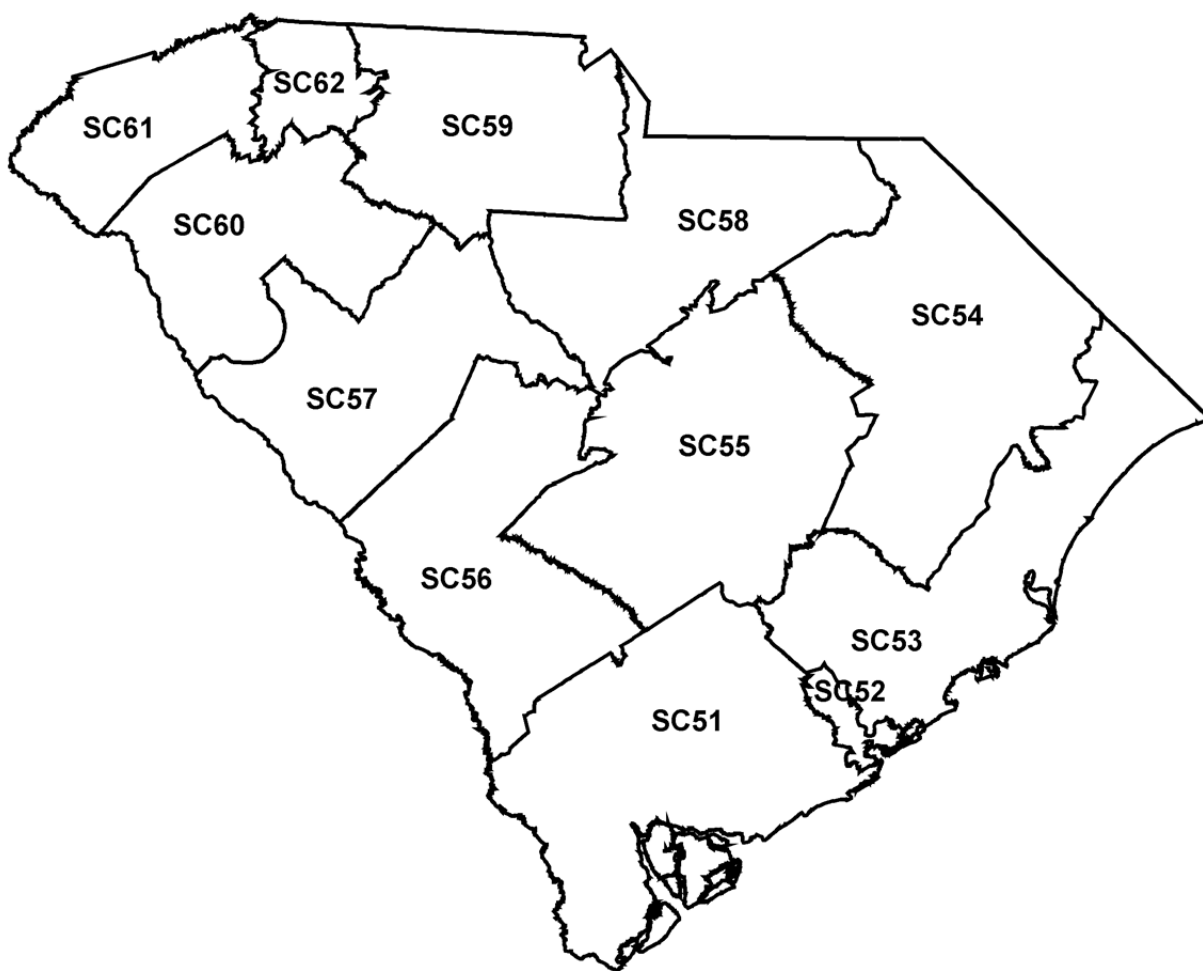
2005-2013 NSDUH State Sampling Regions: Pennsylvania



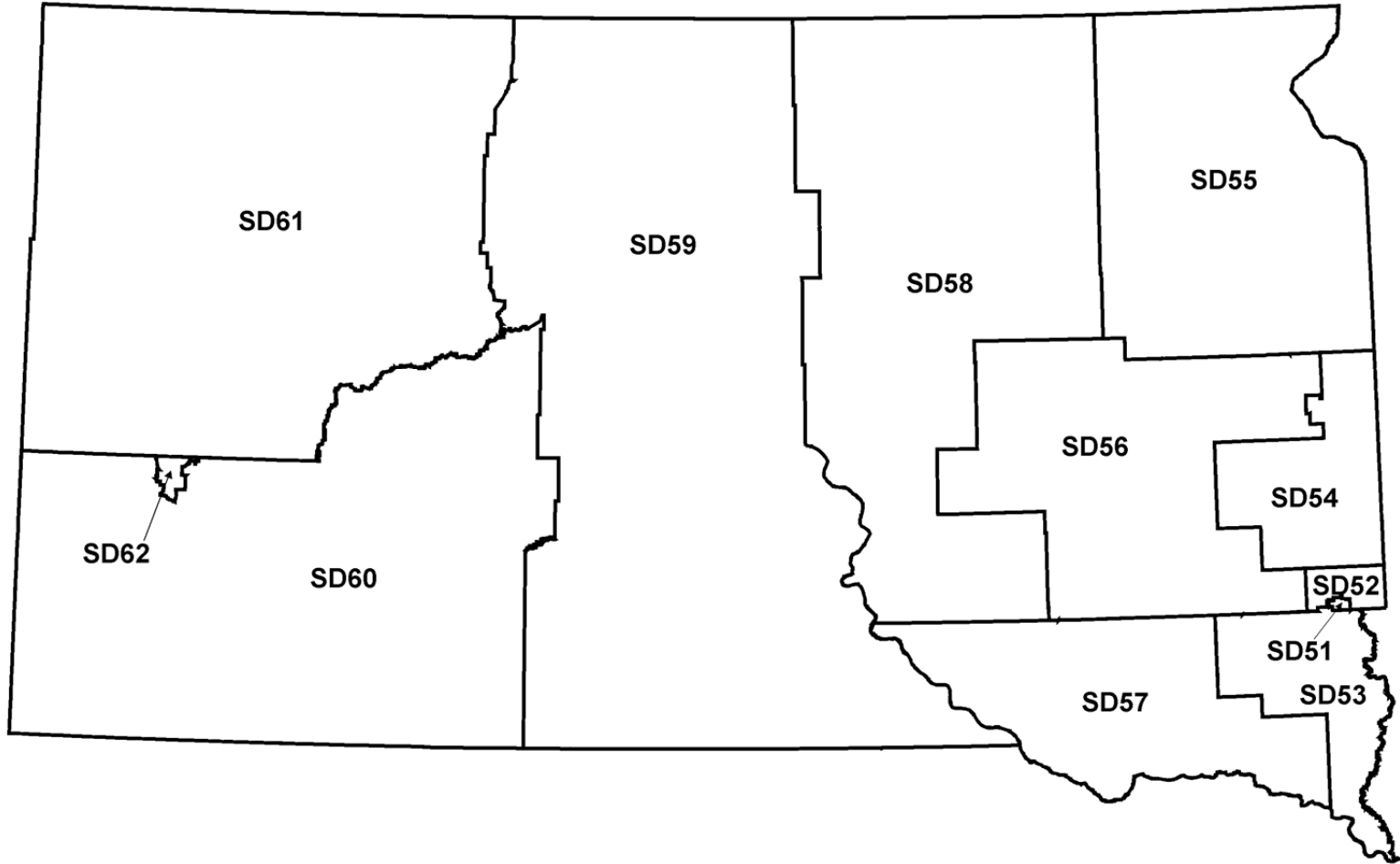
2005-2013 NSDUH State Sampling Regions: Rhode Island



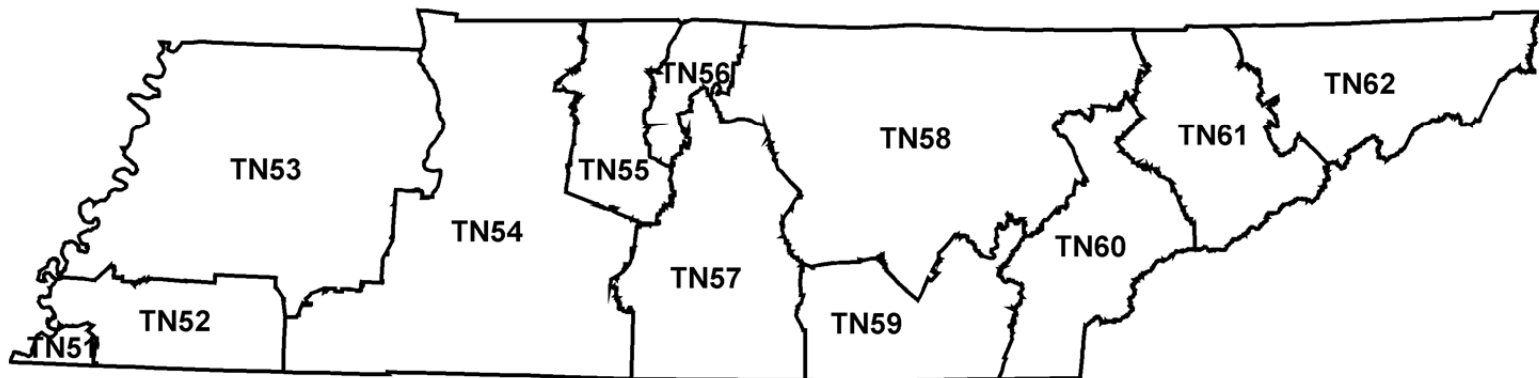
2005-2013 NSDUH State Sampling Regions: South Carolina



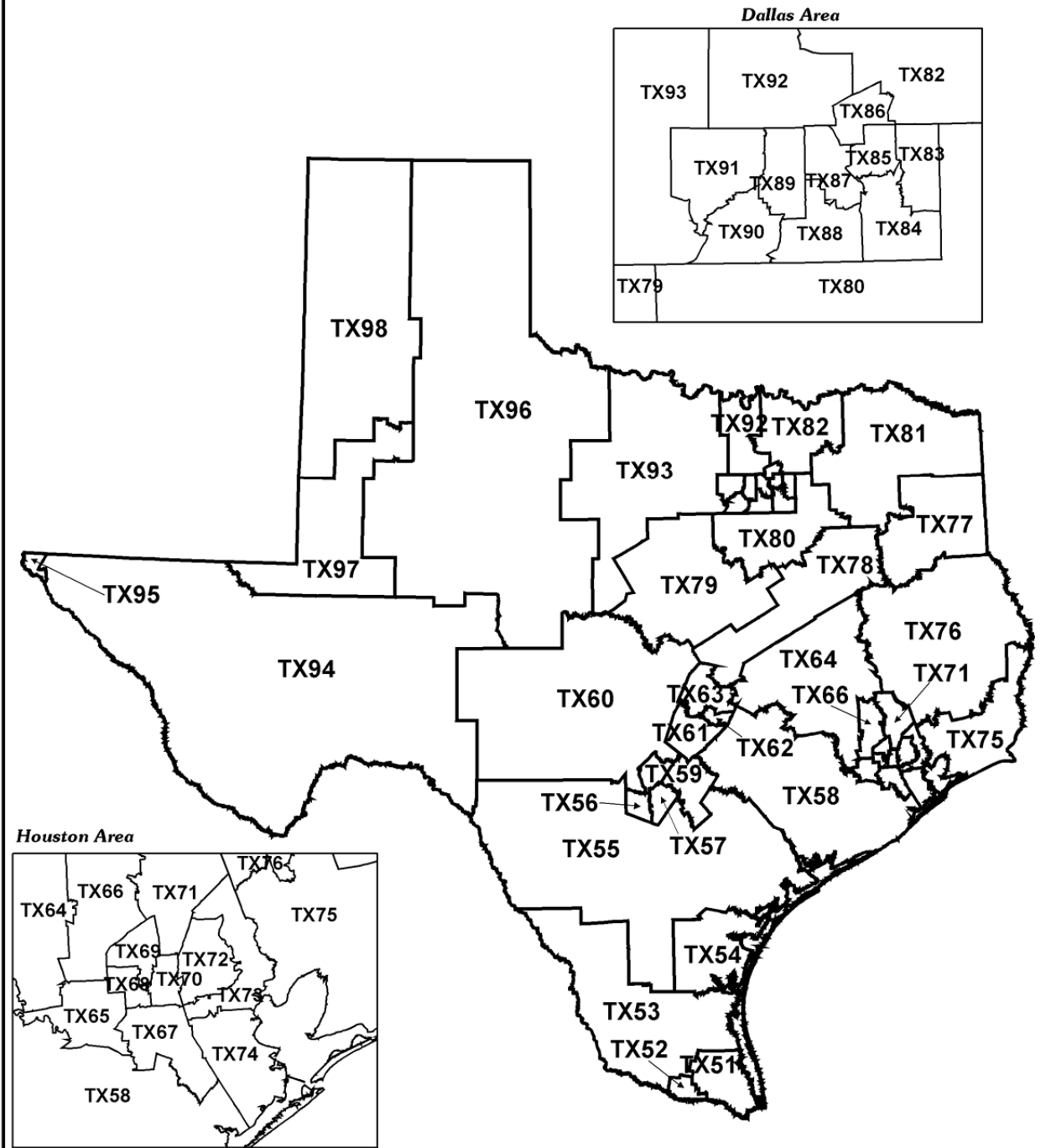
2005-2013 NSDUH State Sampling Regions: South Dakota



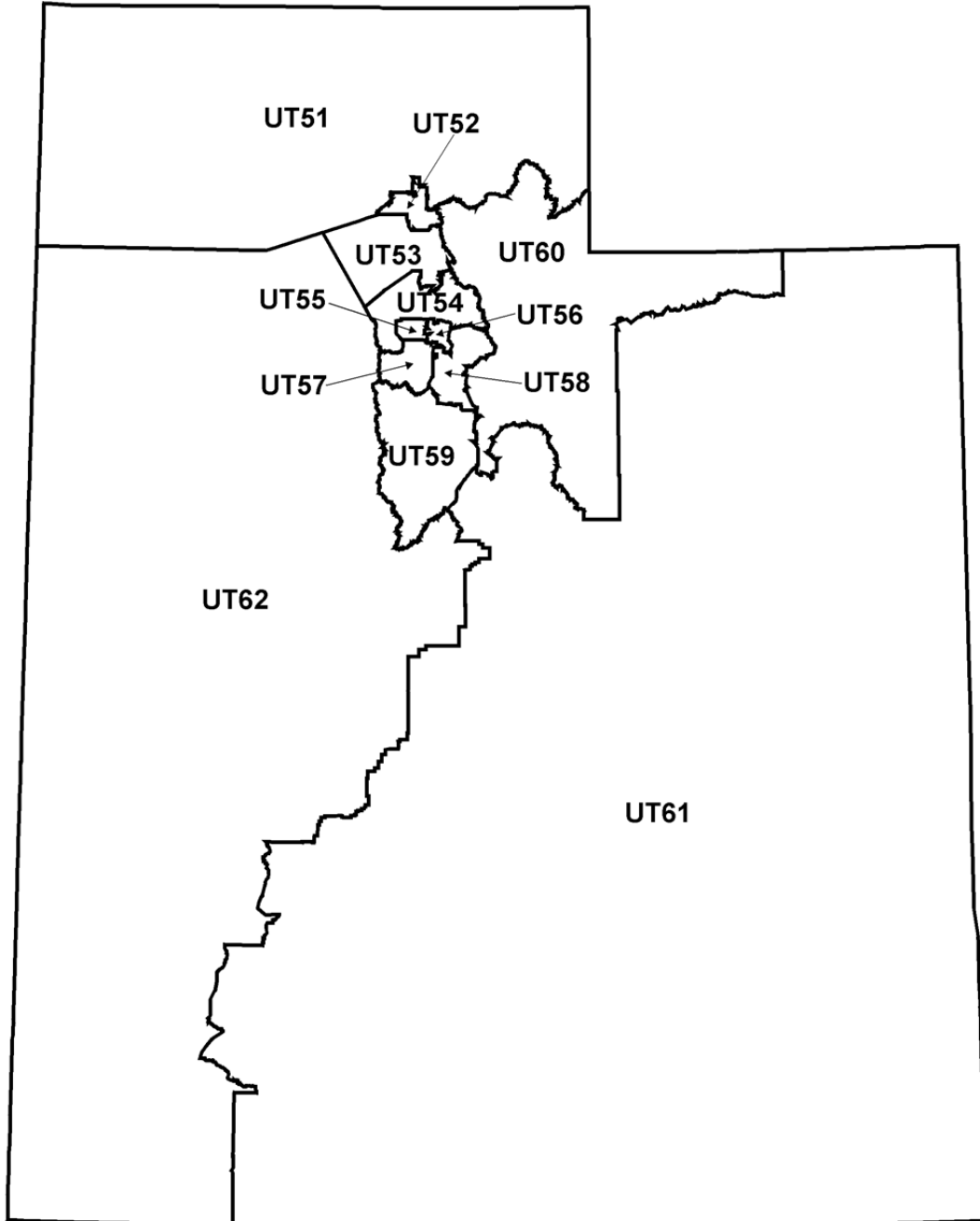
2005-2013 NSDUH State Sampling Regions: Tennessee



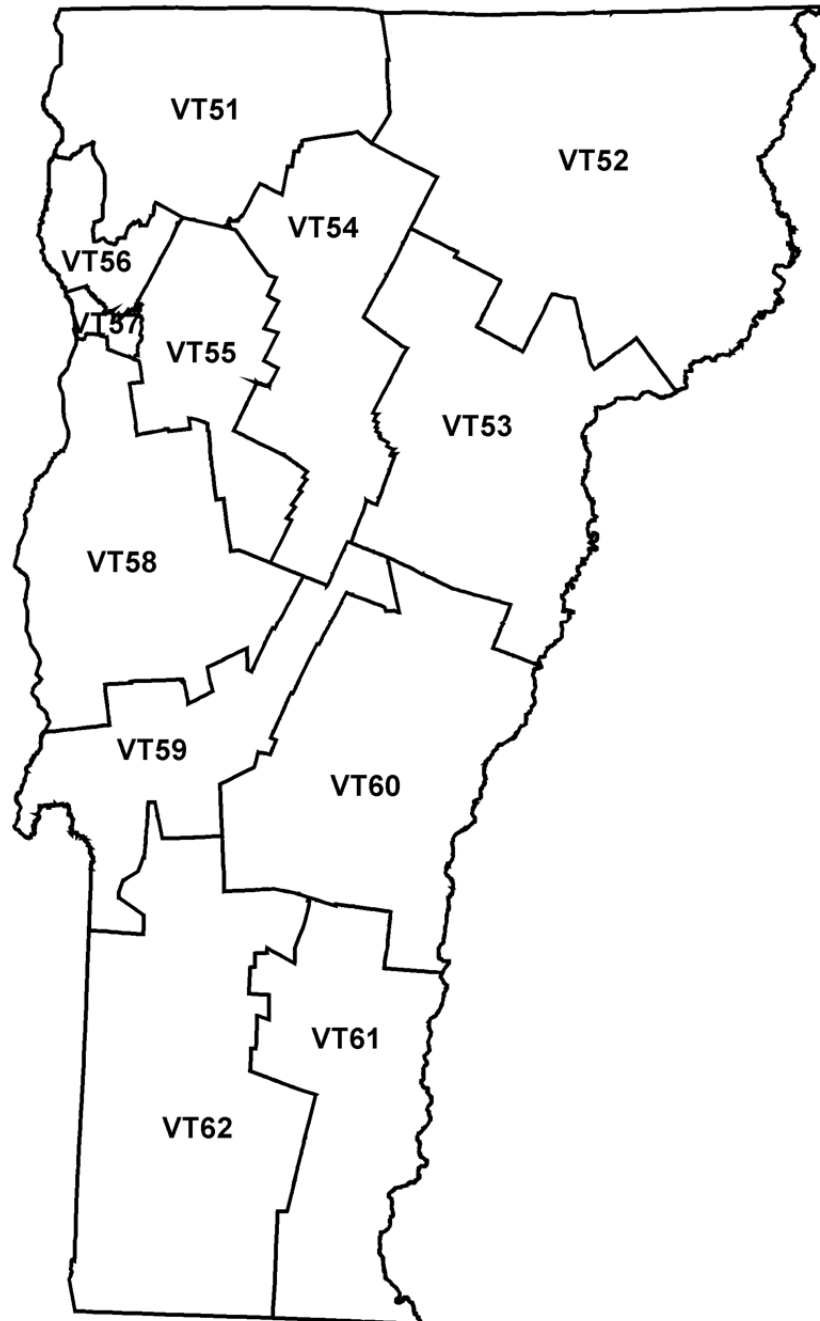
2005-2013 NSDUH State Sampling Regions: Texas



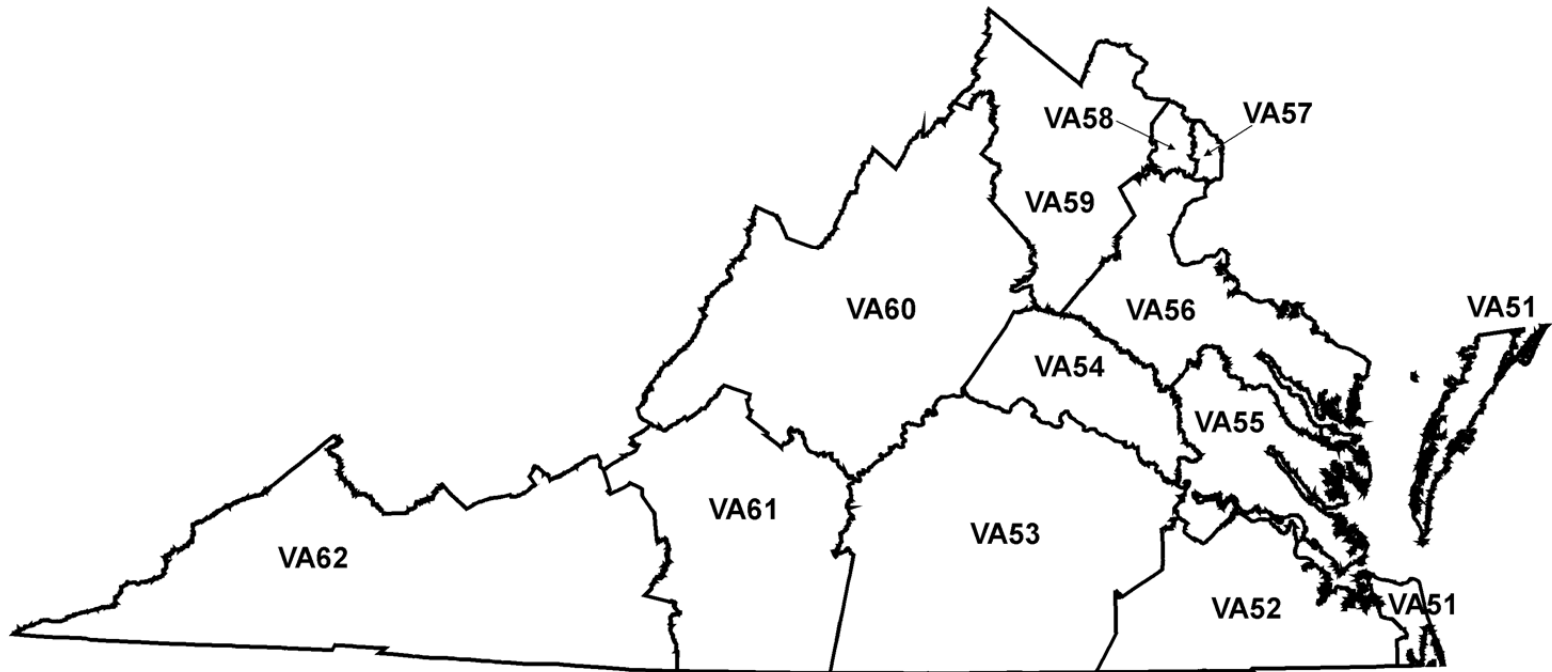
2005-2013 NSDUH State Sampling Regions: Utah



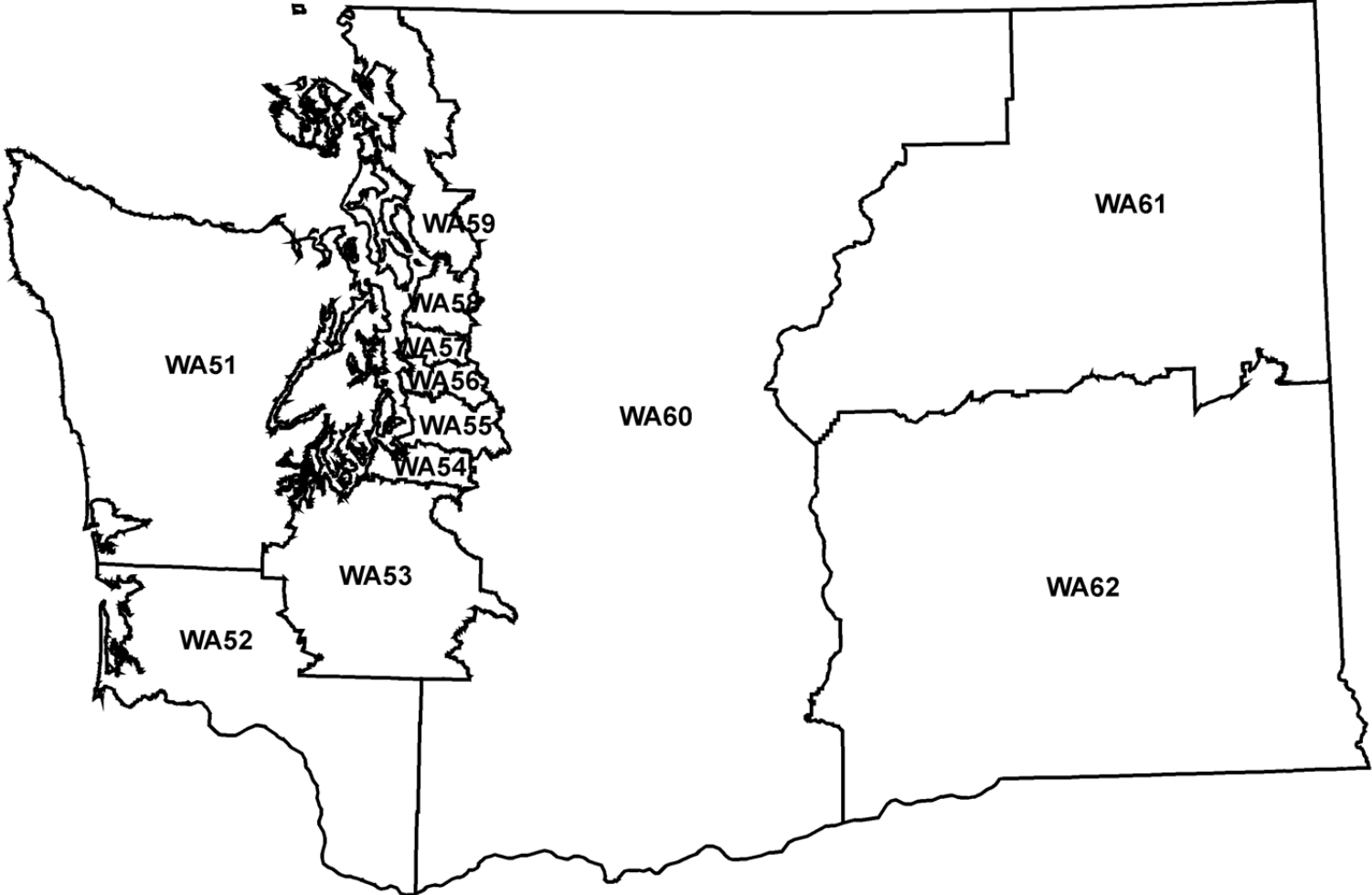
2005-2013 NSDUH State Sampling Regions: Vermont



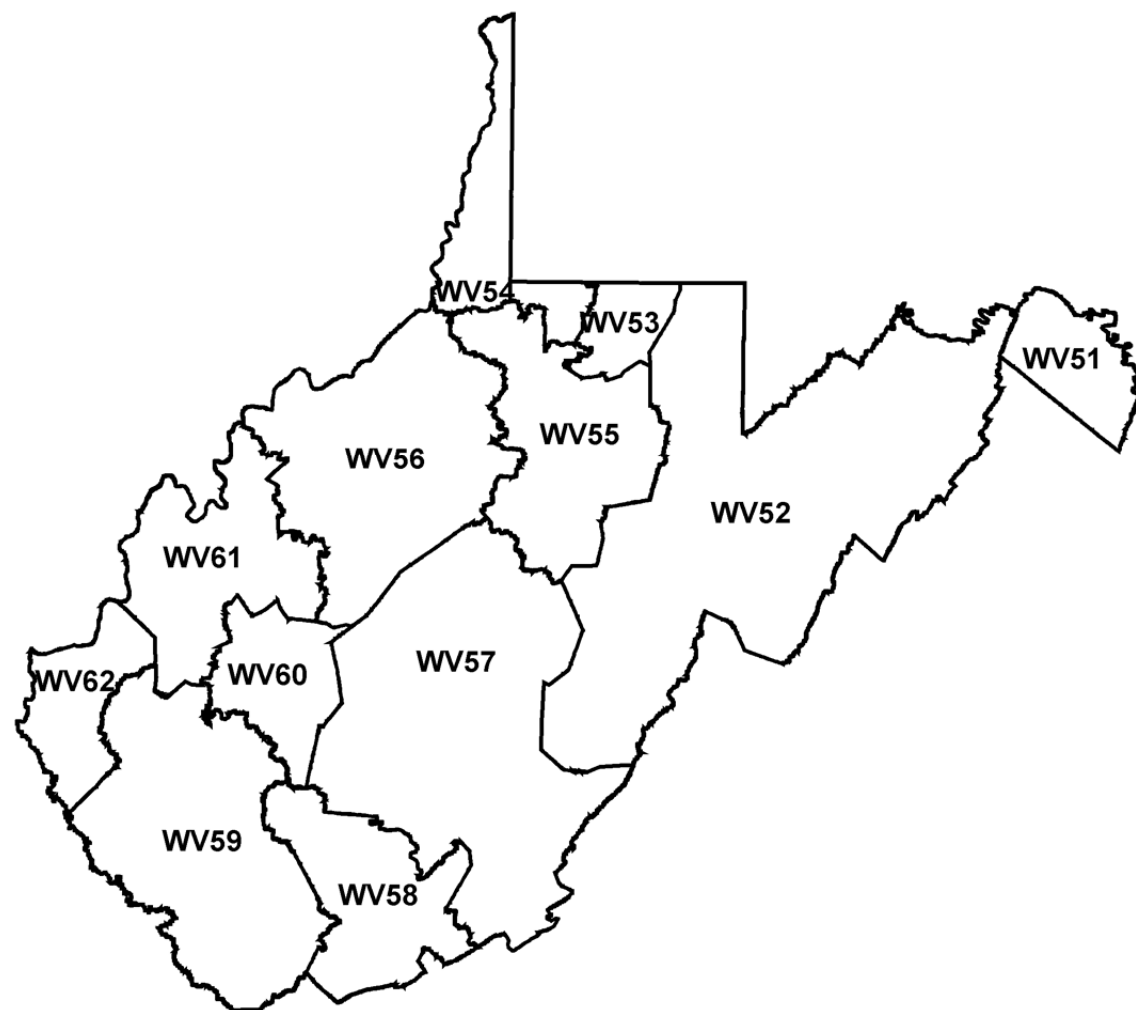
2005-2013 NSDUH State Sampling Regions: Virginia



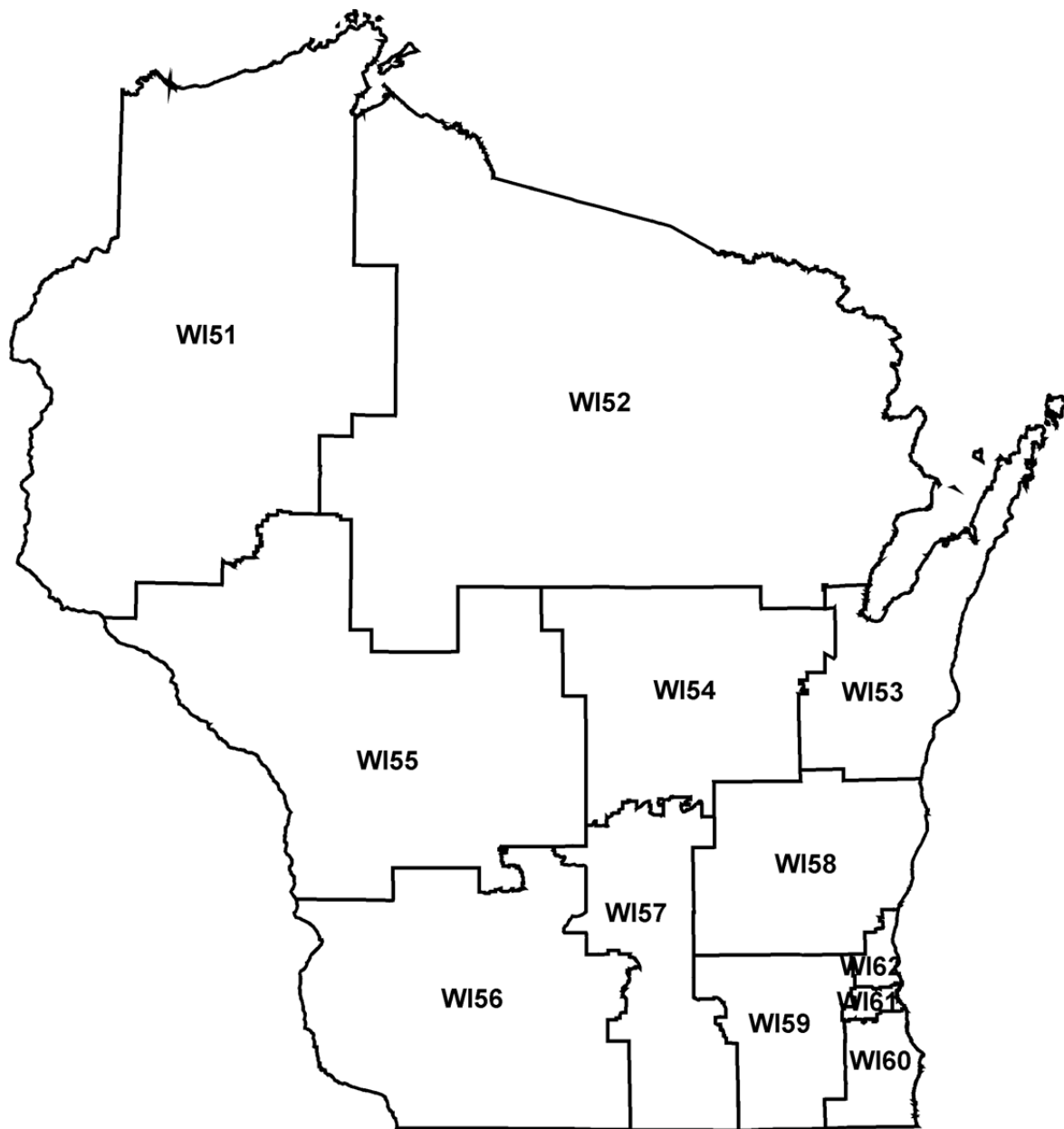
2005-2013 NSDUH State Sampling Regions: Washington



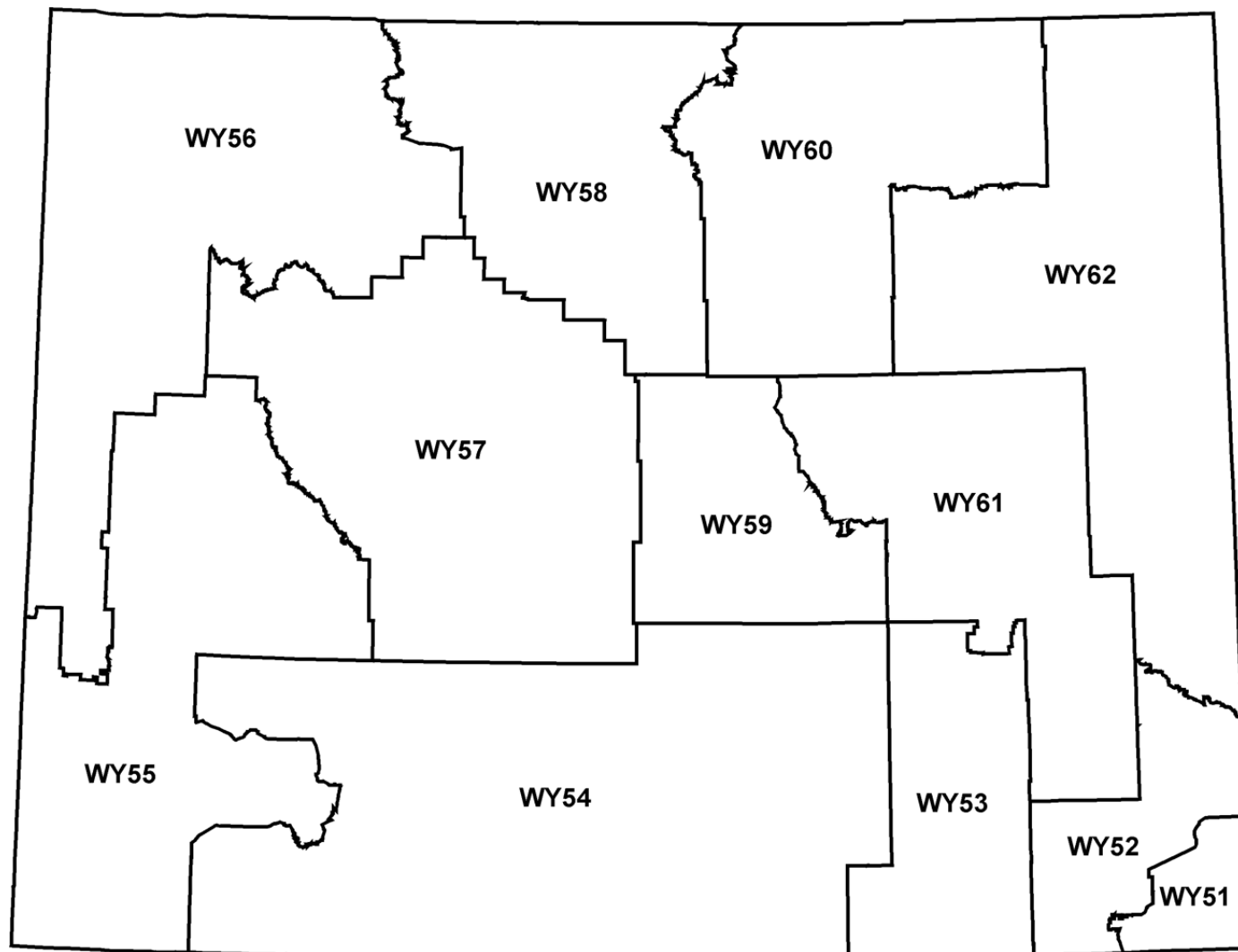
2005-2013 NSDUH State Sampling Regions: West Virginia



2005-2013 NSDUH State Sampling Regions: Wisconsin



2005-2013 NSDUH State Sampling Regions: Wyoming



Appendix B: 2012 NSDUH Procedure for Subsegmenting

B.1 Introduction

Subsegmenting is a statistical process used in the National Survey on Drug Use and Health (NSDUH) to reduce the size of the sample, which reduces the 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, and railroads. 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.

B.2 Determining Subsegmenting In-House

Prior to sending segments to the field for listing, segments that are candidates for in-house subsegmenting are identified based on the number of blocks, square miles, and dwelling units (DUs).³¹ The two criteria for identifying candidate segments are as follows:

- number of blocks > 1, square miles \geq 75, and DU count \geq 150; or
- number of blocks > 1 and DU count \geq 300.

Candidate segments then are evaluated to determine whether they can be subsegmented without input from the field. If feasible, the subsegmenting is performed in-house prior to sending the segment to the field for listing. This step expedites the process and saves time and field expenses for very large segments.

B.3 Determining Subsegmenting While in the Field

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

In the field, some of the segments that were originally subsegmented in-house (as described in Section B.2) may still be too large to list. Additional subsegmenting is required for one of the following reasons: (1) the area experienced high growth, and the census counts used in the initial subsegment were outdated, or (2) there was not enough information available during the first subsegment, and the initial subsegment was still too large to list. In the latter case, the initial subsegment was done to make the counting more manageable, but a second subsegment had to be done to make listing feasible. The initial subsegment then is counted by the lister and sent back to the sampling support office where standard subsegmenting procedures are applied.

³¹ DU counts were obtained from 2000 census data supplemented with revised population counts from Claritas.

B.4 Standard Subsegmenting Procedures

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

- Step 1: On the basis of the count, the segment is divided into areas (list units) containing no fewer than 100 DUs. 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 DUs contained in the smallest unit).
- Step 2: After properly dividing the segment into list units, the units are lettered consecutively with capital letters (A, B, C, ...) starting with the list unit and 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. On the worksheet, the number of DUs in each list unit is recorded and accumulated. A random number generated for each segment is multiplied by the total accumulated DUs. The product then is rounded up, and the list unit whose cumulative DUs is greater than or equal to the product is selected for listing.

After 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: 2012 NSDUH Procedure for Adding Missed Dwelling Units

C.1 Introduction

The 2012 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 (HUs) and group quarters (GQs), was constructed by the counting and listing staff during the summer and fall of 2011. Because the listing was done a short time before the 2012 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.

For all DUs to be given a chance of being selected, 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 (i.e., the half-open interval rule, as described in Section 3.7 of Chapter 3). 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 5 or if the number for the entire segment is less than or equal to 10, the FI is instructed to consider these DUs as part of his or her assignment. However, if either of these limits is exceeded, the FI will contact the sampling support office for subsampling to be considered.

This appendix outlines the procedures for sampling staff to use when discrepant segments are found in the field. For this appendix, procedures for adding missed DUs are classified into three categories: adding HUs, adding GQ units, and "busts."

C.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:

<u>Number of Added DUs</u>	<u>Sampling Rate</u>
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

C.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 5 (or 10) rule is exceeded, the FI should stop screening and interviewing activities on added HUs and contact the sampling support office. 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 the sampling support office, the following is to be done:
 - (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 C.6).
 - (c) If the number of added HUs linked to any *one* nonsampled DU exceeds 50, these units also will be treated using the procedure for "busts" (see Section C.6).
 - (d) Sampling staff 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. The computing division will add the DUs to the system and subsample if necessary:
 - (a) Data entry of the added DUs will be done. Entries will be made 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 then will be entered and a line number will be assigned.³² For DUs obtained through the "bust" procedure, the sampling link number (SLN) also will be recorded. Finally, it will be necessary to check that none of the DUs has already been entered in the iPAQ (i.e., the handheld computer) so that DUs do not appear in the system twice.
 - (b) DUs will be selected from the added DUs at the rate defined above. The subsampling rate will be recorded in a data field.
 - (c) Probabilities of selection will be brought over as appropriate for the segment.
 - (d) A random number will be added for the iPAQ selection algorithm.

³² During the listing process, each DU is written on a separate line on the listing form and assigned a corresponding line number (i.e., the number of lines equals the number of DUs). The added DUs are assigned the next available line number.

4. Selected DUs will be added to the FI's assignment during the next transmission.
5. A sample weight will be assigned to each added DU. If the total number of added DUs is less than or equal to 10, each added DU is assigned the weight of the original selected DUs in the segment. If subsampling is required, the selected DU weight is adjusted by the inverse of the subsampling rate for each added DU.

C.4 Procedure for Adding Group Quarters Structures

In the case of an entire 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 C.6).

C.5 Procedure for Adding Group Quarters Units

In the case of discrepant GQ listings, the number of sampling units (rooms, persons, or beds) and the number of selected units will be known in advance. If the actual number of sampling units equals the amount listed in advance, the iPAQ 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.

C.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 "Listing Error." All other units will remain eligible.

C.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

The additional units then 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:

<u>Unit Number</u>	<u>SLN</u>
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

C.6 "Busts"

A "bust" is 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. In the case of a fictitious listing, a lister 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 an 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 still will 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 nonfinalized DUs as in Section C.3.

C.7 Quality Control

To ensure quality, the sampling support office will employ several quality control checks:

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

