2013 NATIONAL SURVEY ON DRUG USE AND HEALTH

METHODOLOGICAL RESOURCE BOOK SECTION 2: SAMPLE DESIGN REPORT

Substance Abuse and Mental Health Services Administration
Center for Behavioral Health Statistics and Quality
Rockville, Maryland

February 2014

This page intentionally left blank

2013 NATIONAL SURVEY ON DRUG USE AND HEALTH: SAMPLE DESIGN REPORT

Prepared for the 2013 Methodological Resource Book (Section 2)

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.203
Phase II, Deliverable 8

RTI Authors:

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

RTI Project Director:

David Hunter

SAMHSA Project Officer:

Peter Tice

For questions about this report, please email Peter.Tice@samhsa.hhs.gov.

Prepared for Substance Abuse and Mental Health Services Administration, Rockville, Maryland

Prepared by RTI International, Research Triangle Park, North Carolina

February 2014

Recommended Citation: Center for Behavioral Health Statistics and Quality. (2014). *2013 National Survey on Drug Use and Health: Methodological Resource Book (Section 2, Sample Design Report)*. Substance Abuse and Mental Health Services Administration, Rockville, MD.

Acknowledgments

This report was prepared for the Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality, by RTI International (a trade name of Research Triangle Institute). Contributors to this report at RTI include Jeremy Aldworth, Patrick Chen, David Chrest, Phillip Kott, Richard Straw, Valerie Garner, and Teresa Bass.

Table of Contents

Chapter **Page**

Table of Contents (continued)

List of Tables

List of Exhibits

Exhibit **Page**

1. Overview

1.1 Target Population

The respondent universe for the 2013 National Survey on Drug Use and Health[1] (NSDUH) was the civilian, noninstitutionalized population aged 12 years or older residing within the 50 States and the District of Columbia. Consistent with the NSDUH designs since 1991, the 2013 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 2013 universe included those with no fixed household address (e.g., homeless and/or transient persons not in shelters), the active military, 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 that 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 target 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 estimates for 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 2013 NSDUH was 67,838 persons.

Beginning with the 2002 NSDUH and continuing through the 2013 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 2013. 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.

Finally, a Dress Rehearsal (DR) of the 2015 NSDUH redesigned questionnaire protocol was conducted in late 2013. Chapter 4 describes the sample design for the 2013 DR.

[1] This report presents information from the 2013 National Survey on Drug Use and Health (NSDUH). Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

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 2013 NSDUH main sample continues the 50 percent overlap by retaining half of the second-stage units from the 2012 survey. The 2013 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,[2] had samples designed to yield 3,600 respondents per State for the 2013 main study. This sample size was considered adequate to support direct State estimates. The remaining 43 States[3] had samples designed to yield 900 respondents per State in the 2013 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 regions (SSRs) 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 SSRs, whereas the 8 large States were divided into 48 SSRs. Therefore, the partitioning of the United States resulted in the formation of a total of 900 SSRs. Maps for these regions 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 2013 NSDUHs was census tracts.[4] This stage was included to contain sample segments within a single census tract to the extent possible.[5] 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

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

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

[4] 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).

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

aggregated within SSRs until each tract[6] had, at a minimum, 150 DUs[7] in urban areas and 100 DUs in rural areas.[8]

Before selecting census tracts, additional implicit stratification was achieved by sorting the first-stage sampling units by a CBSA/SES[9] (core-based statistical area/socioeconomic status) indicator[10] and by the percentage of the population that is non-Hispanic and white.[11] From this well-ordered sample frame, 48 census tracts per SSR 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 clusters[12] of DUs by aggregating adjacent census blocks.[13] Consistent with the terminology used in previous NSDUHs, these geographic clusters of blocks are referred to as "segments." A sample DU in NSDUH refers to either a housing unit or a group quarters listing unit, such as a dormitory room or a shelter bed. Similar to census tracts, segments were formed to contain a minimum of 150 DUs in urban areas and 100 DUs in rural areas. This minimum DU requirement will support the overlapping sample design and any special supplemental samples or field tests that SAMHSA may wish to conduct.

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

---

[6] 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.

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

[8] 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.

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

[10] 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."

[11] 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.

[12] 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).

[13] 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).

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.

An equal probability subsample of eight segments is used for each NSDUH year. These eight segments are randomly assigned to quarters and to two panels within each quarter. For each survey year after 2005, the first panel segments constitute the overlap sample; that is, they were used in the prior year and were used for the second time in the "current" survey year. The second panel segments constitute the replacement sample and are used for the "current" survey and again the following year, except in 2013 when the second panel segments were used for the 2013 survey only.

1.5 Sample Dwelling Units and Persons

After sample segments for the 2013 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,[14] 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.[15] In addition to these factors, historical data from the 2011, 2012, and 2013 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. DUs that were selected from the first panel

[14] 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.

[15] 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.

(overlap) segments in the prior year were not eligible for selection in the "current" year (i.e., two separate samples were selected with the complement of the prior year's sample serving as the DU frame in the "current" year). In 2005, when there was no overlap with the prior year's sample, the same DUs could have been selected over a 2-year period by chance. Persons may be selected in consecutive years if they move and their new residence is selected the year after their original DU was sampled. No mechanism is currently in place for identifying duplicate persons in a given year, but this number should be small given the restriction on DUs that were sampled in the previous year.

After DU selections were made, an interviewer visited each selected DU to obtain a roster of all persons residing in the DU. As in previous years, during the data collection period, if an interviewer encountered any new DU in a segment or found a DU that was missed during the original counting and listing activities, the new or missed dwellings were selected into the 2013 NSDUH using the half-open interval (HOI) selection technique.[16] 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 2012 designs, one feature that was included in the 2013 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).

[16] In summary, the HOI technique states that, if a DU is selected for the 2013 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 Section 3.7 in Chapter 3 and Appendix D.

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-weighting[17] 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

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

- 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.

[17] 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,...,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 region (SSR) i within each State h. The composite size measure for block k was defined as

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

The composite size measure for segment j was calculated as

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

where N_{hij} equals the number of blocks within segment j of SSR 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 SSRs. Within each State, census tracts were joined to form mutually exclusive and exhaustive SSRs of approximately equal sizes. Prior to forming the SSRs, 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 SSRs 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 2), 12 SSRs were formed; however, 48 SSRs 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 SSRs 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 (SSRs) 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 (SSRs) 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 SSRs 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 SSR. 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.[18] 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 SSR 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 SSR 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 were available for use in 2010-2011 and 2012-2013 NSDUHs.

[18] 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 SSR 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 2013 main survey corresponds to segment suffixes 33 through 40.

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 (SEGID).

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 SSR level, and each SSR 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 SSRs 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 is ≤ 150). Similarly, VESTRQ3 = VESTRQ2 – 150 (+ 900 if VESTRQ2 ≤ 150), and VESTRQ4 = VESTRQ3 – 150 (+ 900 if VESTRQ3 ≤ 150). As an example, an SSR 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 SSRs from four different States.

The 2005 through 2013 definition of variance estimation strata has the effect of increasing the number of degrees of freedom (*df*) 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 SSRs × 4 quarters); therefore, there are 48 degrees of freedom available for State estimates. Similarly, each large sample State is in 192 strata (48 SSRs × 4 quarters) and therefore has 192 degrees of freedom for estimation. As demonstrated in Appendix B, the 2005 through 2013 definition of variance estimation strata achieves variance estimators with the same expected values as those formed by grouping segments across quarters within regions (i.e., the 1999 through 2004 definition of variance estimation strata).

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, 2012b). 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.[19] 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:

[19] 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.

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 group quarter.[20] 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.[21] 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.

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

[21] 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 2013 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.[22] 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.

[22] 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,...,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 2013 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 2013 population (y^*_{ha}) is computed using the original formula and this time allowing x to be 13. Finally, the 2013 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 (HOI) rule (1.01) and the probability of a DU being eligible (ε_h), both determined via historical data. The coefficient adjustment of 1.01 is estimated using historical data and is the proportion of all screened DUs (includes added DUs) over the original total of selected DUs (excluding added DUs). So, $y_{ha} = \{[1.01 * \varepsilon_h * L_{hj} * (1/I_{hj}) / U_{hj}]\} * y^*_{ha}$, where ε_h, L_{hj}, and I_{hj} are defined further below. This adjustment is computed at the census block level and then aggregated to the State level.

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

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

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

I_{hj} = Subsegmentation inflation factor. For segments too large to count and to list efficiently in both time and cost, field listing personnel 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 2011 NSDUH quarters 3 and 4 and 2012 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 2011 NSDUH and quarters 1 and 2 of the 2012 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 2011 NSDUH quarters 3 and 4 and 2012 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 2011 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 2013 NSDUH design, are presented in Table 3.1, which was updated using 2011 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 2012 surveys, racial groups were not oversampled for the 2013 NSDUH. Consistent with previous surveys, the 2013 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	22.06	1.54	2.45	7.76	3.33	2.90
Past Month Alcohol Use	51.78	0.84	2.60	13.32	2.49	2.92
Past Month Use of Illicit Drugs	8.72	2.18	2.02	10.10	2.82	2.83
Past Month Use of Illicit Drugs Other Than Marijuana	3.11	3.53	1.90	4.11	4.45	2.76
Past Month Cocaine Use	0.53	8.81	1.93	0.27	17.79	2.77
Past Year, Dependent on Illicit Drugs	1.77	4.36	1.76	2.53	5.75	2.78
Past Year, Dependent on Alcohol but Not Illicit Drugs	2.56	4.13	2.01	0.97	9.11	2.71
Past Year, Received Treatment for Illicit Drug Use	0.78	7.50	2.00	0.91	9.72	2.80
Past Year, Received Treatment for Alcohol Use but Not for Illicit Drug Use	0.56	9.79	2.20	0.12	25.88	2.71
Average RSE	N/A	N/A	2.10	N/A	N/A	2.80
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	33.49	1.39	2.96	21.88	1.79	2.84
Past Month Alcohol Use	60.68	0.80	3.00	55.08	0.90	2.98
Past Month Use of Illicit Drugs	21.43	1.84	2.89	6.34	3.42	2.67
Past Month Use of Illicit Drugs Other Than Marijuana	6.95	3.42	2.81	2.32	5.61	2.60
Past Month Cocaine Use	1.35	8.01	2.81	0.42	13.19	2.58
Past Year, Dependent on Illicit Drugs	5.30	3.98	2.83	1.07	8.18	2.55
Past Year, Dependent on Alcohol but Not Illicit Drugs	4.68	4.13	2.74	2.39	5.51	2.59
Past Year, Received Treatment for Illicit Drug Use	1.69	7.25	2.85	0.61	11.26	2.64
Past Year, Received Treatment for Alcohol Use but Not for Illicit Drug Use	0.64	11.46	2.75	0.60	11.54	2.68
Average RSE	N/A	N/A	2.85	N/A	N/A	2.68
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	SSRs	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	SSRs	Total Segments	Total Respondents					
				12-17	18-25	26-34	35-49	50+	Total
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; SSR = State sampling region.

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 2013 NSDUH sample is summarized in Table 3.4. For more information on the subsegmenting procedures, see Appendix C.

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	42,892	6.0
Alabama	96	617	6.4
Alaska	96	641	6.7
Arizona	96	550	5.7
Arkansas	96	625	6.5
California	384	1,396	3.6
Colorado	96	480	5.0
Connecticut	96	465	4.8
Delaware	96	419	4.4
District of Columbia	96	305	3.2
Florida	384	2,004	5.2
Georgia	96	584	6.1
Hawaii	96	396	4.1
Idaho	96	818	8.5
Illinois	384	2,340	6.1
Indiana	96	546	5.7
Iowa	96	742	7.7
Kansas	96	838	8.7
Kentucky	96	537	5.6
Louisiana	96	615	6.4
Maine	96	599	6.2
Maryland	96	379	3.9
Massachusetts	96	470	4.9
Michigan	384	2,029	5.3
Minnesota	96	604	6.3
Mississippi	96	734	7.6
Missouri	96	619	6.4
Montana	96	1,082	11.3
Nebraska	96	910	9.5
Nevada	96	452	4.7
New Hampshire	96	491	5.1
New Jersey	96	422	4.4
New Mexico	96	950	9.9
New York	384	1,510	3.9
North Carolina	96	529	5.5
North Dakota	96	1,063	11.1
Ohio	384	2,143	5.6
Oklahoma	96	738	7.7
Oregon	96	587	6.1
Pennsylvania	384	2,249	5.9
Rhode Island	96	490	5.1
South Carolina	96	615	6.4
South Dakota	96	984	10.3
Tennessee	96	519	5.4
Texas	384	2,136	5.6

(continued)

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

State	Total Segments	Cumulative Number of Map Pages per State	Average Number of Map Pages per Segment
Utah	96	501	5.2
Vermont	96	616	6.4
Virginia	96	467	4.9
Washington	96	559	5.8
West Virginia	96	664	6.9
Wisconsin	96	577	6.0
Wyoming	96	1,286	13.4

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	720	4	1,562,599	225,176	1,907
Alabama	96	11	0	21,366	3,105	5
Alaska	96	21	0	21,600	3,119	58
Arizona	96	10	0	20,290	3,009	4
Arkansas	96	6	0	18,607	2,714	7
California	384	34	1	82,686	9,951	43
Colorado	96	12	0	21,432	2,776	14
Connecticut	96	8	0	21,707	2,958	31
Delaware	96	7	0	22,433	3,030	12
District of Columbia	96	14	1	28,475	5,398	68
Florida	384	68	0	90,256	14,125	49
Georgia	96	9	0	21,181	2,652	8
Hawaii	96	14	0	25,726	3,266	28
Idaho	96	13	0	19,232	2,381	7
Illinois	384	18	0	78,098	11,695	72
Indiana	96	8	0	18,227	2,962	30
Iowa	96	6	0	19,522	2,667	33
Kansas	96	5	0	18,452	2,583	25
Kentucky	96	8	0	20,478	3,053	32
Louisiana	96	10	0	21,008	2,864	13
Maine	96	7	0	21,880	3,565	59
Maryland	96	9	0	25,429	2,740	19
Massachusetts	96	9	0	20,991	2,981	26
Michigan	384	30	1	81,297	11,966	114
Minnesota	96	7	0	19,339	2,582	13
Mississippi	96	5	0	18,450	2,430	11
Missouri	96	10	0	17,917	3,093	51
Montana	96	18	0	17,181	2,968	23
Nebraska	96	10	0	18,332	3,045	7

(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	20	0	23,367	2,748	5
New Hampshire	96	8	0	20,672	3,407	81
New Jersey	96	5	0	22,417	3,153	11
New Mexico	96	12	1	18,511	2,864	4
New York	384	53	0	93,064	15,001	164
North Carolina	96	10	0	22,221	2,826	46
North Dakota	96	12	0	17,847	3,595	39
Ohio	384	33	0	81,473	11,486	54
Oklahoma	96	10	0	20,315	2,804	26
Oregon	96	5	0	20,248	2,734	36
Pennsylvania	384	25	0	82,160	13,168	124
Rhode Island	96	4	0	19,367	2,935	34
South Carolina	96	5	0	21,167	3,253	38
South Dakota	96	12	0	18,390	2,693	35
Tennessee	96	17	0	20,215	2,938	29
Texas	384	40	0	85,442	9,288	35
Utah	96	6	0	21,334	2,025	7
Vermont	96	13	0	20,335	3,540	82
Virginia	96	14	0	22,342	2,740	52
Washington	96	11	0	22,003	2,566	32
West Virginia	96	4	0	20,135	3,487	39
Wisconsin	96	4	0	17,882	2,829	36
Wyoming	96	10	0	20,100	3,418	36

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,[23] 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 HOI procedure for recovering missed DUs, as described in Section 3.7. A detailed description of the counting and listing procedures is provided in the 2013 counting and listing general manual (RTI International, 2012a).

[23] 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:

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

At this point in the procedure, only two components in the formula are unknown: D_{hj} and S_{hja}. Selection probabilities are segment- and age group-specific, and to maximize the number of selected persons within a DU, the age group whose adjusted sampling fraction [f_{ha} / (ϕ_h * λ_{ha} * δ_{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

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

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

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

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

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

1. Number of selected DUs for screening: < 100 or < ½ L_{hj}. Adjustments were made by adjusting the D_{hj} counts to equal the minimum of 100 or ½ L_{hj}.
2. Number of selected DUs: > 5. For cost purposes, if at least five DUs remain in the segment, the minimum number of selected DUs was set to five.
3. Expected number of interviews: < 40.

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

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

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

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

Similarly, the interview adjustment is computed as

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

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

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

3.6 Dwelling Unit Selection and Release Partitioning

After derivation of the required DU sample size 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 Procedures for Adding Dwelling Units

To guarantee that every DU had a chance of selection and to eliminate any bias associated with incomplete frames, two procedures for adding missed DUs were implemented. First, during the screening interview, the interviewer asked the screening respondent about other units on the property of the sampled DU (e.g., a garage apartment). Next, the interviewer implemented the HOI rule. This procedure required that the interviewer look between each selected DU and the next listed DU for any unlisted units. When found on the property of a sampled DU or in the geographic interval following the DU (HOI), the unlisted units became part of the sample (added DUs) and were considered "linked" to that DU. 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 D. The total number of added DUs identified during the screening interview or through implementing the HOI rule is summarized in the last column of Table 3.4. An evaluation of 2010 NSDUH data found that the HOI procedure accounted for only 0.2 percent of the total DUs on the NSDUH frame (Iannacchione, McMichael, Shook-Sa, & Morton, 2012).

Table 3.5 Quarterly Sample Sizes and Percentages Released

Region/State	Quarter 1			Quarter 2		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
Total Population	61,181	51,360	84	67,414	60,202	89
Northeast						
Connecticut	836	700	84	922	760	82
Maine	1,033	864	84	1,204	852	71
Massachusetts	789	722	92	935	777	83
New Hampshire	816	751	92	916	916	100
New Jersey	834	693	83	865	825	95
New York	4,289	3,573	83	4,568	3,807	83
Pennsylvania	3,465	2,897	84	3,873	3,544	92
Rhode Island	863	722	84	909	717	79
Vermont	971	806	83	1,093	1,045	96
Midwest						
Illinois	3,179	2,649	83	3,405	2,979	87
Indiana	764	635	83	877	877	100
Iowa	712	650	91	791	726	92
Kansas	692	578	84	709	623	88
Michigan	3,249	2,711	83	3,634	3,634	100
Minnesota	706	582	82	766	673	88
Missouri	767	641	84	859	859	100
Nebraska	761	636	84	846	846	100
North Dakota	897	821	92	1,022	1,022	100
Ohio	3,094	2,561	83	3,438	2,999	87
South Dakota	773	648	84	932	775	83
Wisconsin	708	586	83	812	742	91

(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	884	739	84	977	775	79
Arkansas	745	619	83	818	818	100
Delaware	788	659	84	890	706	79
District of Columbia	1,489	1,239	83	1,444	1,444	100
Florida	3,687	3,077	83	4,209	3,682	87
Georgia	694	581	84	791	791	100
Kentucky	806	673	83	827	827	100
Louisiana	824	688	83	863	828	96
Maryland	779	646	83	859	825	96
Mississippi	713	593	83	775	677	87
North Carolina	790	724	92	900	863	96
Oklahoma	846	704	83	945	707	75
South Carolina	918	769	84	983	862	88
Tennessee	768	638	83	878	766	87
Texas	2,496	2,072	83	2,660	2,329	88
Virginia	692	578	84	742	742	100
West Virginia	891	738	83	982	938	96
West						
Alaska	784	655	84	993	827	83
Arizona	735	607	83	876	838	96
California	2,790	2,323	83	3,163	2,626	83
Colorado	774	642	83	873	728	83
Hawaii	898	748	83	1,065	887	83
Idaho	652	543	83	680	601	88
Montana	900	750	83	981	815	83
Nevada	824	750	91	752	627	83
New Mexico	813	677	83	919	733	80
Oregon	725	605	83	815	746	92
Utah	521	439	84	583	583	100
Washington	792	657	83	786	686	87
Wyoming	965	801	83	1,009	927	92

(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	67,104	58,922	88	63,914	54,692	86
Northeast						
Connecticut	806	738	92	760	760	100
Maine	1,135	995	88	1,134	854	75
Massachusetts	788	687	87	873	795	91
New Hampshire	937	824	88	1,094	916	84
New Jersey	843	809	96	860	826	96
New York	4,729	4,321	91	4,184	3,300	79
Pennsylvania	3,792	3,466	91	3,898	3,261	84
Rhode Island	908	871	96	784	625	80
Vermont	1,040	825	79	1,043	864	83
Midwest						
Illinois	3,526	3,092	88	3,244	2,975	92
Indiana	824	824	100	830	626	75
Iowa	787	720	91	763	571	75
Kansas	724	724	100	658	658	100
Michigan	3,408	2,978	87	3,339	2,643	79
Minnesota	852	783	92	724	544	75
Missouri	828	828	100	916	765	84
Nebraska	841	841	100	818	722	88
North Dakota	1,227	1,075	88	1,009	677	67
Ohio	3,336	3,060	92	3,128	2,866	92
South Dakota	887	626	71	771	644	84
Wisconsin	877	766	87	842	735	87

(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	972	849	87	844	742	88
Arkansas	798	702	88	769	575	75
Delaware	894	857	96	808	808	100
District of Columbia	1,390	1,390	100	1,515	1,325	87
Florida	4,246	3,709	87	4,383	3,657	83
Georgia	769	639	83	699	641	92
Kentucky	872	797	91	856	756	88
Louisiana	856	714	83	800	634	79
Maryland	857	745	87	785	524	67
Mississippi	654	541	83	705	619	88
North Carolina	932	622	67	780	617	79
Oklahoma	952	719	76	900	674	75
South Carolina	968	851	88	839	771	92
Tennessee	931	776	83	827	758	92
Texas	2,698	2,469	92	2,636	2,418	92
Virginia	775	775	100	706	645	91
West Virginia	990	868	88	1,032	943	91
West						
Alaska	884	738	83	899	899	100
Arizona	932	852	91	815	712	87
California	3,079	2,314	75	2,812	2,688	96
Colorado	870	724	83	779	682	88
Hawaii	977	817	84	1,086	814	75
Idaho	656	656	100	633	581	92
Montana	852	679	80	863	724	84
Nevada	859	750	87	677	621	92
New Mexico	912	723	79	882	731	83
Oregon	941	828	88	701	555	79
Utah	597	496	83	526	507	96
Washington	864	648	75	690	575	83
Wyoming	1,032	821	80	995	869	87

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 2012 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 2012 NSDUH and quarter 1 of the 2013 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 2012 NSDUH and quarters 1 and 2 of the 2013 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 2013 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 HOI 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 2010 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 2010 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 2013 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 report on the person-level sampling weight calibration for the 2013 NSDUH (Chen et al., 2014).

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

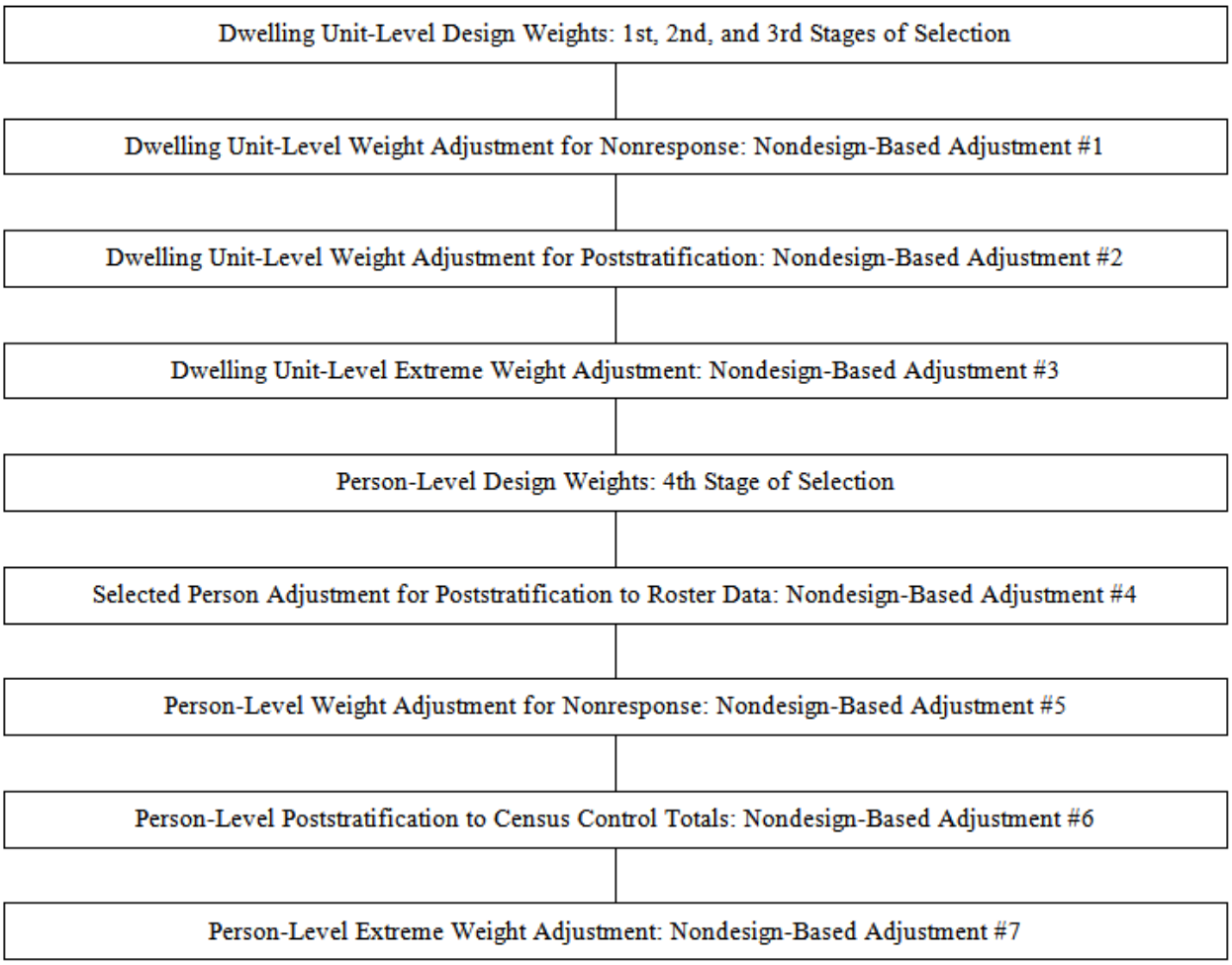
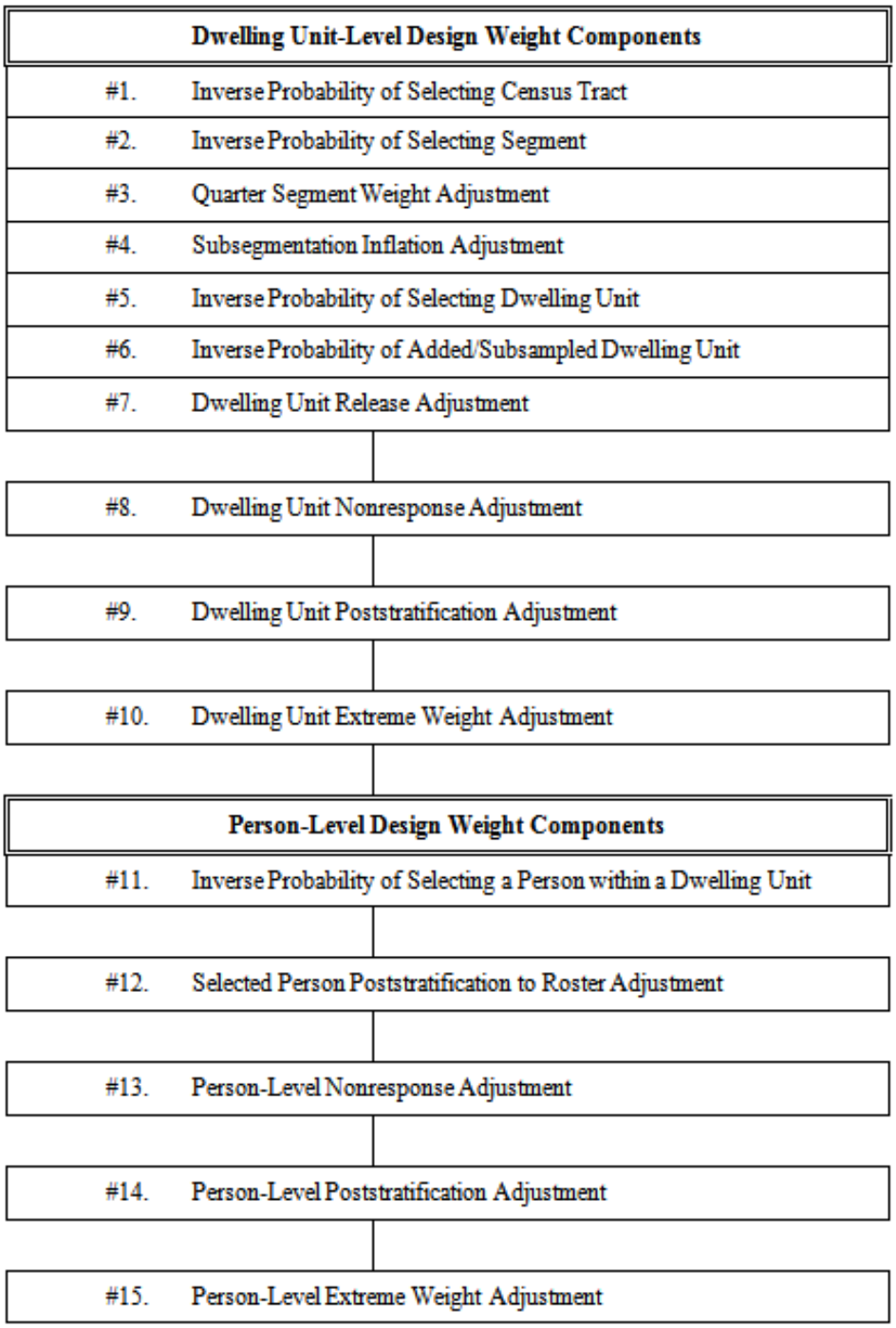


Exhibit 2 Sample Weight Components



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 that 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 that 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 Procedures for the Dress Rehearsal

A Questionnaire Field Test (QFT) was conducted in 2012 that tested revisions to the NSDUH respondent materials, questionnaire, procedures, and equipment associated with the 2015 partial redesign goals (Currivan et al., 2013). Following the QFT, a Dress Rehearsal (DR) was conducted during the 2013 National Survey on Drug Use and Health (NSDUH), which aimed to further test this same set of revisions made to the QFT materials, questionnaire, and procedures, as well as some further revisions made specifically for the DR (Currivan et al., 2014). Two major differences between the QFT and the DR were the addition of Spanish-language interviews and a test of new lightweight laptop computers. The DR provided another opportunity to further refine and improve the redesigned questionnaire, materials, and procedures prior to any full-scale changes for the 2015 partial redesign. It also allowed for an evaluation of the Spanish-language version of the redesigned questionnaire. The DR overlapped with quarters 3 and 4 of the main study, with approximately 2,000 persons interviewed between September 1 and October 31, 2013. Data from the DR were compared with data from the 2012 NSDUH, quarters 3 and 4 of the 2013 NSDUH, and the QFT. This chapter provides the sampling procedures for the DR.

4.1 Target Population

Similar to the main study of the NSDUH, the respondent universe for the DR was the civilian, noninstitutionalized population aged 12 or older. In order to control costs, persons residing in Alaska and Hawaii were excluded from the DR. Therefore, the sample is representative of the noninstitutionalized population aged 12 or older in the contiguous United States.

4.2 Spanish-Language Interview Oversample

One primary goal of the DR was to evaluate the Spanish-language questionnaire, so it was critical to complete enough DR interviews in Spanish to allow for this evaluation. To achieve a higher yield of Spanish-language interviews than what would be observed with a probability proportional to size (PPS) sample, a special certainty stratum was created that comprised the State sampling regions (SSRs) with an historically high percentage of interviews conducted in Spanish. SSRs that had 10 percent or more of their 2011 NSDUH interviews conducted in Spanish were assigned to the certainty stratum. The percentage of interviews conducted in Spanish was calculated at the SSR level rather than the segment level because sample sizes at the segment level were too small to provide reliable estimates. A total of 101 of the NSDUH SSRs fell into the certainty stratum and were selected for the DR with certainty.

Because of the oversampling of areas with historically high concentrations of Spanish-language interviews, 207 of the 2,000 total interviews were expected to be completed in Spanish. Table 4.1 presents the expected number of interviews and estimated precision of survey estimates for the total interviews and for the Spanish-language interviews. Although this oversampling approach led to a higher yield of Spanish-language interviews compared with a design where all of the segments were selected PPS, it decreased the precision of the overall estimates by increasing

the design effects. Areas with high concentrations of Spanish-language interviews had a much higher probability of selection under this design than they would have had under a PPS design. This design balanced the goals of testing the Spanish-language questionnaire and producing efficient overall estimates.

Table 4.1 Expected Number of Interviews and Precision of Dress Rehearsal Estimates

	Total Interviews	Spanish-Language Interviews
Expected Number of Interviews	2,000	207
Standard Errors (SEs) of Estimates[1]	1.30%	4.08%
Relative Standard Errors (RSEs) of Estimates[1]	12.98%	40.83%

[1] SE and RSE calculations assume a design effect of 2.5 and a prevalence of $p = 0.10$.

4.3 Selection of State Sampling Regions and Segments

Because NSDUH is designed to yield 67,500 interviews from 7,200 segments each calendar year, an estimated 200 segments were needed to yield approximately 2,000 completed DR interviews. As discussed in Section 4.2, a special certainty stratum was developed to ensure that a sufficient number of DR interviews would be completed in Spanish. As mentioned in Section 4.2, 101 of the NSDUH SSRs fell into the certainty stratum and were selected with certainty. To ensure national representation, the remaining 775 SSRs were stratified by census region, and 99 SSRs were selected PPS for inclusion in the DR. Implicit stratification was achieved by sorting the frame of SSRs by the percentage urban and the percentage of interviews completed in Spanish in 2011 prior to selecting the sample.

This design had the benefit of placing much of the sample in heavily populated areas where a sufficient mix of field interviewers (FIs) with various experience levels were available to meet the DR staffing needs. As shown in Table 4.2, a large portion of the sample was selected from the eight largest States (i.e., California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas). In addition, the majority of the Spanish-language interviews were expected to be completed in States where bilingual FIs were already employed.

Within each selected SSR, a sample of DUs was drawn from the segment that was retired from use in quarter 1 of the 2013 NSDUH. DUs that were not selected for the main study in 2012 and 2013 were eligible for selection in the field test. If an insufficient number of DUs remained in a segment, or if significant access problems were expected, the segment was replaced with the quarter 3 or quarter 4 2012 retired segment in the same SSR. One segment was replaced because it had fewer than 10 DUs remaining, and 10 segments were replaced because of anticipated access problems in the segments.

Table 4.2 Number of 2013 Dress Rehearsal State Sampling Regions and Achieved Sample Sizes, by State

State	Population Rank (12 or Older)	Current Design	NSDUH SSR Regions	Number of DR SSR Regions/ Segments	2013 DR Total Interviews	2013 DR Spanish-Language Interviews
CA	1	3,600	48	38	640	60
TX	2	3,600	48	23	265	59
NY	3	3,600	48	15	142	11
FL	4	3,600	48	13	121	22
IL	5	3,600	48	12	96	8
PA	6	3,600	48	4	41	1
OH	7	3,600	48	5	49	2
MI	8	3,600	48	3	29	0
GA	9	900	12	4	23	0
NC	10	900	12	4	44	0
NJ	11	900	12	3	19	0
VA	12	900	12	4	28	0
MA	13	900	12	4	42	0
WA	14	900	12	1	5	0
IN	15	900	12	2	19	0
AZ	16	900	12	7	58	7
TN	17	900	12	3	22	0
MO	18	900	12	2	22	0
WI	19	900	12	5	51	1
MD	20	900	12	2	22	3
MN	21	900	12	2	20	0
CO	22	900	12	3	10	0
AL	23	900	12	3	32	0
SC	24	900	12	1	9	0
KY	25	900	12	1	8	0
LA	26	900	12	2	17	0
OR	27	900	12	1	12	0
OK	28	900	12	1	8	0
CT	29	900	12	1	2	0
IA	30	900	12	1	13	0
MS	31	900	12	0	0	0
AR	32	900	12	1	11	0
KS	33	900	12	2	19	0
NV	34	900	12	5	31	4
UT	35	900	12	2	11	0
NM	36	900	12	4	16	1

(continued)

Table 4.2 Number of 2013 Dress Rehearsal State Sampling Regions and Sample Sizes, by State (continued)

State	Population Rank (12 or Older)	Current Design	NSDUH SSR Regions	Number of DR SSR Regions/ Segments	2013 DR Total Respondents	2013 DR Spanish-Language Respondents
WV	37	900	12	3	35	0
NE	38	900	12	1	7	2
ID	39	900	12	1	1	0
ME	40	900	12	3	34	0
NH	41	900	12	0	0	0
HI	42	900	12	0	0	0
RI	43	900	12	3	7	0
MT	44	900	12	0	0	0
DE	45	900	12	1	4	2
SD	46	900	12	1	15	0
AK	47	900	12	0	0	0
VT	48	900	12	1	14	0
ND	49	900	12	0	0	0
DC	50	900	12	1	4	2
WY	51	900	12	1	9	0
	Total	67,500	900	200	2,087	185

DR = Dress Rehearsal; NSDUH = National Survey on Drug Use and Health; SSR = State sampling region.

4.4 Selection of Dwelling Units

The starting sample size and the sample allocation across the segments were determined based on anticipated eligibility, nonresponse, and the person-level sample selection procedures. Similar to the main study, a small reserve sample (15 percent) of DUs from each segment was selected, and the total sample was partitioned into four probability subsamples within each segment: 100 percent and three 5 percent partitions, for a total of 115 percent. Although the majority of the sample (100/115) was released at the beginning of the DR data collection period, having the additional sample partitions allowed for greater flexibility in controlling the sample size and provided the ability to ensure that the data collection goals were attained within the field period. No additional sample partitions were needed to achieve the target of 2,000 completed interviews.

A total of 5,016 DUs were sampled and yielded 2,087 completed interviews (Table 4.3). As shown in Table 4.2, 185 Spanish-language interviews were yielded from the DR sample. The half-open interval (HOI) procedure for missed dwelling units (DUs) was implemented during the DR, but it is not scheduled to be implemented in the 2014 or 2015 NSDUHs. Table 4.3 compares the expected DR sampling rates and yields to the actual rates and yields.

Table 4.3 Summary of the Dress Rehearsal Sample Design and Results

Statistic	Expected		Actual	
	Total	Rate[1]	Total	Rate
State Sampling Regions	200	N/A	200	N/A
Segments	200	N/A	200	N/A
Selected Dwelling Units	5,146	N/A	5,016	N/A
Eligible Dwelling Units	4,426	0.86	4,392	0.88
Completed Screening Interviews	3,673	0.83	3,511	0.80
Selected Persons	2,703		2,808	
Completed Interviews	2,000	0.74	2,087	0.74

[1] Expected eligibility and screening rates are the observed rates from the 2012 Questionnaire Field Test (QFT) (unweighted). The expected interview response rate is the observed rate from the QFT adjusted with 2011 National Survey on Drug Use and Health (NSDUH) rates to account for the oversampling of high Spanish-language interview areas.

4.5 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. Among the 26 or older age groups, 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 2014 NSDUH and the QFT. One implication of the respondent sample allocation by age groups is a potential impact on DR response rates. As with the QFT, retaining more of the 26 or older adults identified in households to complete interviews had a negative effect on unweighted interview response rates. Both the weighted and unweighted interview response rates for persons younger than 26 were higher than the response rates for persons aged 26 or older. The unweighted interview response rate for the DR sample was 74.32 percent compared with 78.01 percent for the 2012 main study comparison sample and 80.71 percent for the 2013 quarters 3 and 4 main study comparison sample. Weighted interview response rates are not affected by the change in age allocation. 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., relative to the total population in this age group, 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.

4.6 Selection of Persons

After DUs were selected within each DR segment, an FI 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 program, which automatically implemented this stage of selection based on the segment and age group sampling parameters. As indicated in Table 4.3, 2,808 people were selected from within 3,511 screened and eligible DUs, which yielded 2,087 completed interviews.

4.7 Creation of Dress Rehearsal Variance Estimation Strata and Replicates

The nature of the stratified, clustered sampling design of the NSDUH main study, DR, and QFT samples requires that the design structure be taken into consideration when computing variances of survey estimates. Because the DR and QFT samples are assumed to be independent, two sets of key nesting variables (pseudo-strata and replicates) were utilized in the analyses. One set captured the design structure of the DR, and the other captured the design structure of the QFT. Both sets of nesting variables were mapped to the main study comparison data to allow for comparisons with the DR and QFT samples. The development of the QFT nesting variables is further discussed in the 2012 QFT final report (Currivan et al., 2013). This section outlines the creation of the DR nesting variables.

To allow for comparisons between the DR and main study samples, a common set of stratification and clustering variables were defined. Because SSRs serve as strata for the main study samples and as primary sampling units (PSUs) for the DR 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 of analysis was followed in developing a design structure that could be used to simultaneously analyze all three samples (Gfroerer, Eyerman, & Chromy, 2002). This methodology was also used to create the nesting variables for the QFT. Steps in the process were as follows:

- Within the five DR sampling strata (high Spanish certainty stratum and four noncertainty census region strata), variance pseudo-strata were formed by assigning two sequential DR-selected SSRs to the same variance pseudo-strata on the sorted sampling frame. Each sampled SSR was then assigned to a replicate (1 or 2). However, there were three DR SSRs per variance pseudo-strata for four randomly selected pseudo-strata. This was necessary because an odd number of DR SSRs were selected in four of the strata. Within these four pseudo-strata, the third SSR was randomly assigned to either replicate 1 or replicate 2. This led to a total of 98 DR variance pseudo-strata, with two replicates per pseudo-strata.
- The main study SSRs that comprised the DR certainty stratum received the same pseudo-strata assignments as the DR certainty SSRs. For the noncertainty SSRs, the main study SSRs not selected for the DR were assigned to DR sampling pseudo-strata sequentially on the sorted SSR frame, in accordance with the assignments of selected DR SSRs. These assignments kept the number of SSRs per pseudo-strata as equal as possible given the distribution of DR-sampled SSRs within the sorted SSR frame. For both certainty and noncertainty SSRs, the original replicate assignments of either replicate 1 or replicate 2 were maintained for the main study.

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 SSRs between the DR and main study samples.

References

Chen, P., Cribb, D., Dai, L., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (2014). 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. 41, 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. 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.

Currivan, D., Chen, P., Granger, B., Heller, D., Kroutil, L., LeBaron, P., McHenry, G., McKamey, A., Moore, A., Morton, K., Myers, S., Payne, S., Shook-Sa, B., Stringer, C., Terrey, S., Wang, K., & Zelko, H. (2014, January). *National Survey on Drug Use and Health: 2013 Dress Rehearsal draft final report* (prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283201000003C, RTI 0212800). RTI International: Research Triangle Park, NC.

Currivan, D., Kroutil, L., Wang, K., Chen, P., Davis, T., Granger, B., Johnson, I., LeBaron, P., McHenry, G., McKamey, A., Morton, K., Myers, S., Scott, V., Shook-Sa, B., Terrey, S., & Zelko, H. (2013). *National Survey on Drug Use and Health: 2012 Questionnaire Field Test report* (prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283200800004C, RTI 0211838.102.003.008.005). Research Triangle Park, NC: RTI International.

Gfroerer, J., Eyerman, J., & Chromy, J. (Eds.). (2002). *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Iannacchione, V., McMichael, J., Shook-Sa, B., & Morton, K. (2012). *A proposed hybrid sampling frame for the National Survey on Drug Use and Health* (prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-2004-00022). Research Triangle Park, NC: RTI International.

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

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

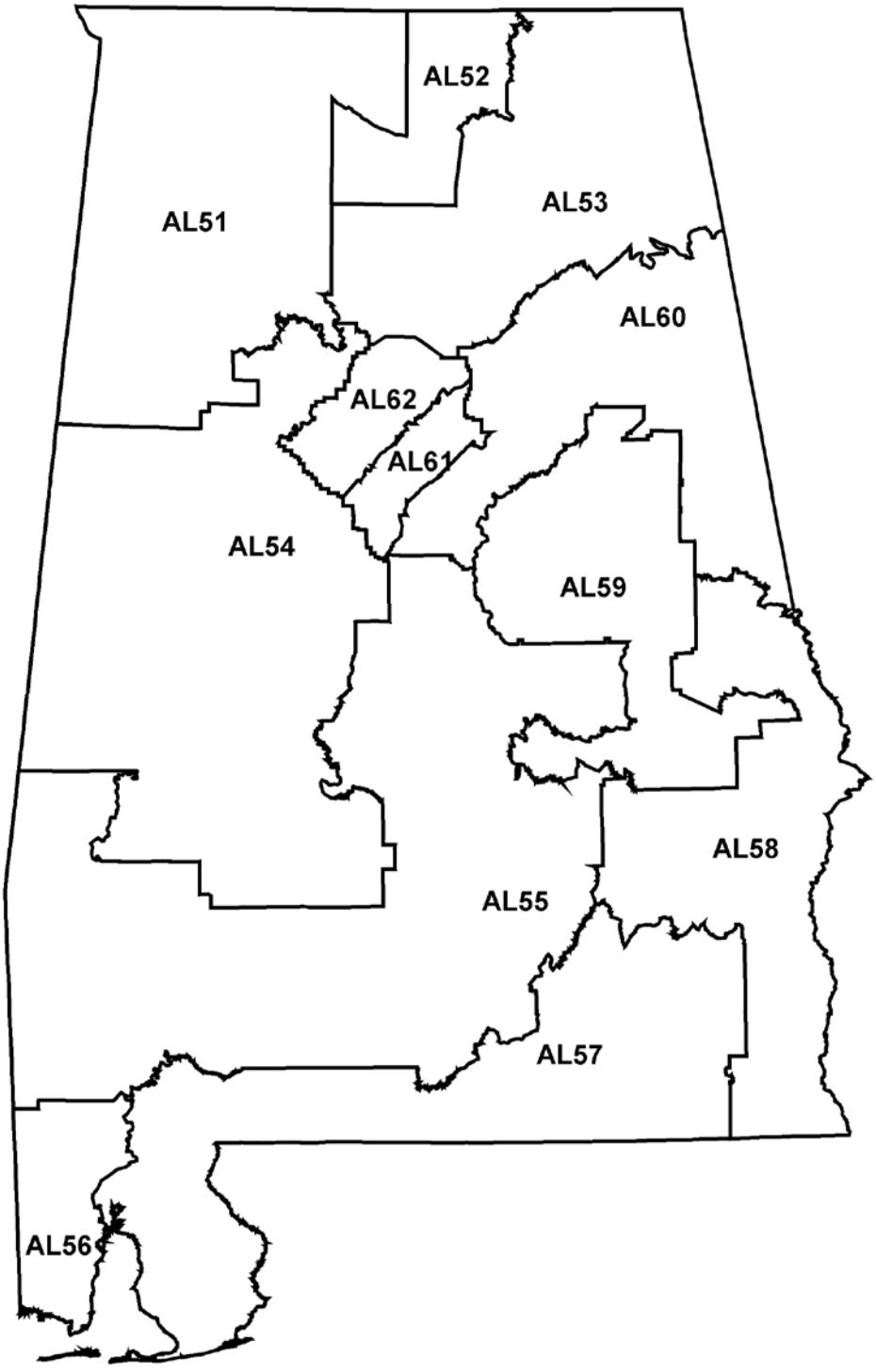
RTI International. (2012a). *2013 National Survey on Drug Use and Health: Counting and listing general manual* (RTI/0212800, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283201000003C). Research Triangle Park, NC: Author.

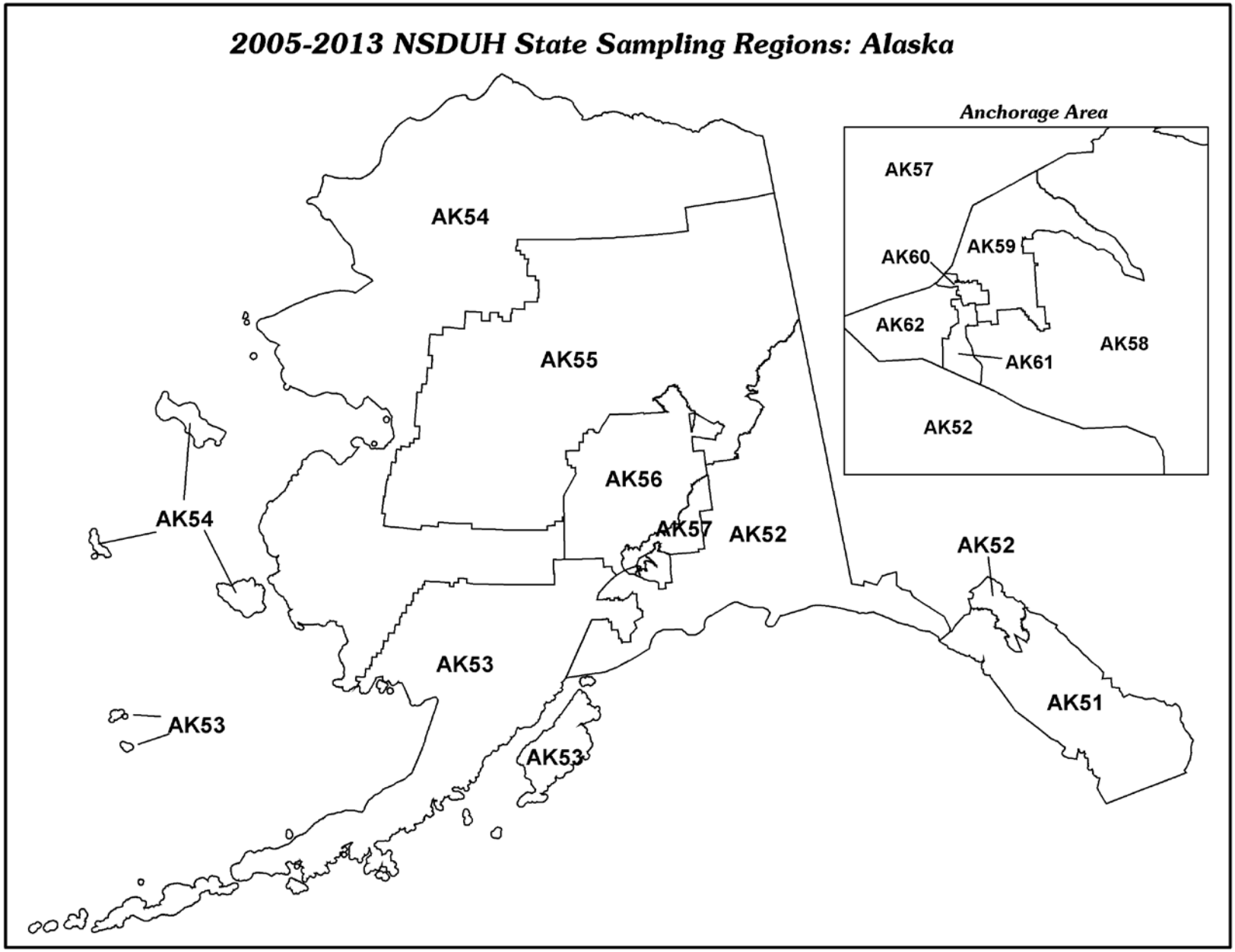
RTI International. (2012b). *SUDAAN®, Release 11.0* [computer software]. Research Triangle Park, NC: Author.

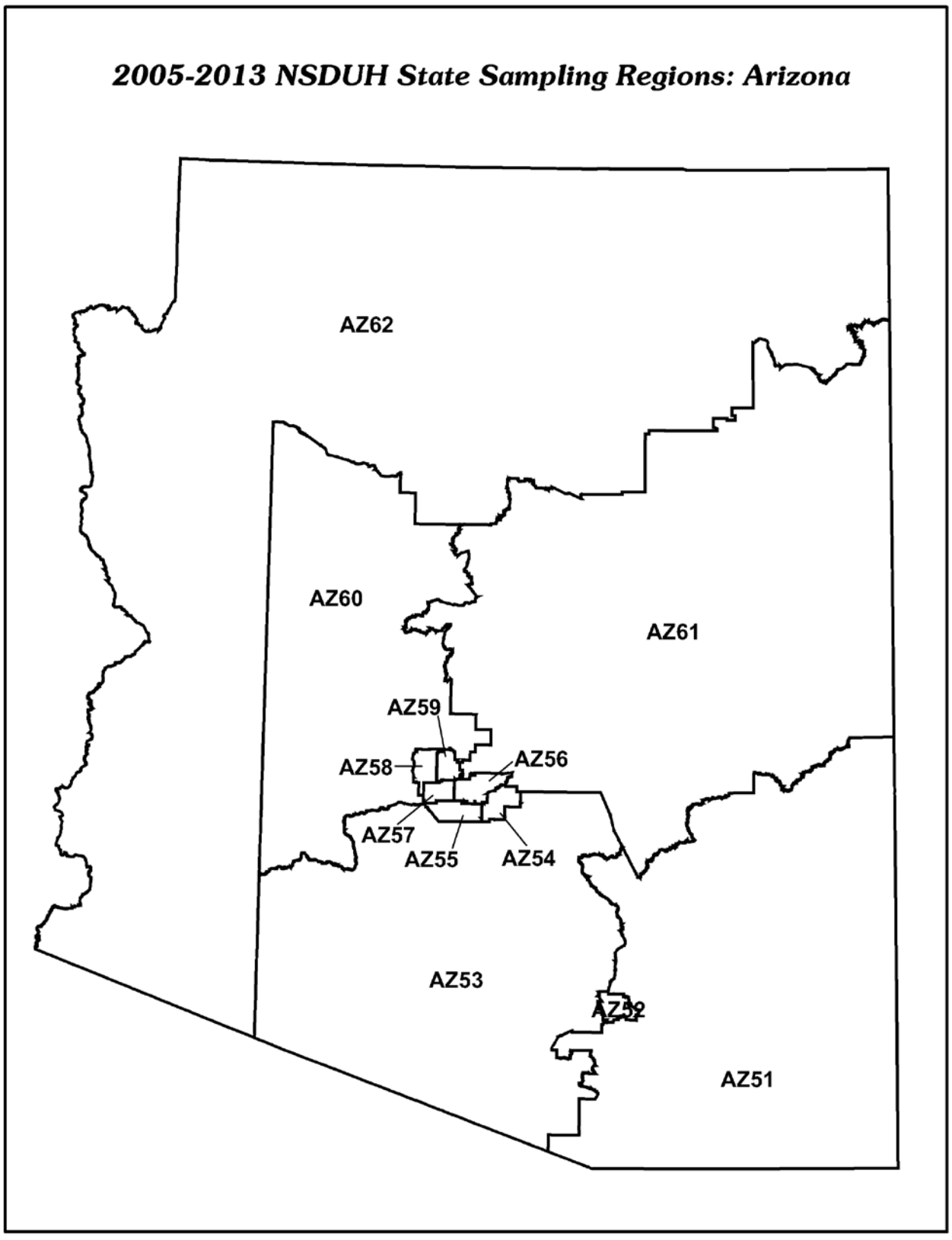
U.S. Census Bureau, Redistricting Data Office. (2009). Appendix A: Geographic terms and concepts. In *2008 Redistricting data prototype (Public Law 94-171) summary file*. Washington, DC: Author. [Available as a PDF at http://www.census.gov/rdo/; see http://www.census.gov/rdo/pdf/pl94-171.pdf]

Appendix A:
2005 through 2013 NSDUH State Sampling Regions

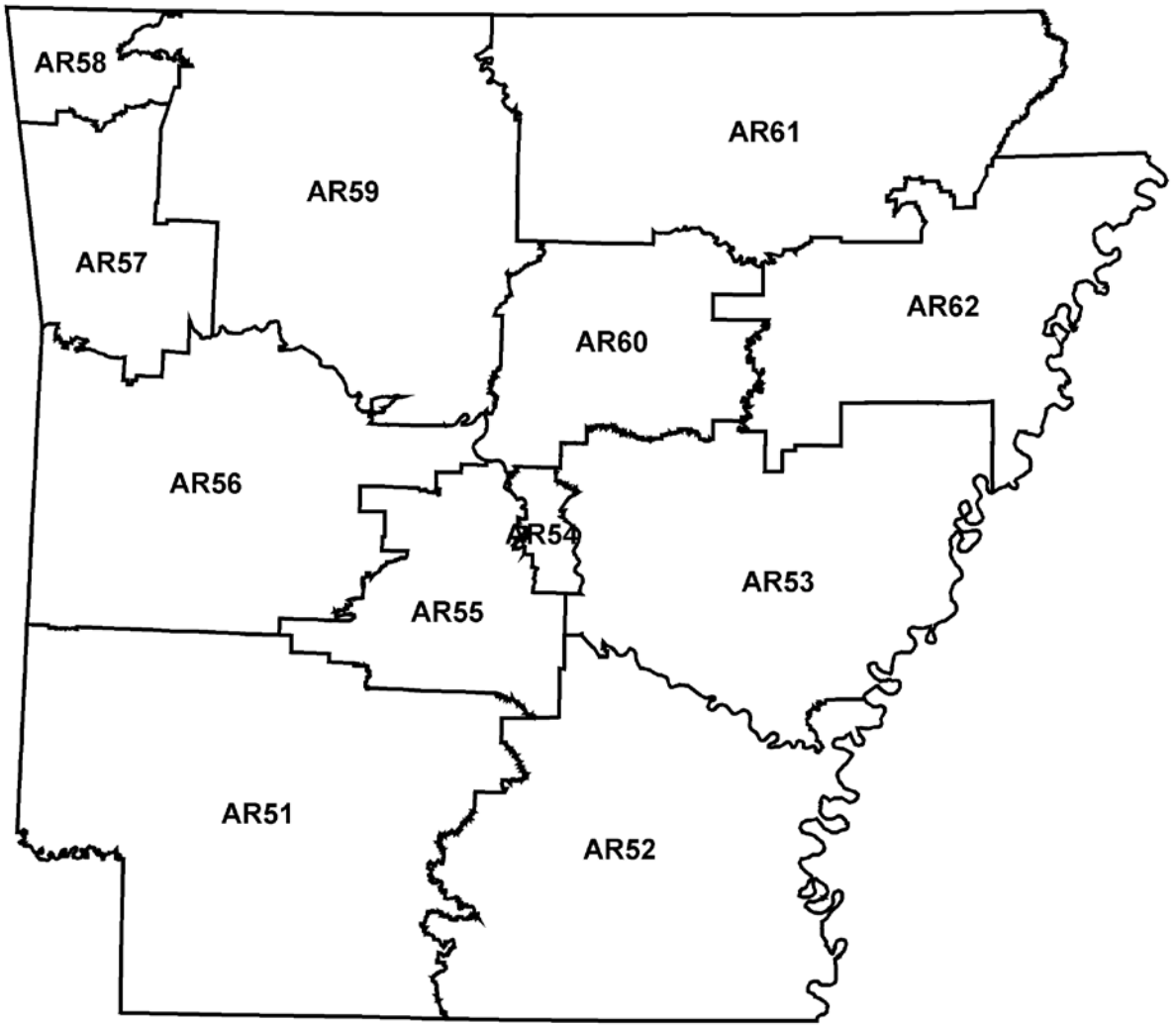
2005-2013 NSDUH State Sampling Regions: Alabama



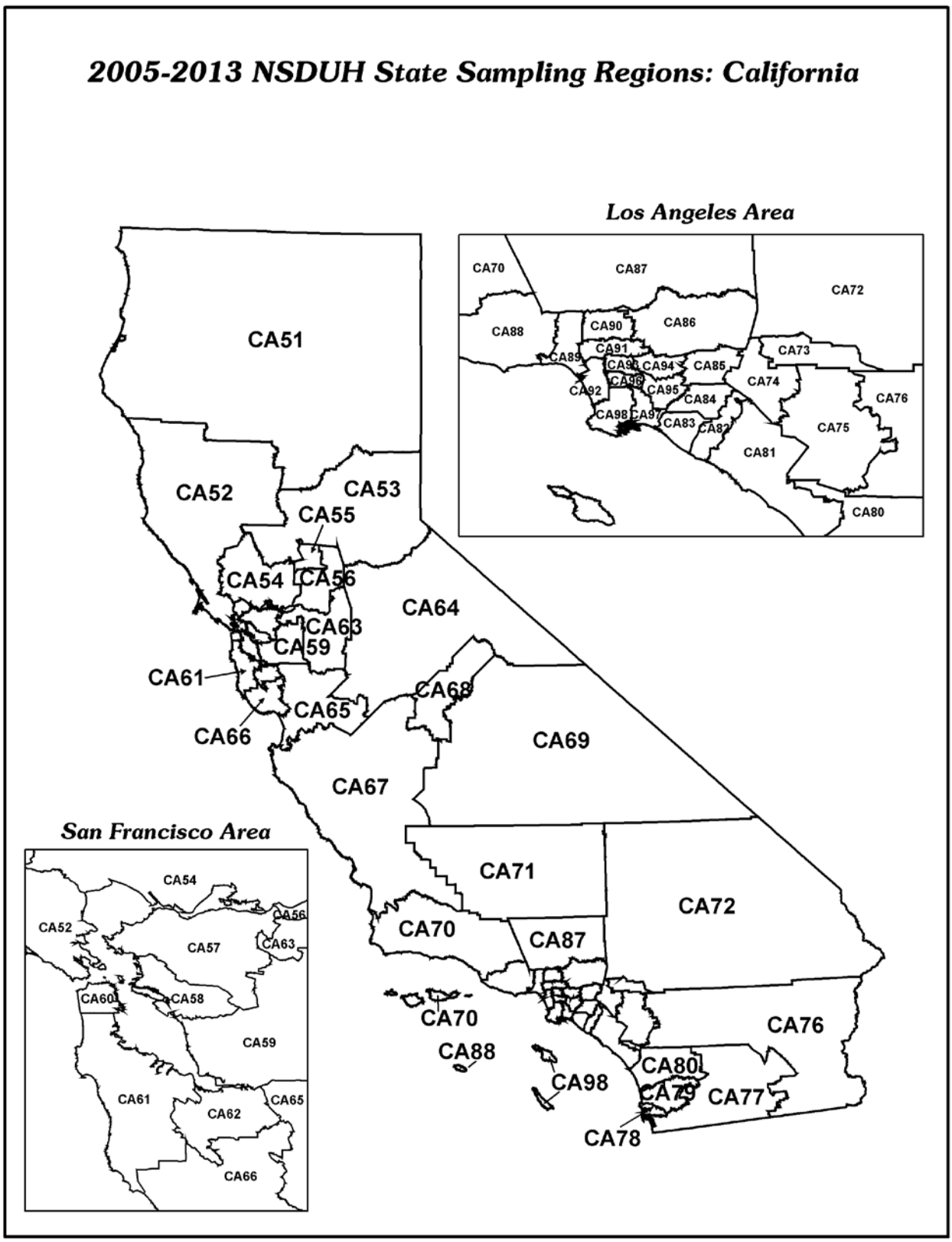

2005-2013 NSDUH State Sampling Regions: Alaska
Anchorage Area
AK57
AK60
AK59
AK62
AK58
AK61
AK52
AK54
AK55
AK56
AK57
AK52
AK54
AK52
AK53
AK51
AK53
AK53

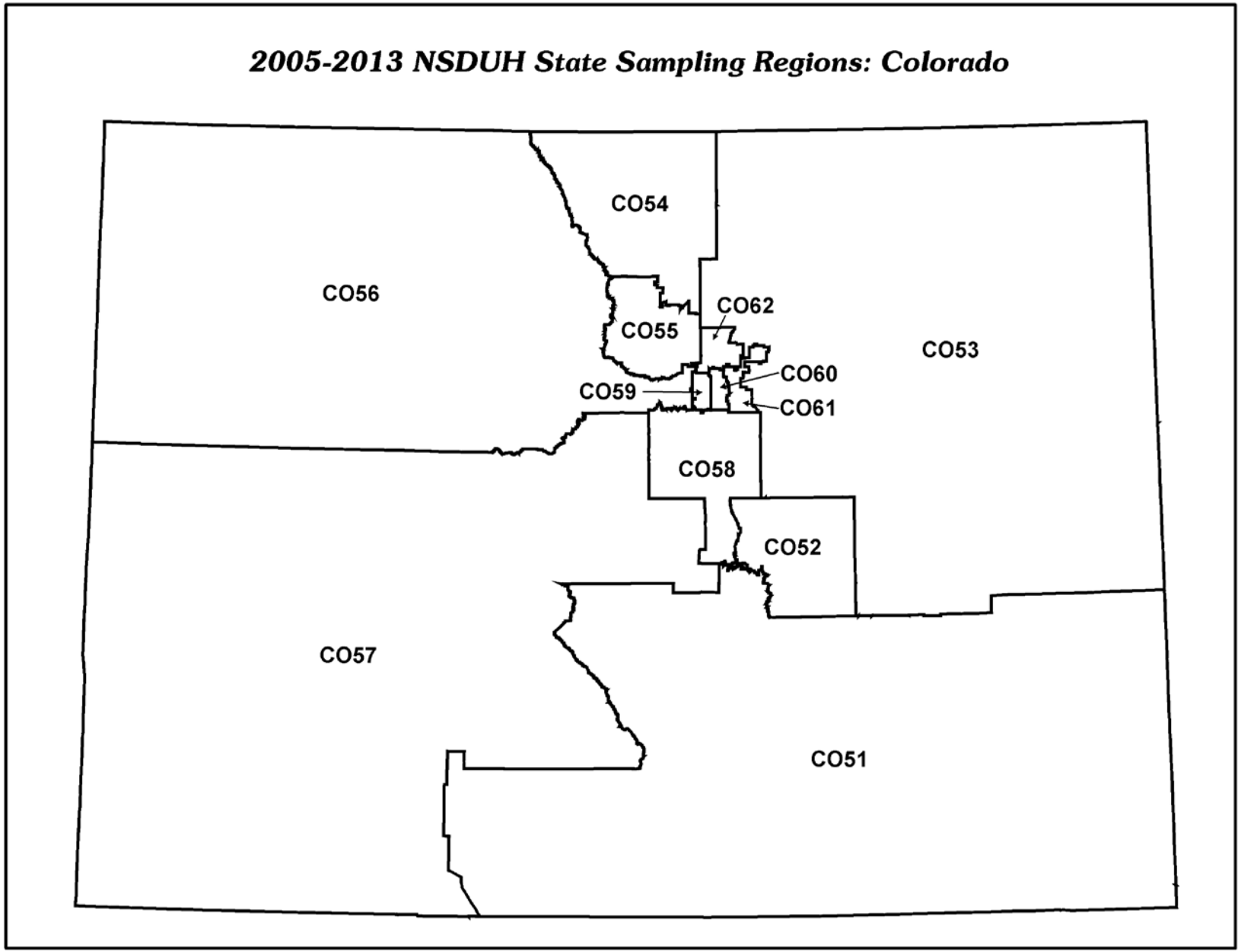

2005-2013 NSDUH State Sampling Regions: Arizona
AZ62
AZ60
AZ61
AZ59
AZ58
AZ56
AZ57
AZ55
AZ54
AZ53
AZ52
AZ51

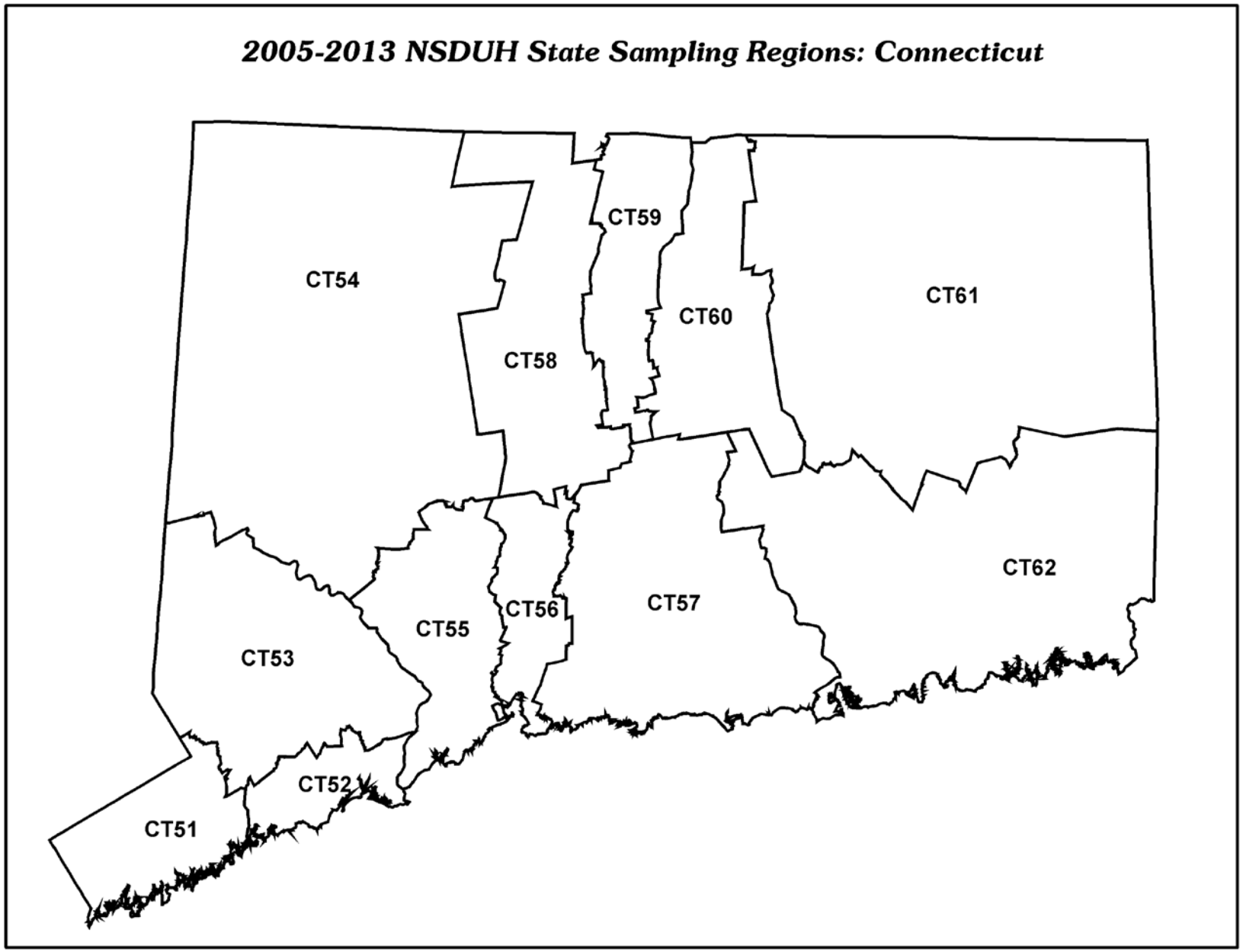
2005-2013 NSDUH State Sampling Regions: Arkansas



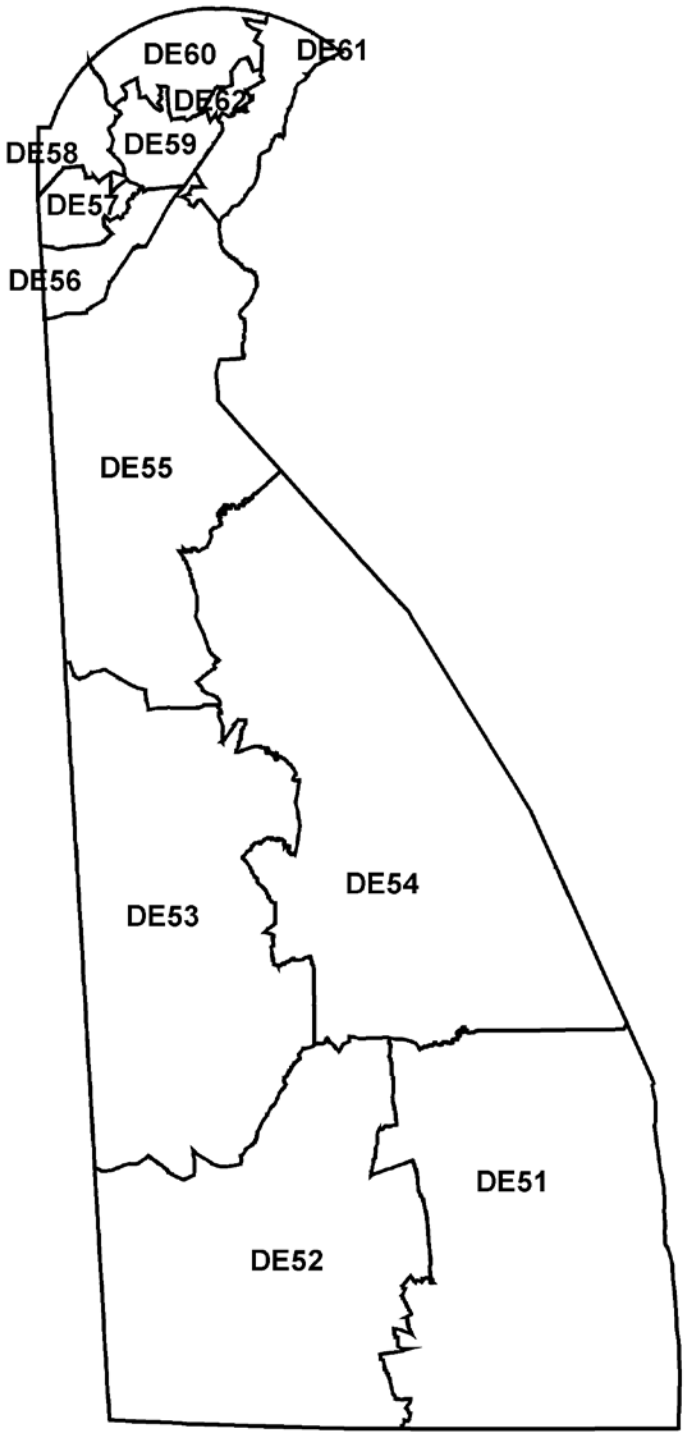
2005-2013 NSDUH State Sampling Regions: California



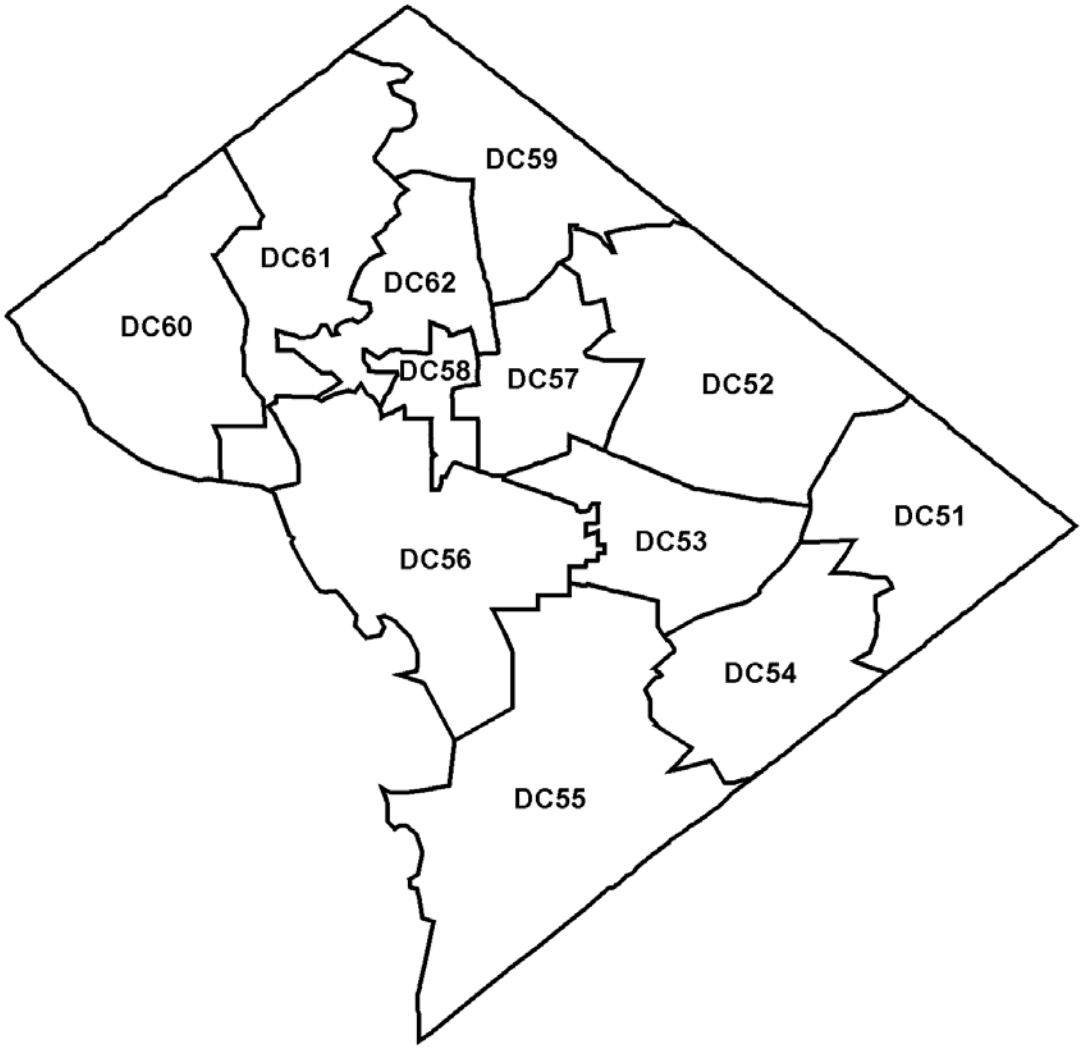

2005-2013 NSDUH State Sampling Regions: Colorado
CO54
CO56
CO62
CO55
CO53
CO60
CO59
CO61
CO58
CO52
CO57
CO51

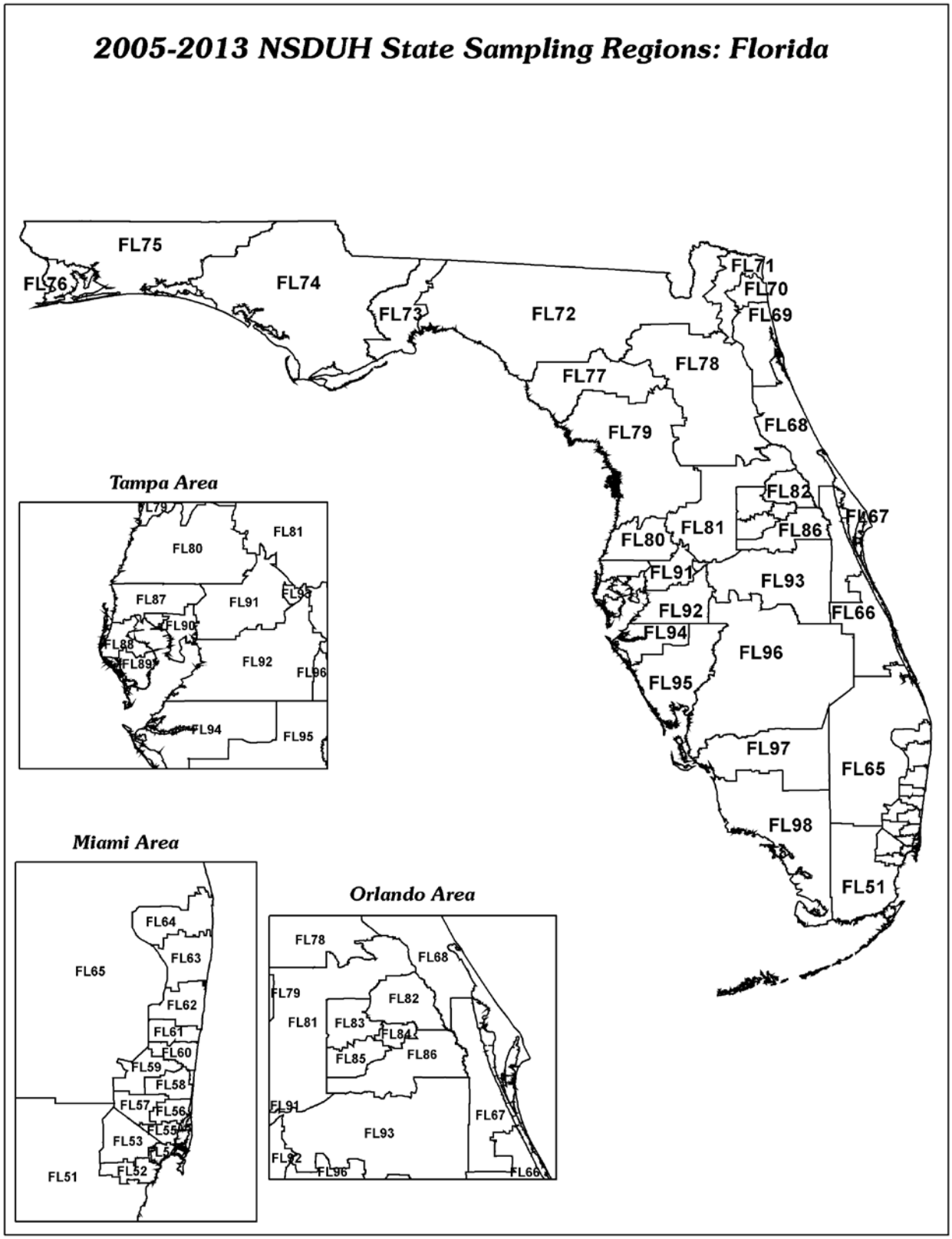

2005-2013 NSDUH State Sampling Regions: Connecticut
CT54
CT59
CT60
CT61
CT58
CT62
CT57
CT56
CT55
CT53
CT52
CT51

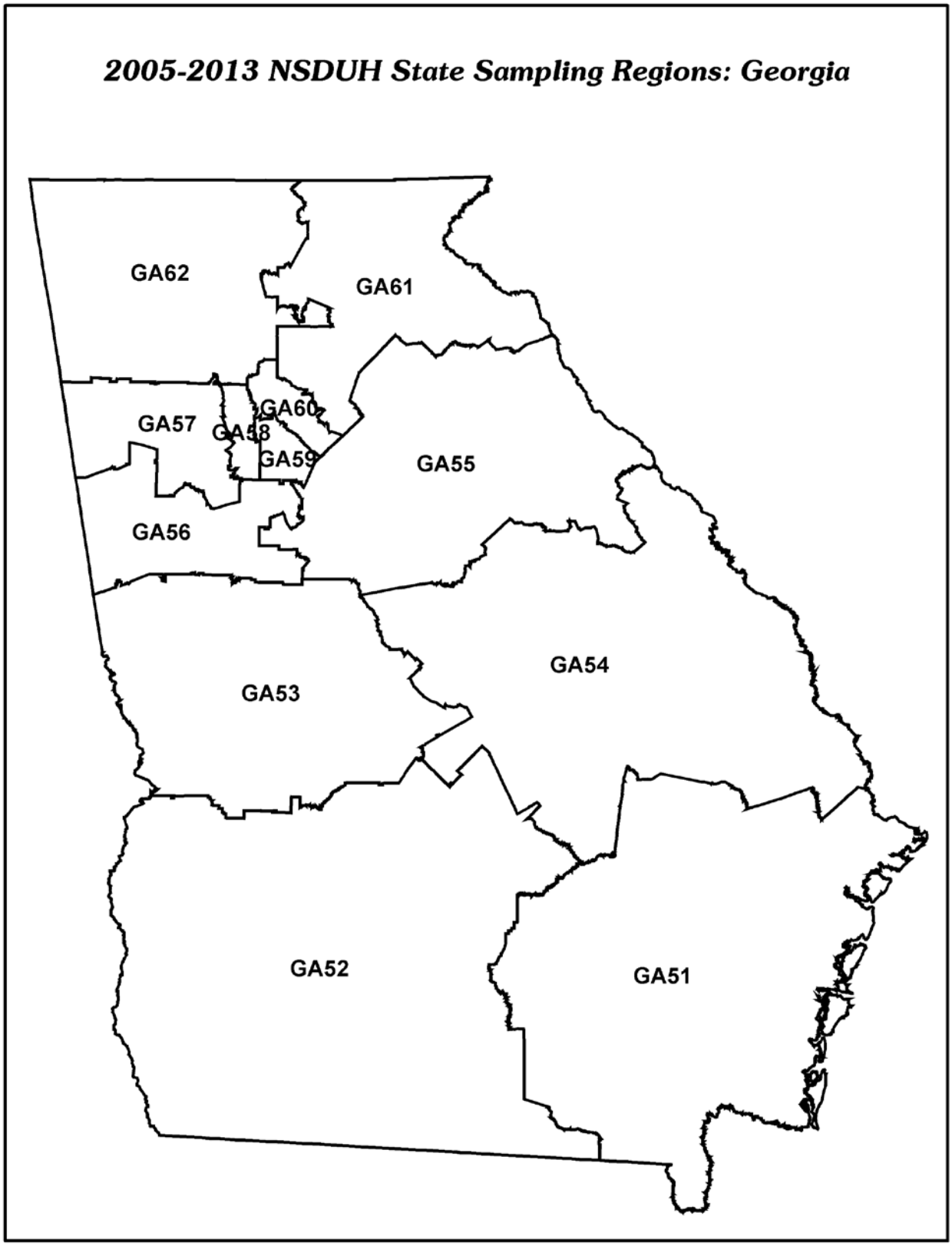
2005-2013 NSDUH State Sampling Regions: Delaware

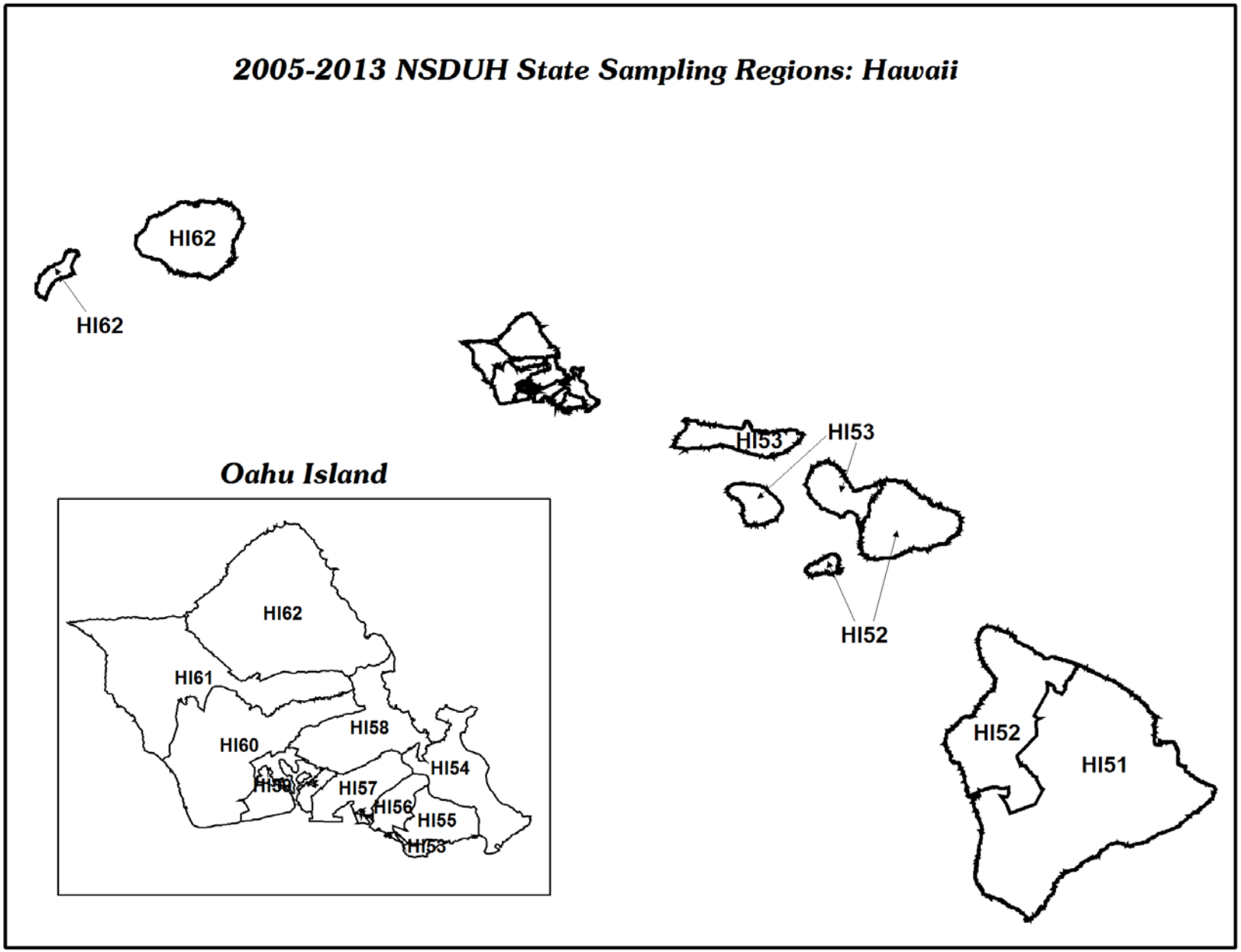


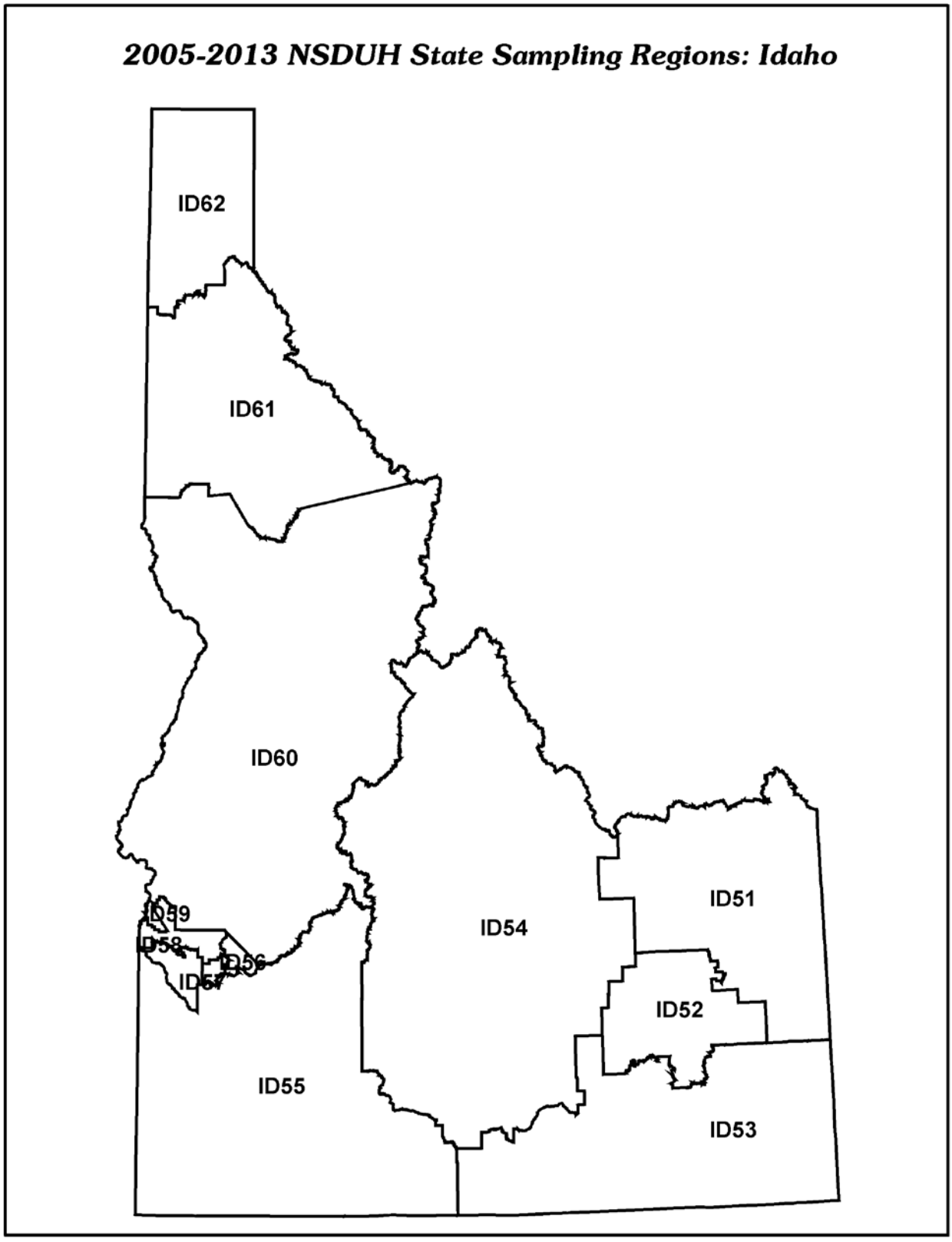
2005-2013 NSDUH State Sampling Regions: District of Columbia

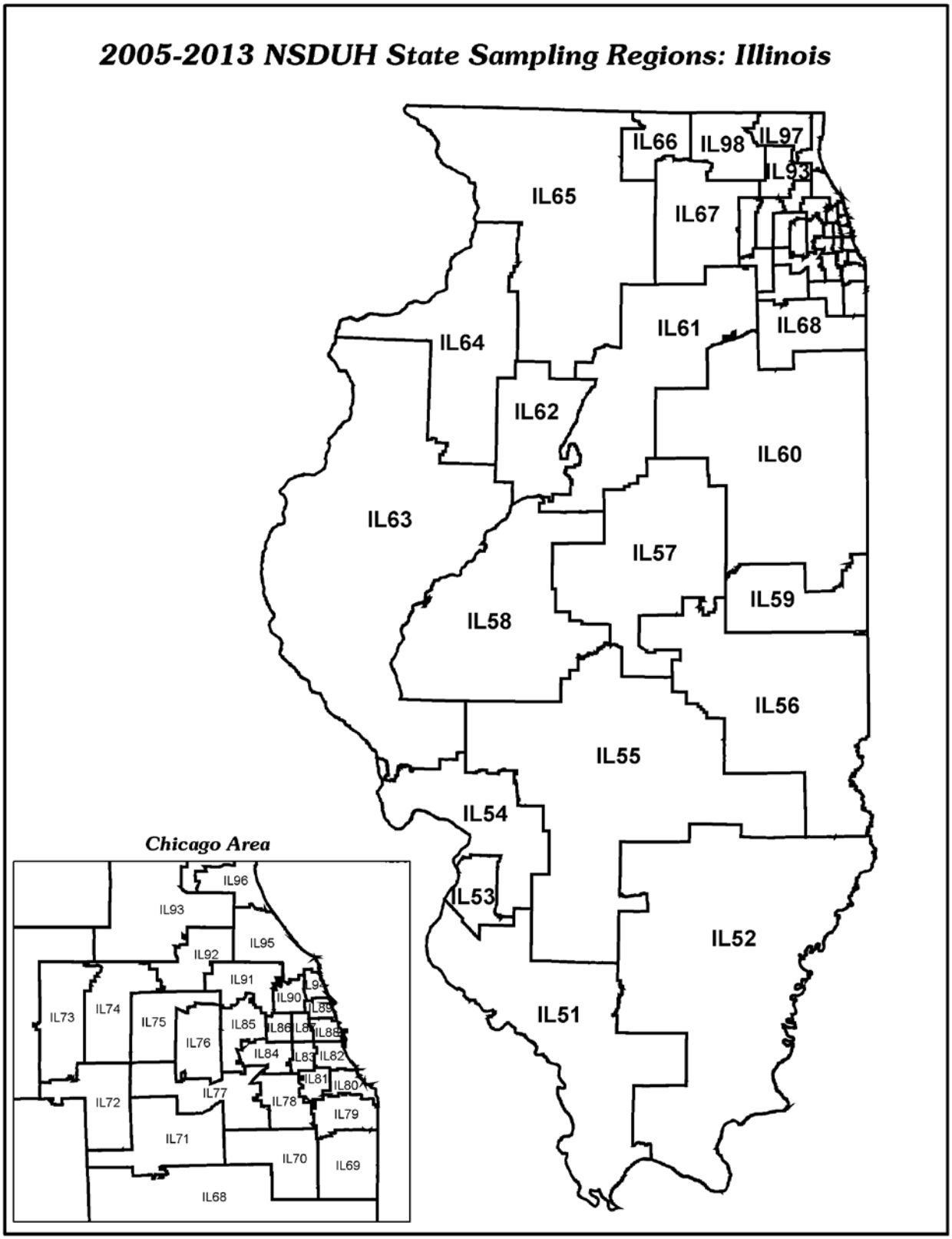



2005-2013 NSDUH State Sampling Regions: Florida
FL75
FL76
FL74
FL73
FL72
FL71
FL70
FL69
FL77
FL78
FL79
FL68
FL82
FL67
FL81
FL80
FL86
FL91
FL93
FL92
FL66
FL94
FL96
FL95
FL97
FL65
FL98
FL51
Tampa Area
FL79
FL81
FL80
FL87
FL91
FL93
FL90
FL88
FL89
FL92
FL96
FL94
FL95
Miami Area
FL64
FL63
FL65
FL62
FL61
FL60
FL59
FL58
FL57
FL56
FL55
FL53
FL52
FL51
Orlando Area
FL78
FL68
FL79
FL82
FL81
FL83
FL84
FL85
FL86
FL91
FL67
FL93
FL92
FL96
FL66

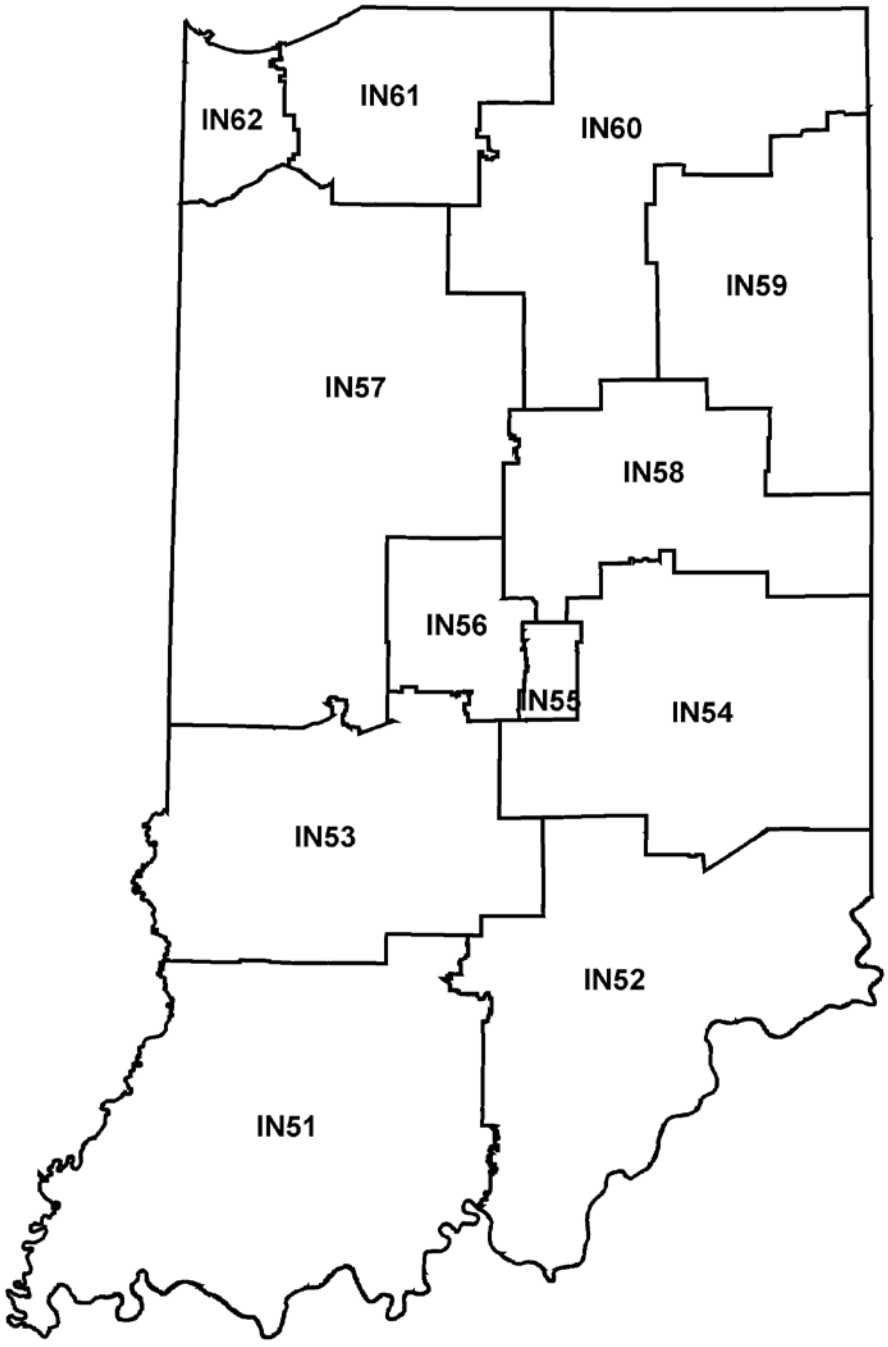

2005-2013 NSDUH State Sampling Regions: Georgia
GA62
GA61
GA60
GA57
GA58
GA59
GA55
GA56
GA54
GA53
GA52
GA51


2005-2013 NSDUH State Sampling Regions: Hawaii
HI62
HI62
HI53
HI53
HI52
HI52
HI51
Oahu Island
HI62
HI61
HI58
HI60
HI54
HI57
HI56
HI55
HI53

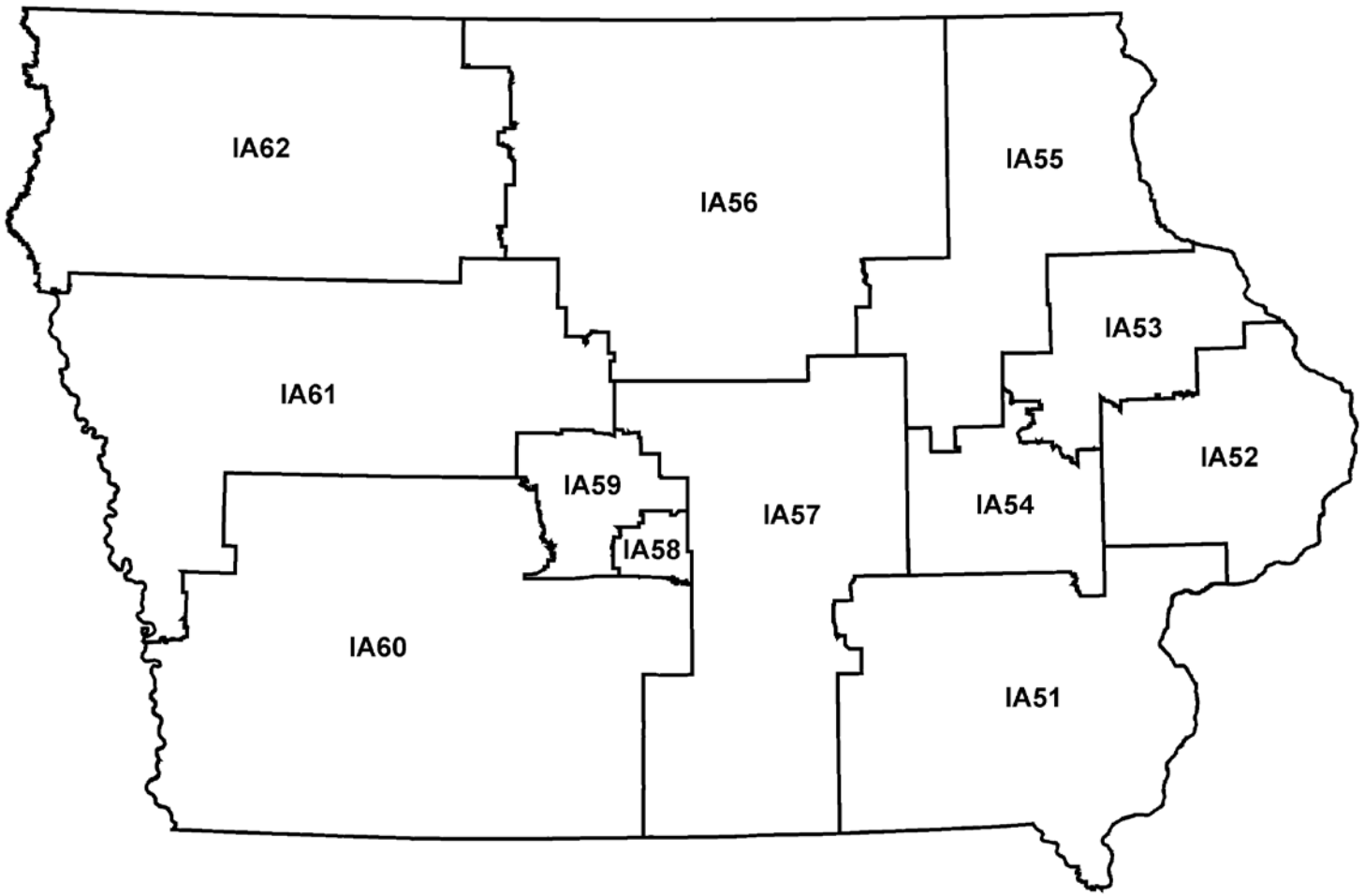

2005-2013 NSDUH State Sampling Regions: Idaho
ID62
ID61
ID60
ID59
ID58
ID56
ID57
ID54
ID51
ID52
ID55
ID53

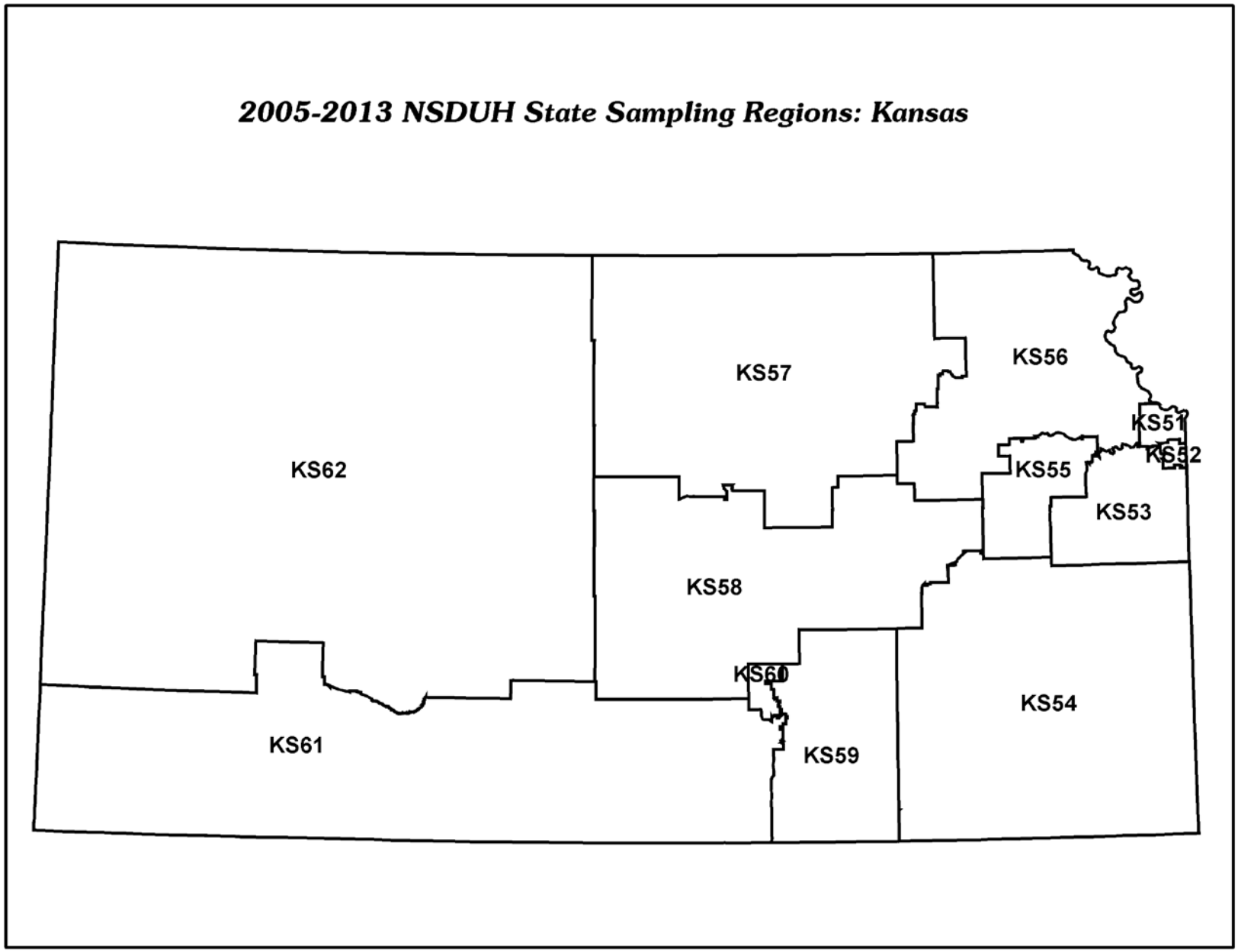

2005-2013 NSDUH State Sampling Regions: Illinois
IL66
IL98
IL97
IL93
IL65
IL67
IL64
IL61
IL68
IL62
IL60
IL63
IL57
IL59
IL58
IL56
IL55
IL54
IL53
IL52
IL51
Chicago Area
IL96
IL93
IL95
IL92
IL91
IL90
IL73
IL74
IL75
IL85
IL76
IL84
IL83
IL82
IL81
IL80
IL77
IL78
IL79
IL72
IL71
IL70
IL69
IL68

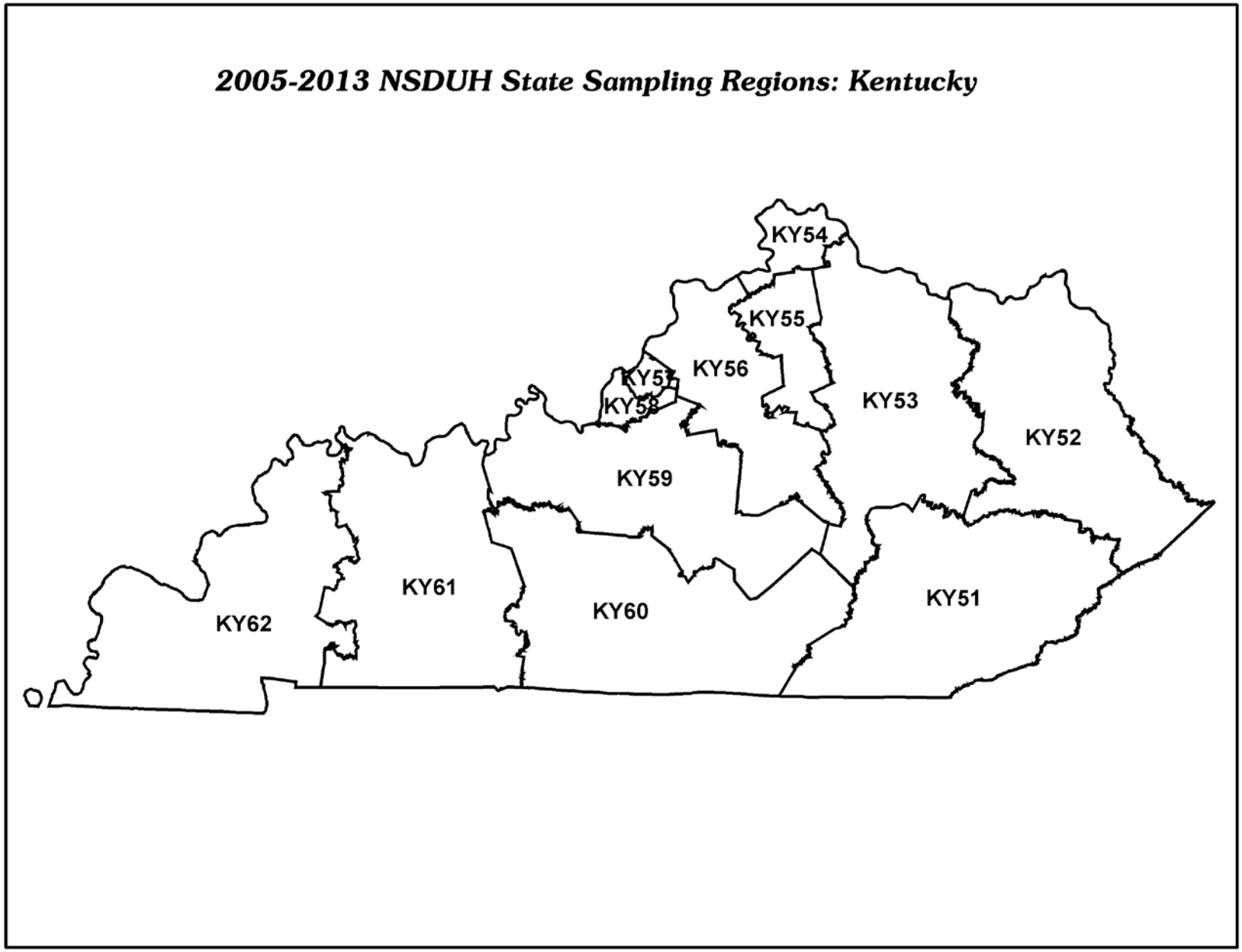
2005-2013 NSDUH State Sampling Regions: Indiana

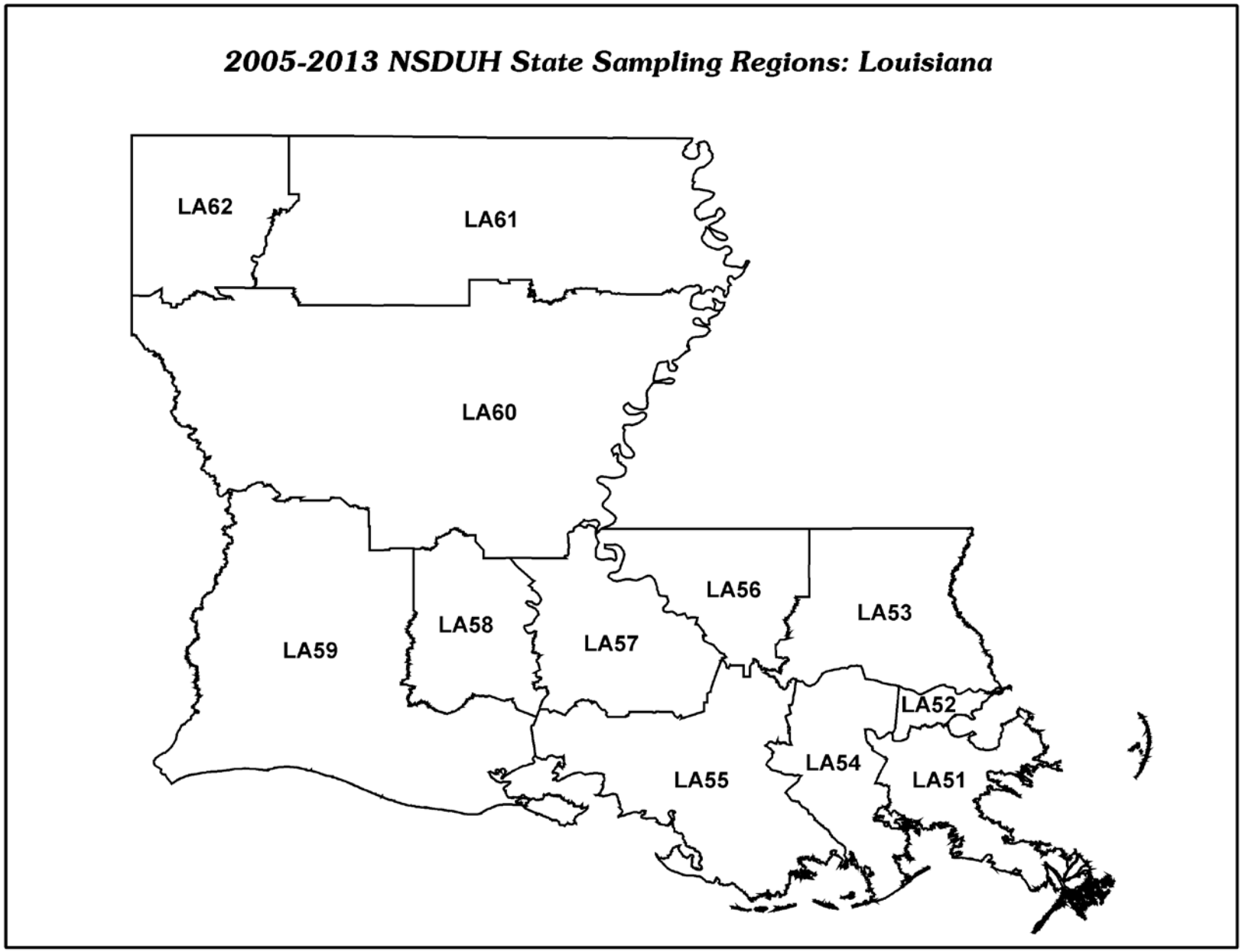


2005-2013 NSDUH State Sampling Regions: Iowa

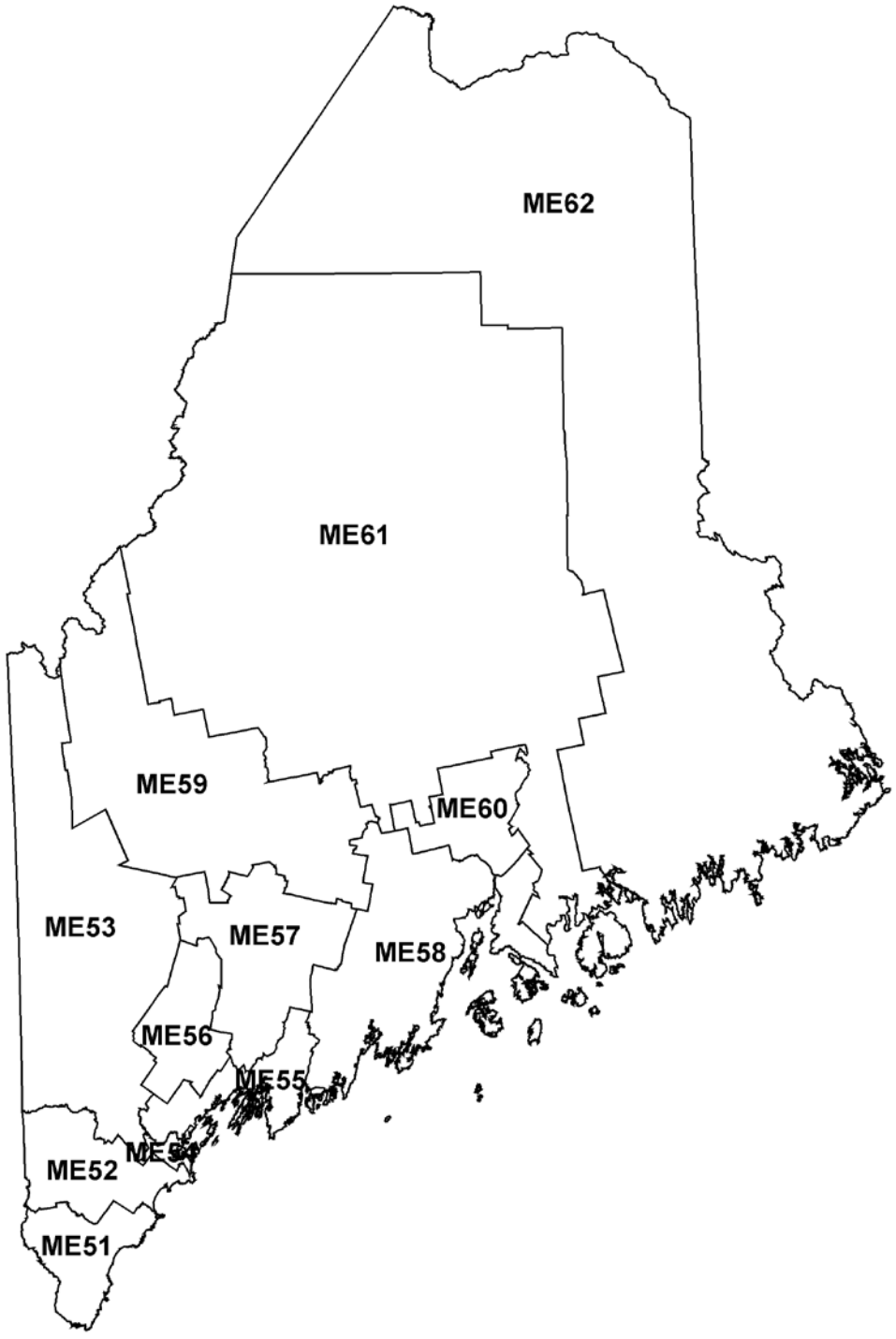


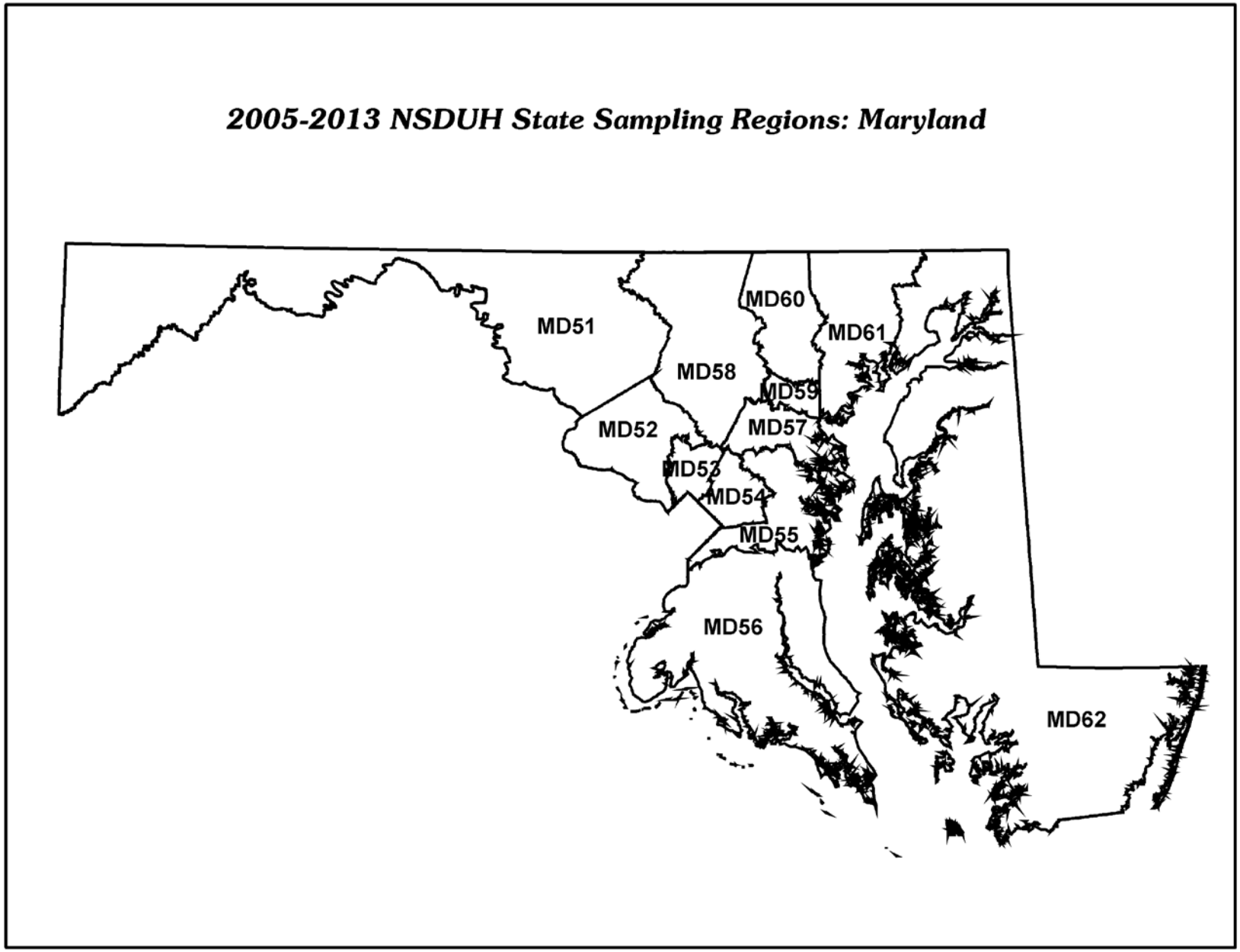

2005-2013 NSDUH State Sampling Regions: Kansas
KS62
KS57
KS56
KS51
KS52
KS55
KS53
KS58
KS60
KS54
KS61
KS59

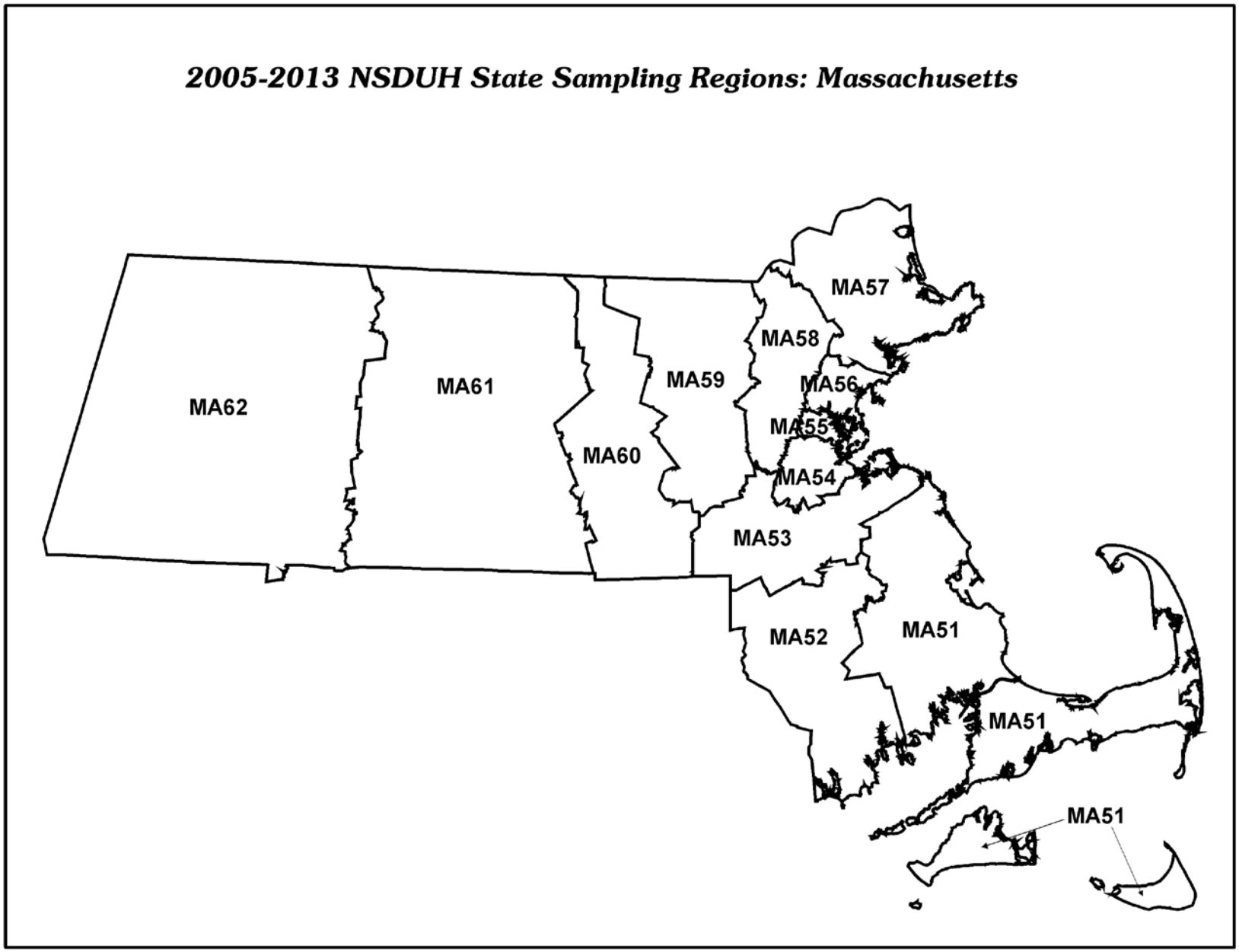

2005-2013 NSDUH State Sampling Regions: Kentucky
KY54
KY55
KY56
KY57
KY58
KY53
KY52
KY59
KY61
KY60
KY51
KY62

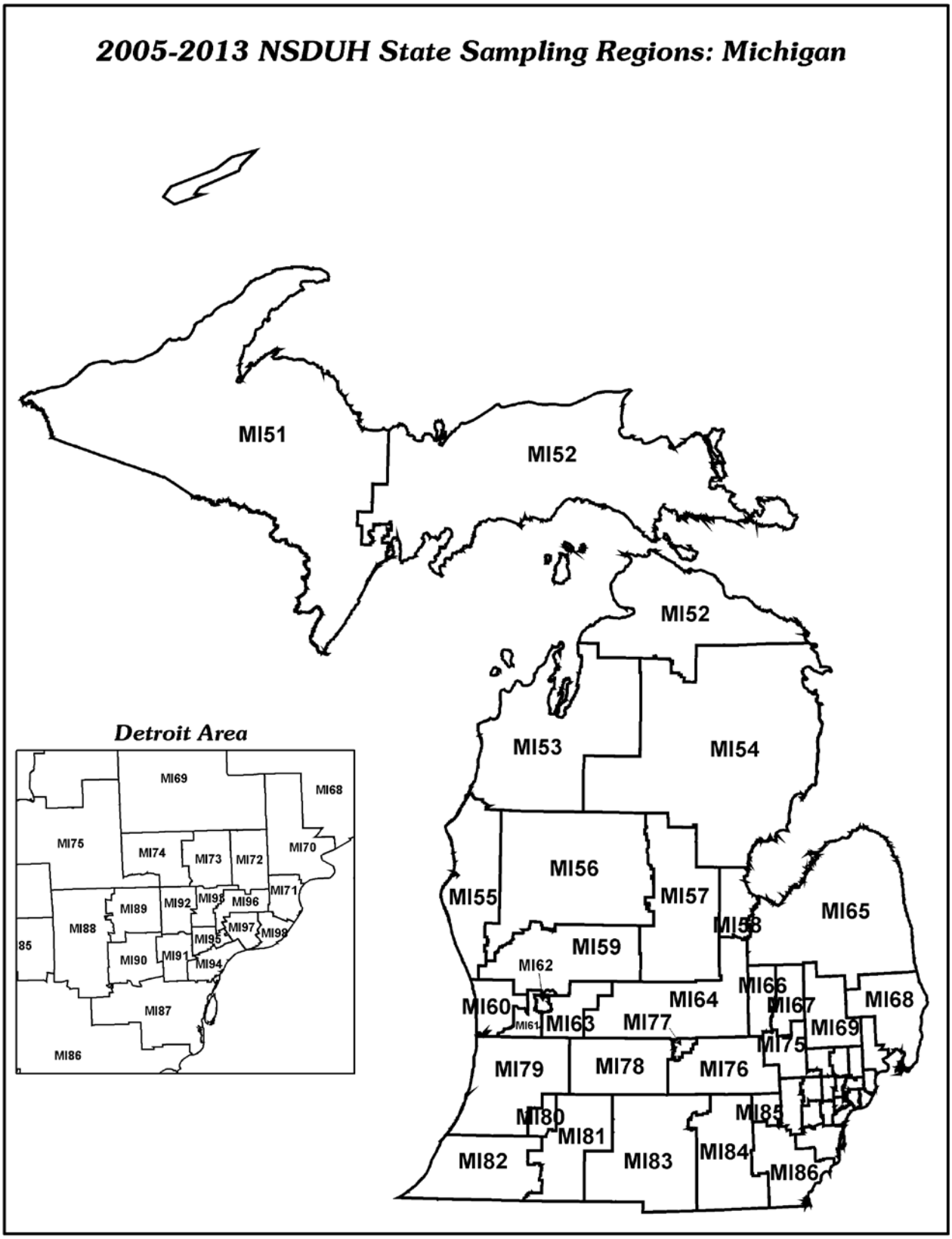

2005-2013 NSDUH State Sampling Regions: Louisiana
LA62
LA61
LA60
LA59
LA58
LA57
LA56
LA53
LA52
LA54
LA55
LA51

2005-2013 NSDUH State Sampling Regions: Maine

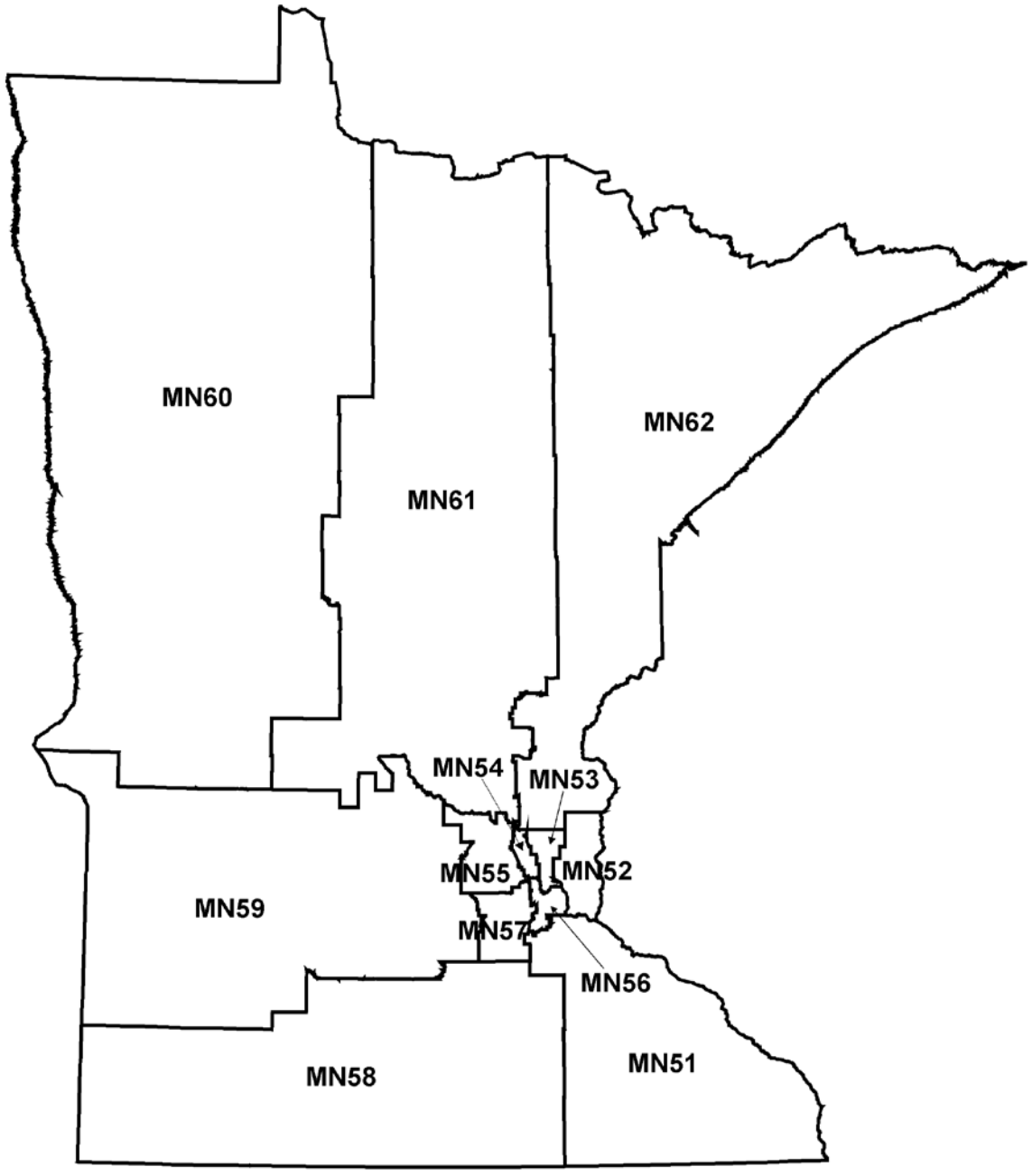



2005-2013 NSDUH State Sampling Regions: Maryland
MD51
MD52
MD53
MD54
MD55
MD56
MD57
MD58
MD59
MD60
MD61
MD62

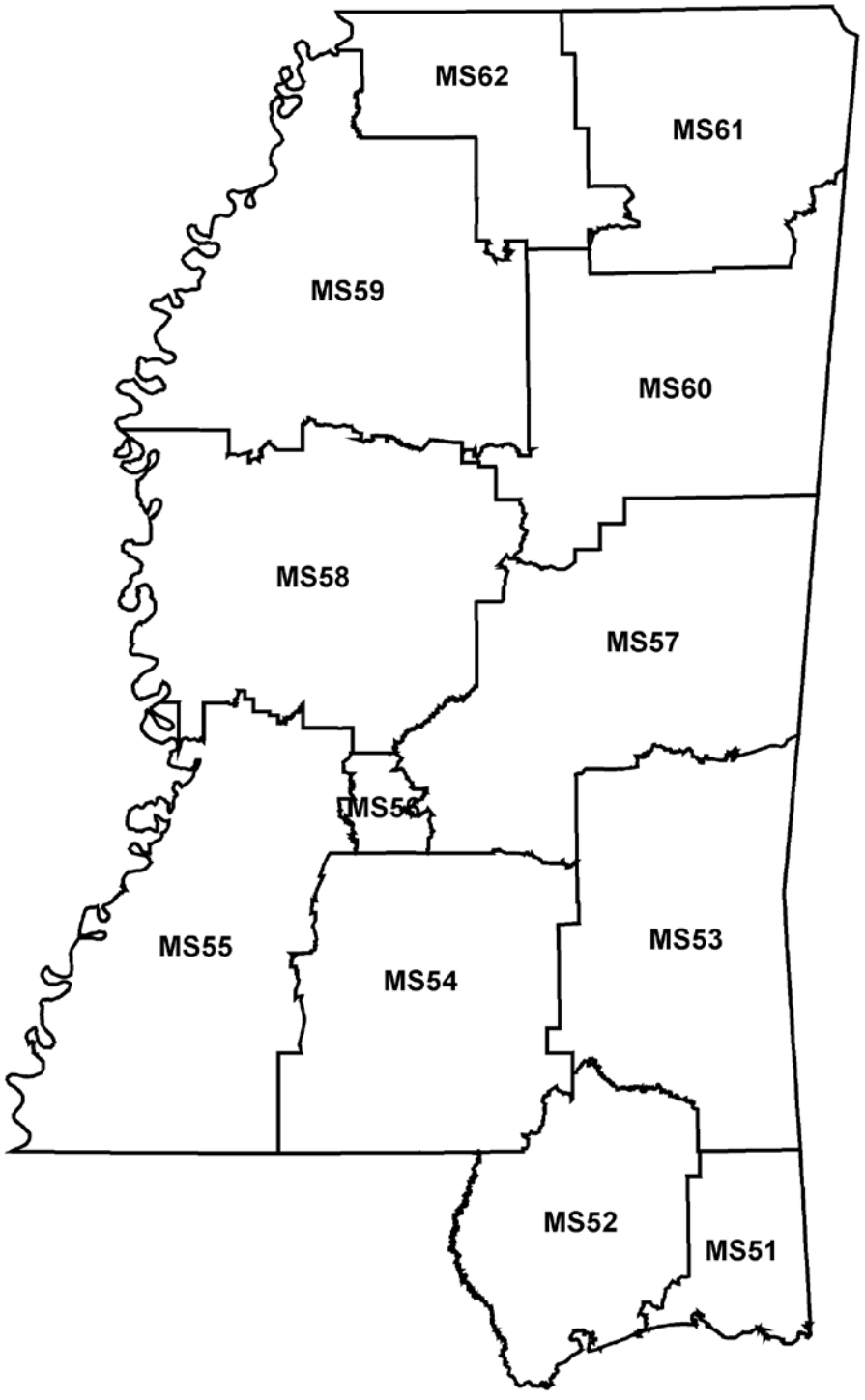

2005-2013 NSDUH State Sampling Regions: Massachusetts
MA62
MA61
MA60
MA59
MA57
MA58
MA56
MA55
MA54
MA53
MA52
MA51
MA51
MA51


2005-2013 NSDUH State Sampling Regions: Michigan
MI51
MI52
MI52
MI53
MI54
MI55
MI56
MI57
MI58
MI59
MI60
MI61
MI62
MI63
MI64
MI65
MI66
MI67
MI68
MI69
MI75
MI76
MI77
MI78
MI79
MI80
MI81
MI82
MI83
MI84
MI85
MI86
Detroit Area
MI69
MI68
MI75
MI74
MI73
MI72
MI70
MI71
MI89
MI92
MI95
MI96
MI88
MI97
MI99
MI95
85
MI90
MI91
MI94
MI87
MI86

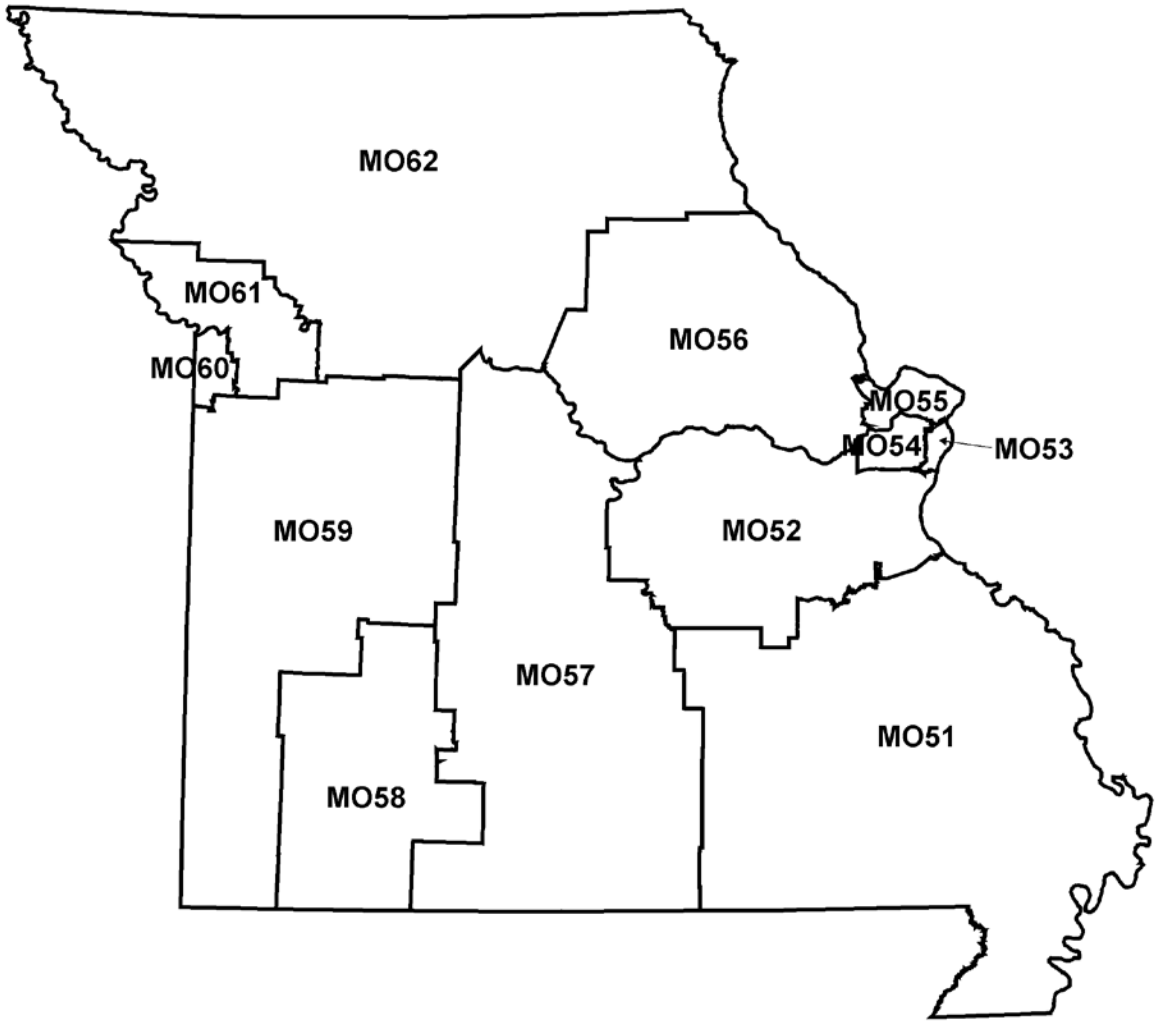
2005-2013 NSDUH State Sampling Regions: Minnesota

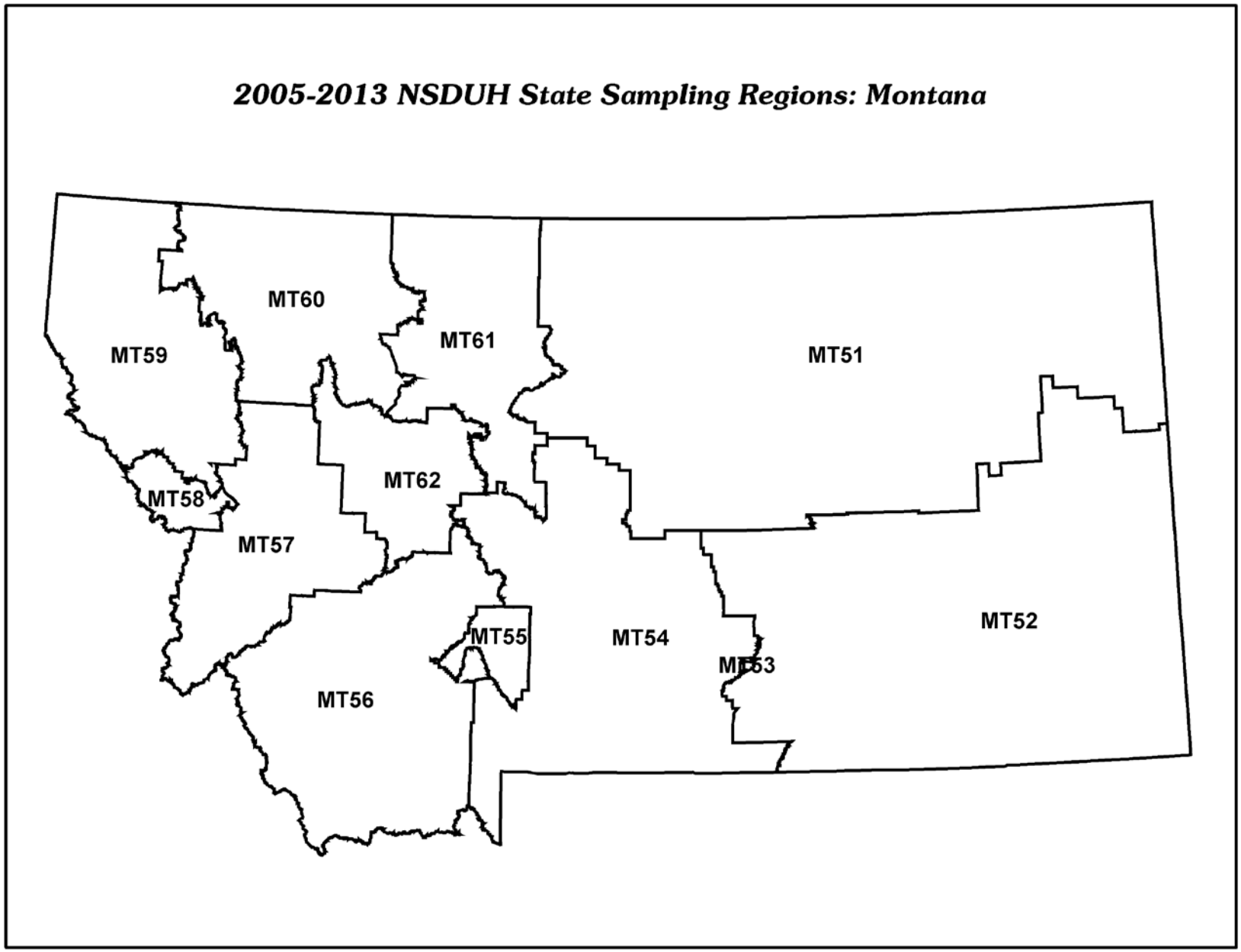


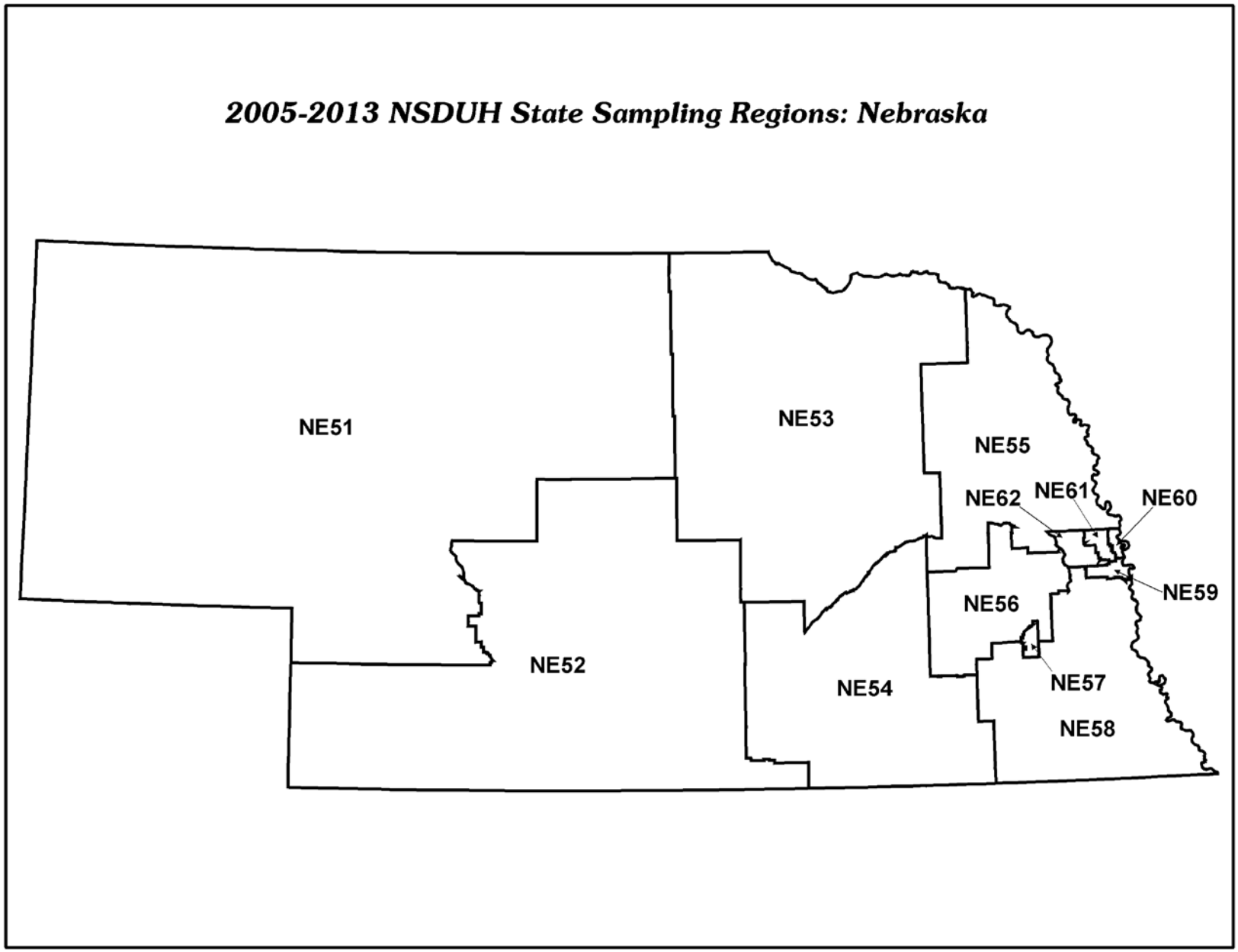
2005-2013 NSDUH State Sampling Regions: Mississippi

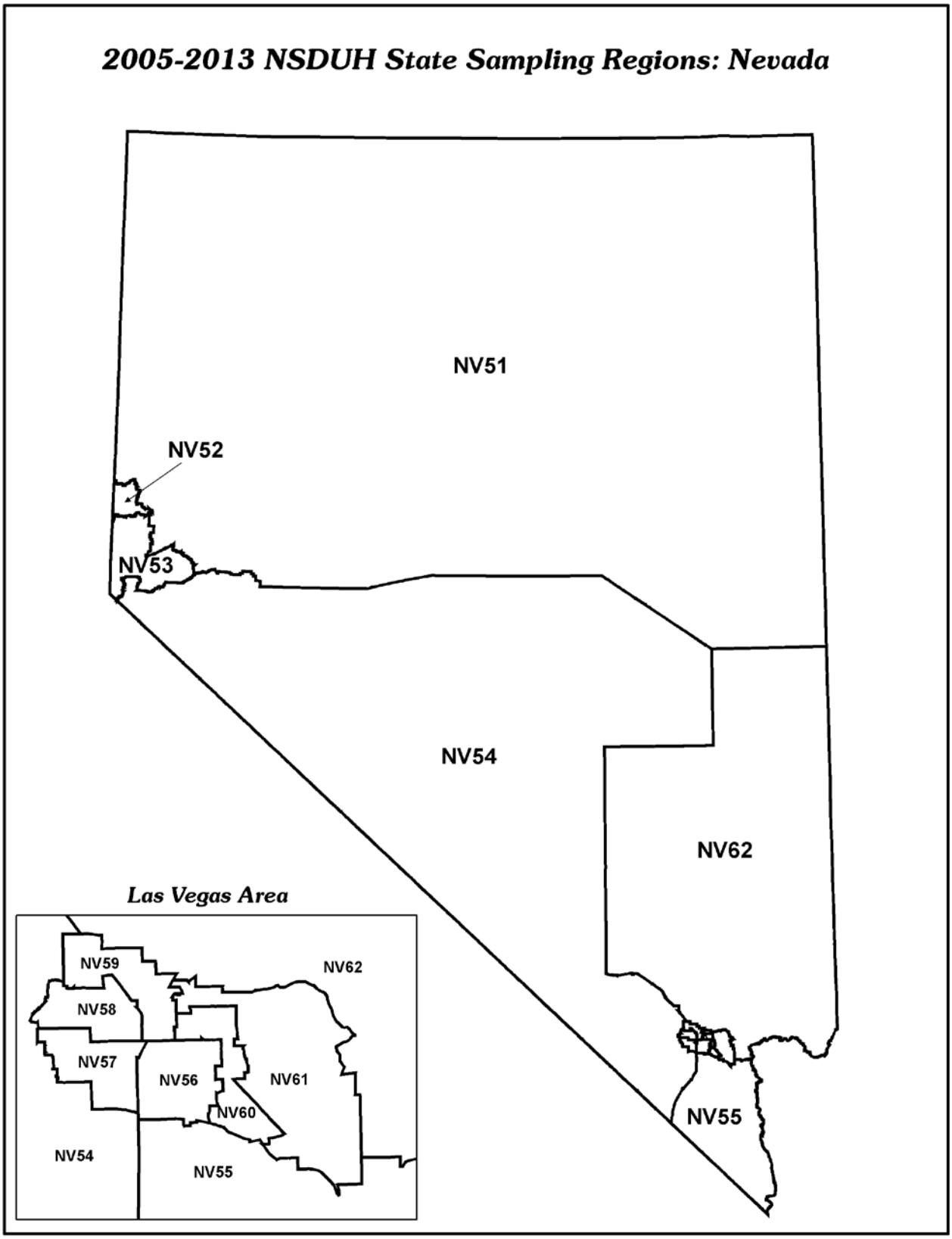


2005-2013 NSDUH State Sampling Regions: Missouri

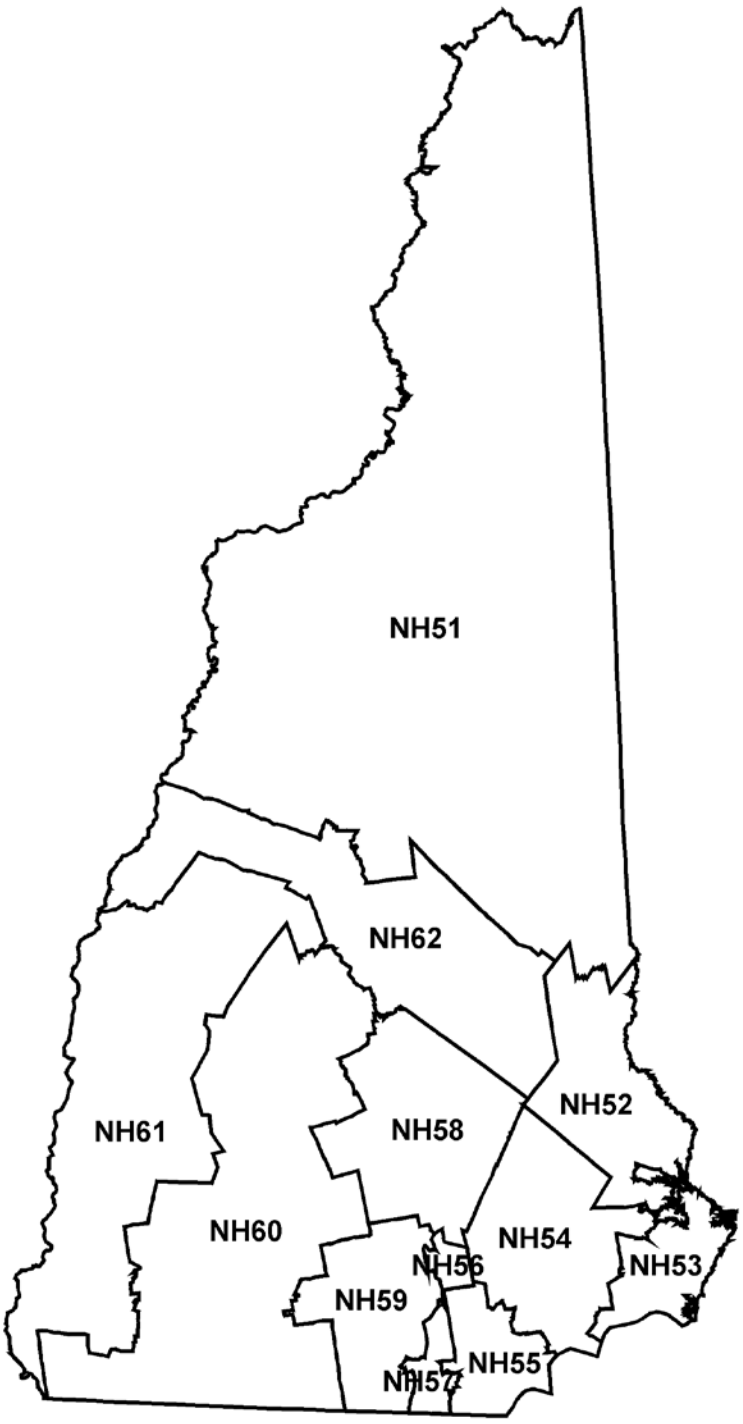



2005-2013 NSDUH State Sampling Regions: Montana
MT60
MT59
MT61
MT51
MT62
MT58
MT57
MT55
MT54
MT52
MT53
MT56

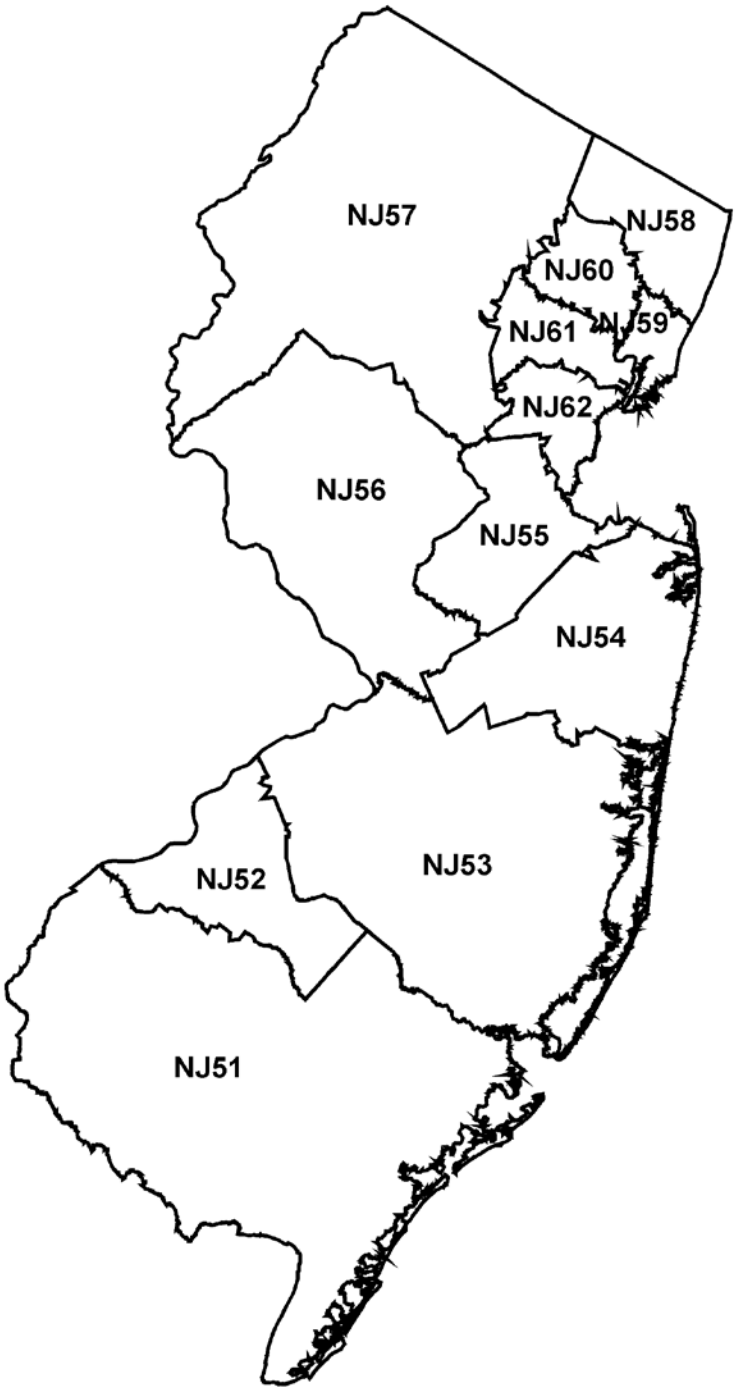

2005-2013 NSDUH State Sampling Regions: Nebraska
NE51
NE53
NE55
NE62
NE61
NE60
NE59
NE56
NE52
NE54
NE57
NE58


2005-2013 NSDUH State Sampling Regions: Nevada
NV51
NV52
NV53
NV54
NV62
NV55
Las Vegas Area
NV59
NV62
NV58
NV57
NV56
NV61
NV60
NV54
NV55

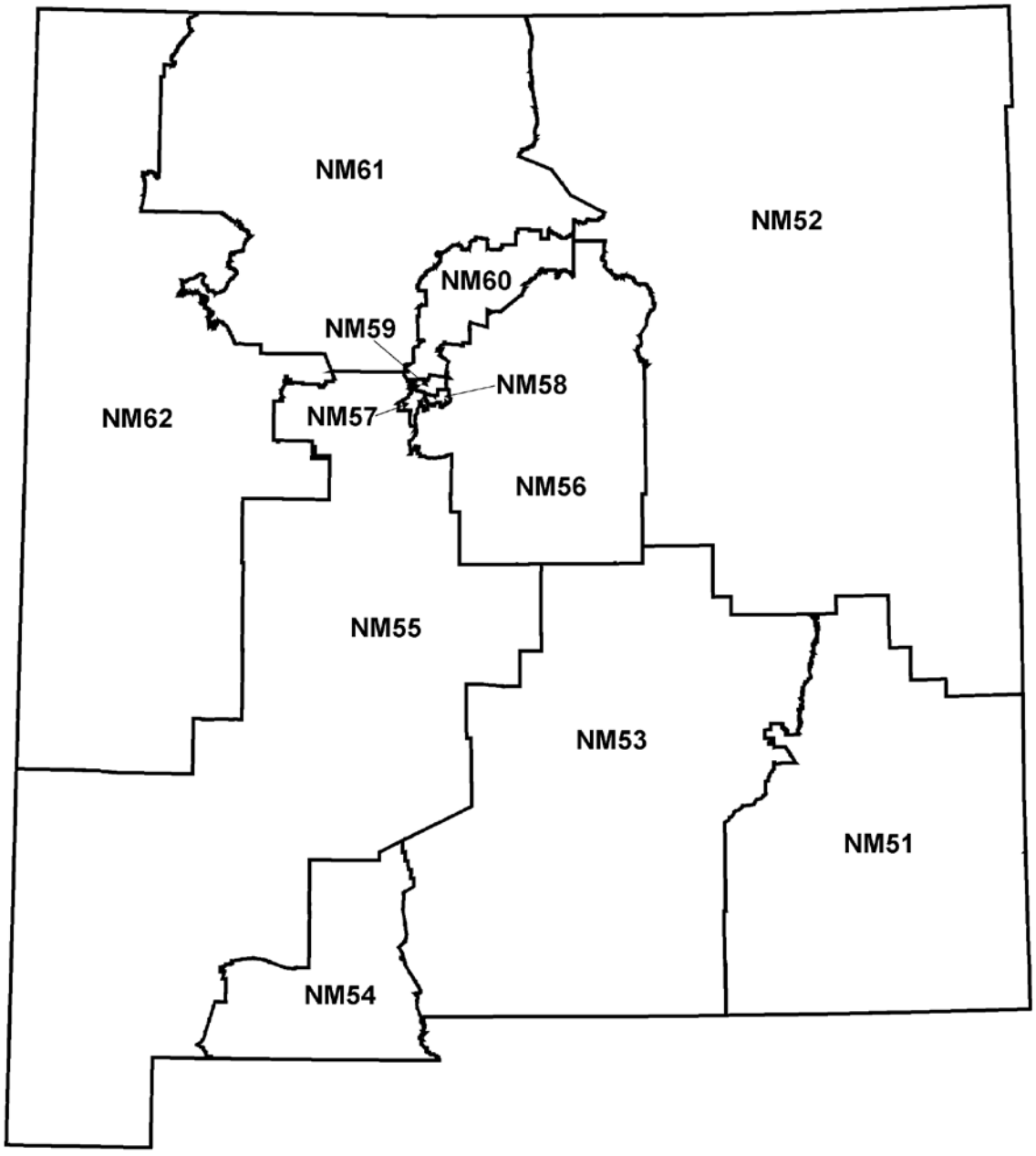
2005-2013 NSDUH State Sampling Regions: New Hampshire

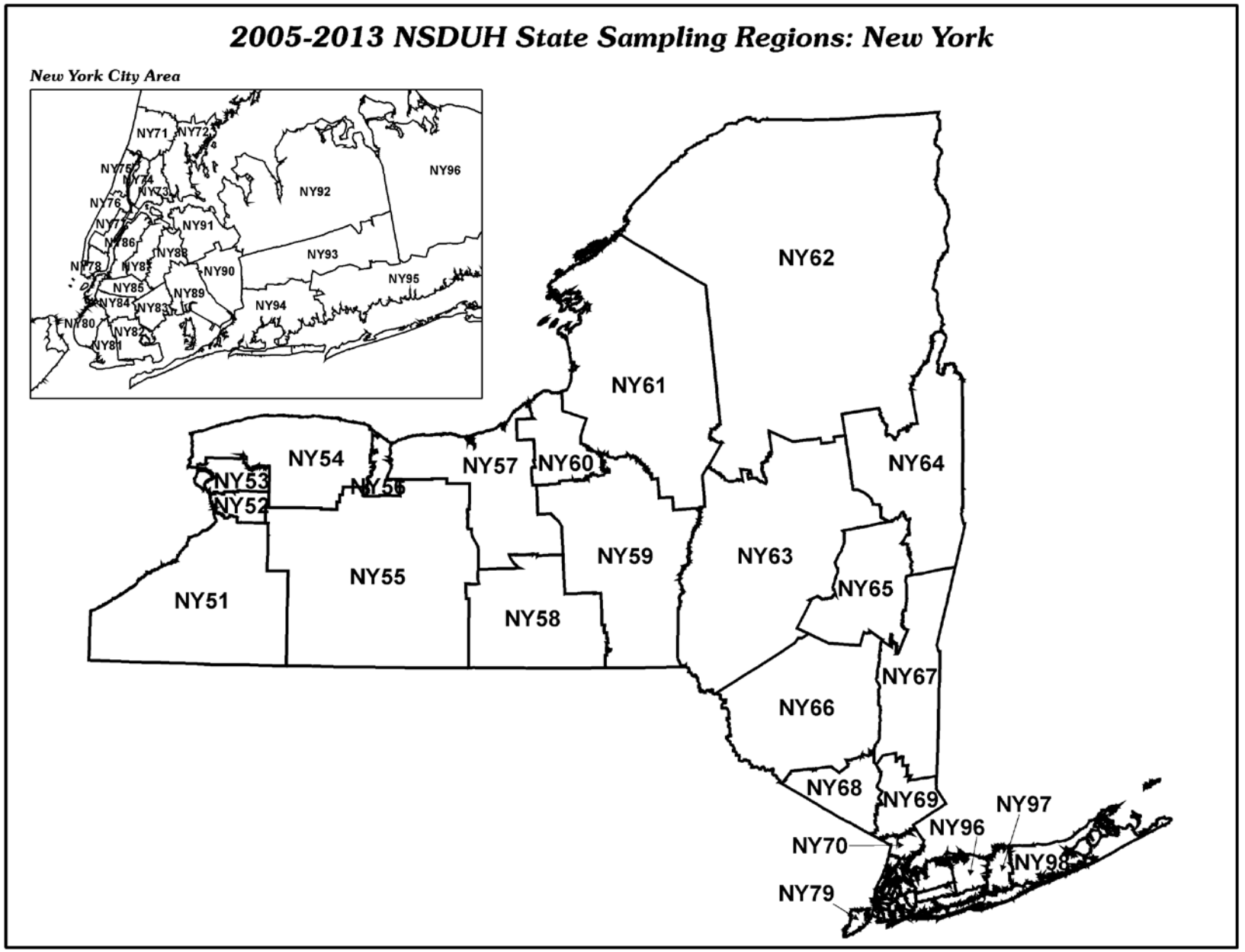


2005-2013 NSDUH State Sampling Regions: New Jersey

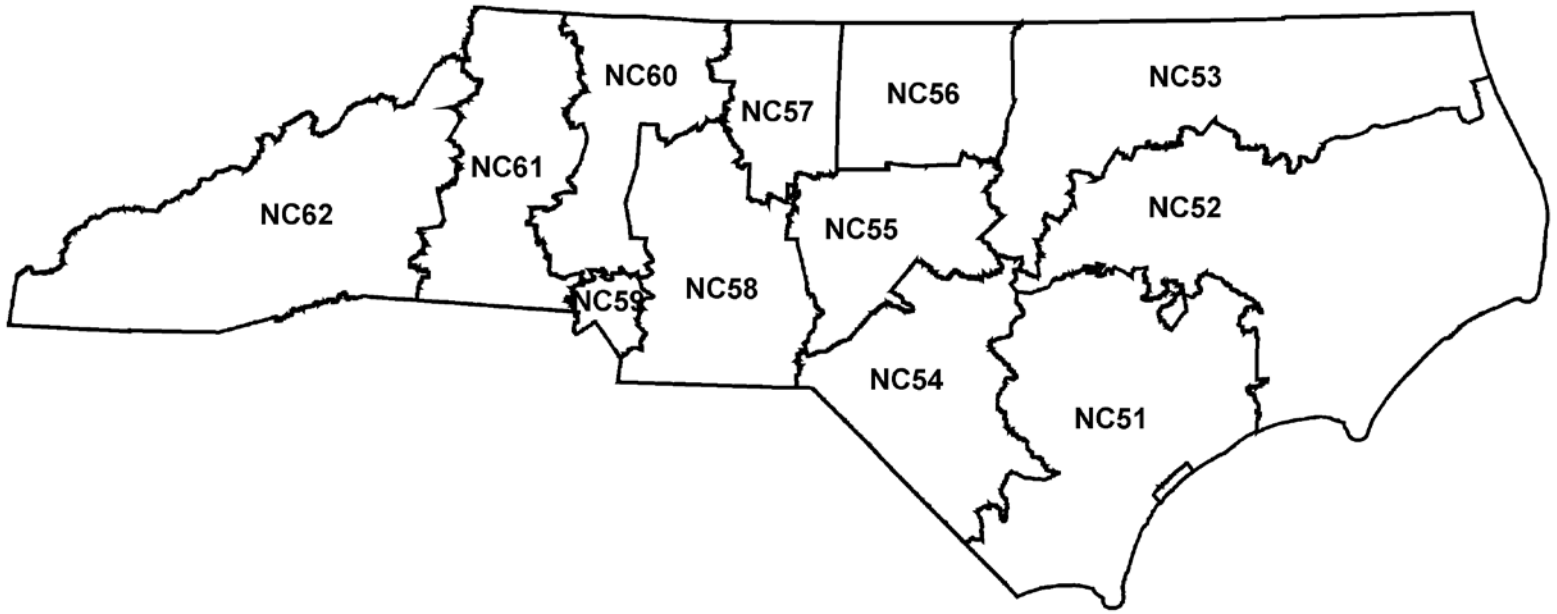


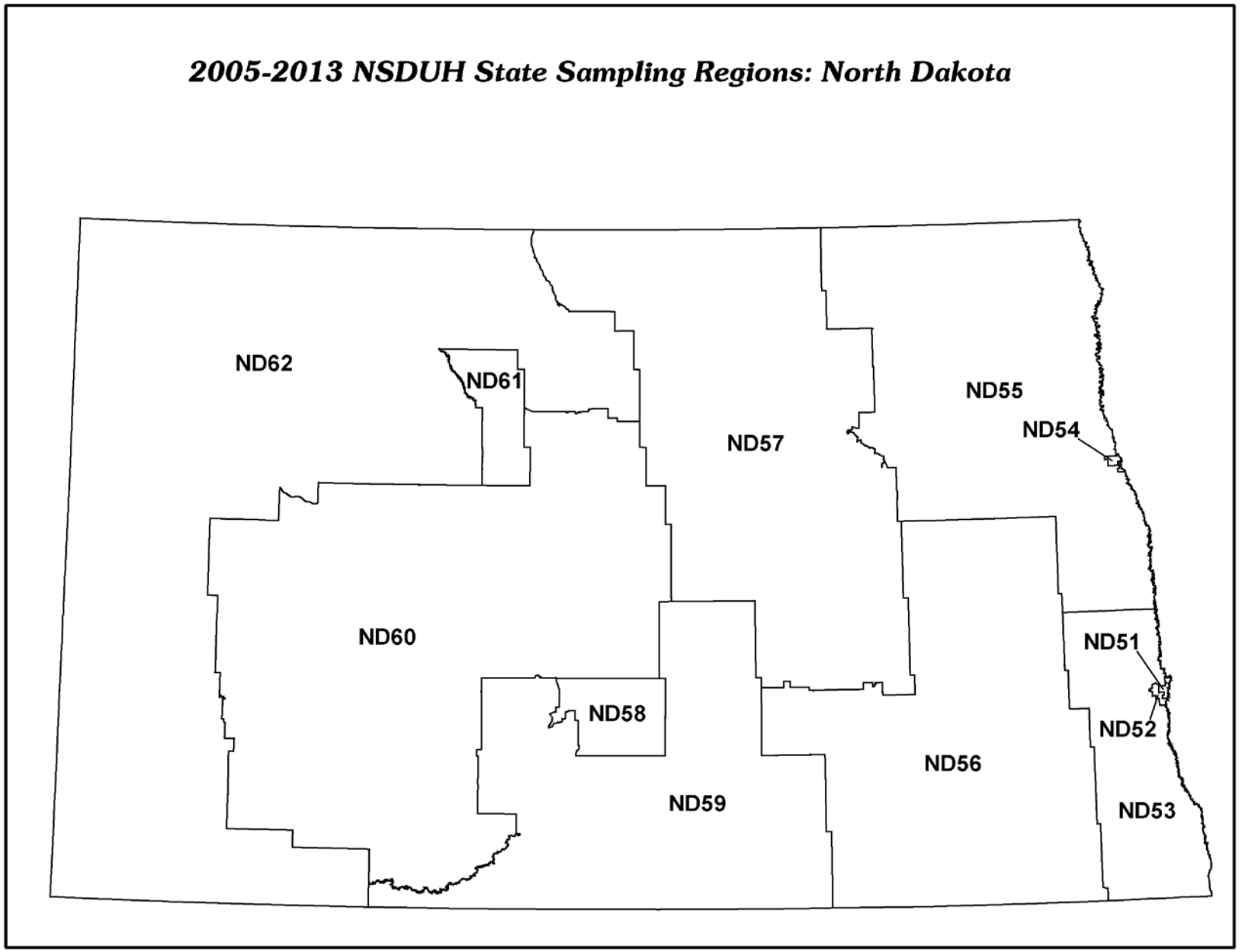
2005-2013 NSDUH State Sampling Regions: New Mexico



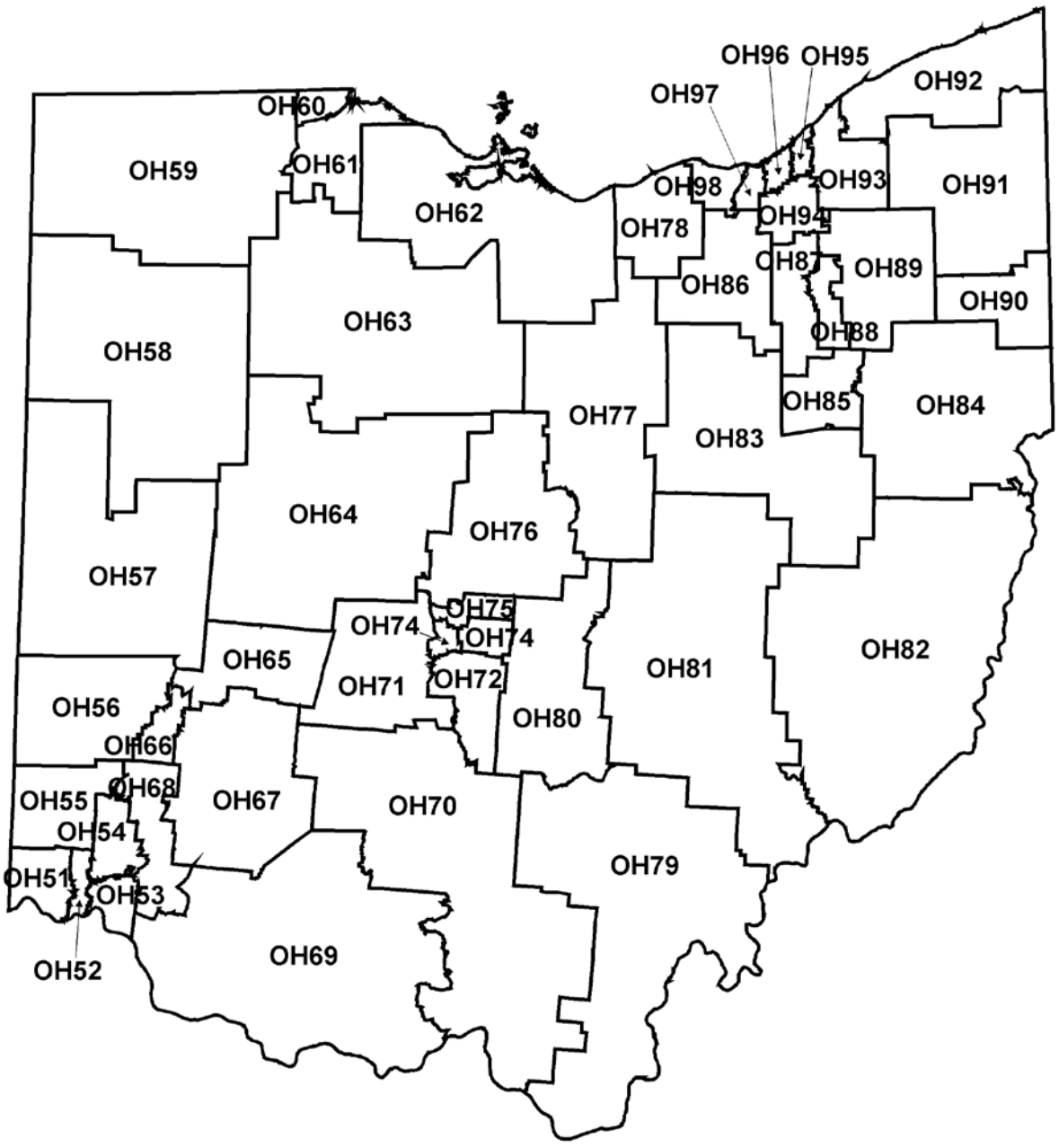

2005-2013 NSDUH State Sampling Regions: New York
New York City Area
NY71
NY72
NY75
NY74
NY73
NY76
NY77
NY91
NY92
NY96
NY86
NY88
NY93
NY78
NY81
NY90
NY95
NY85
NY89
NY84
NY83
NY94
NY80
NY82
NY81
NY62
NY61
NY54
NY57
NY60
NY64
NY53
NY56
NY52
NY59
NY63
NY55
NY65
NY51
NY58
NY67
NY66
NY68
NY69
NY97
NY96
NY70
NY98
NY79

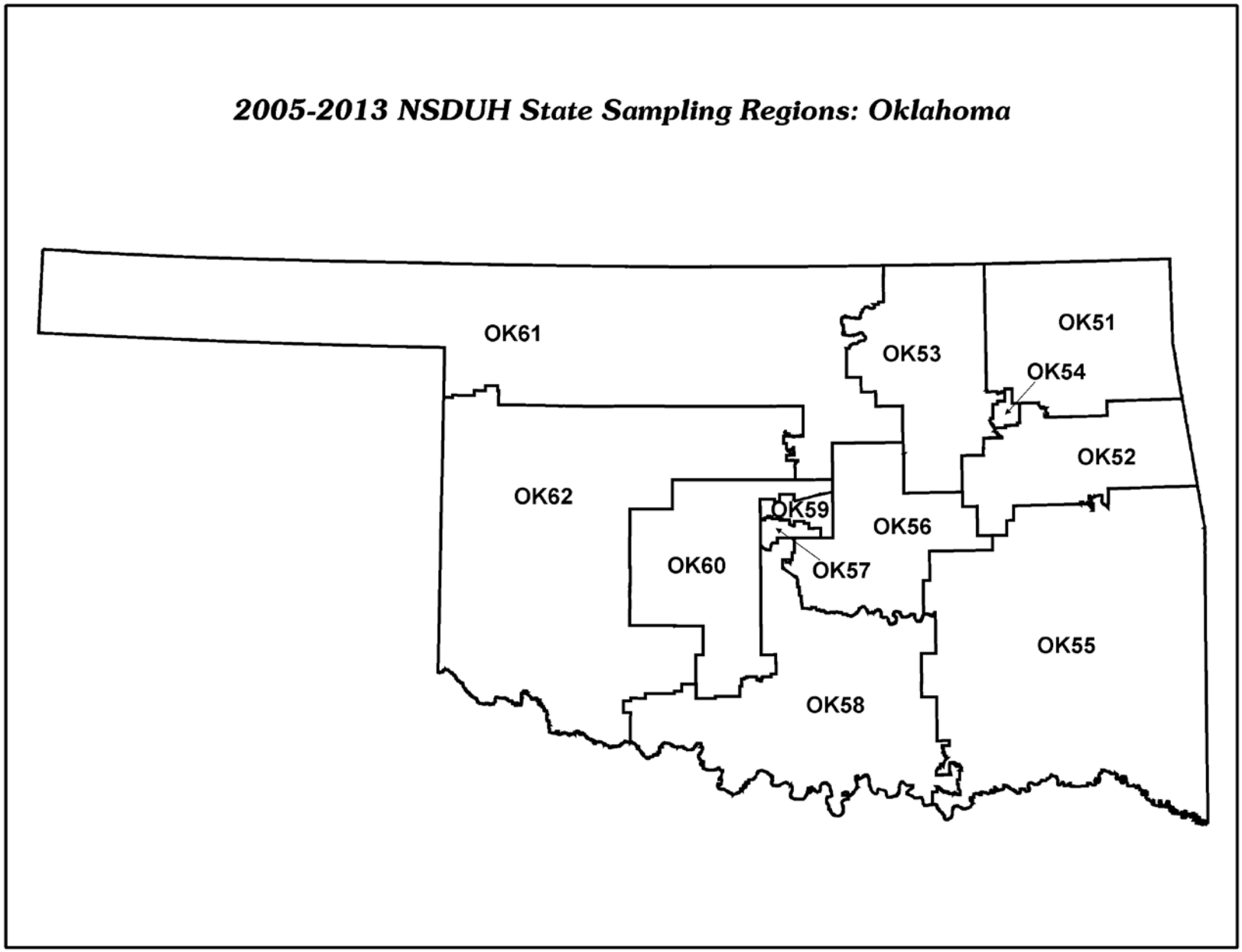
2005-2013 NSDUH State Sampling Regions: North Carolina

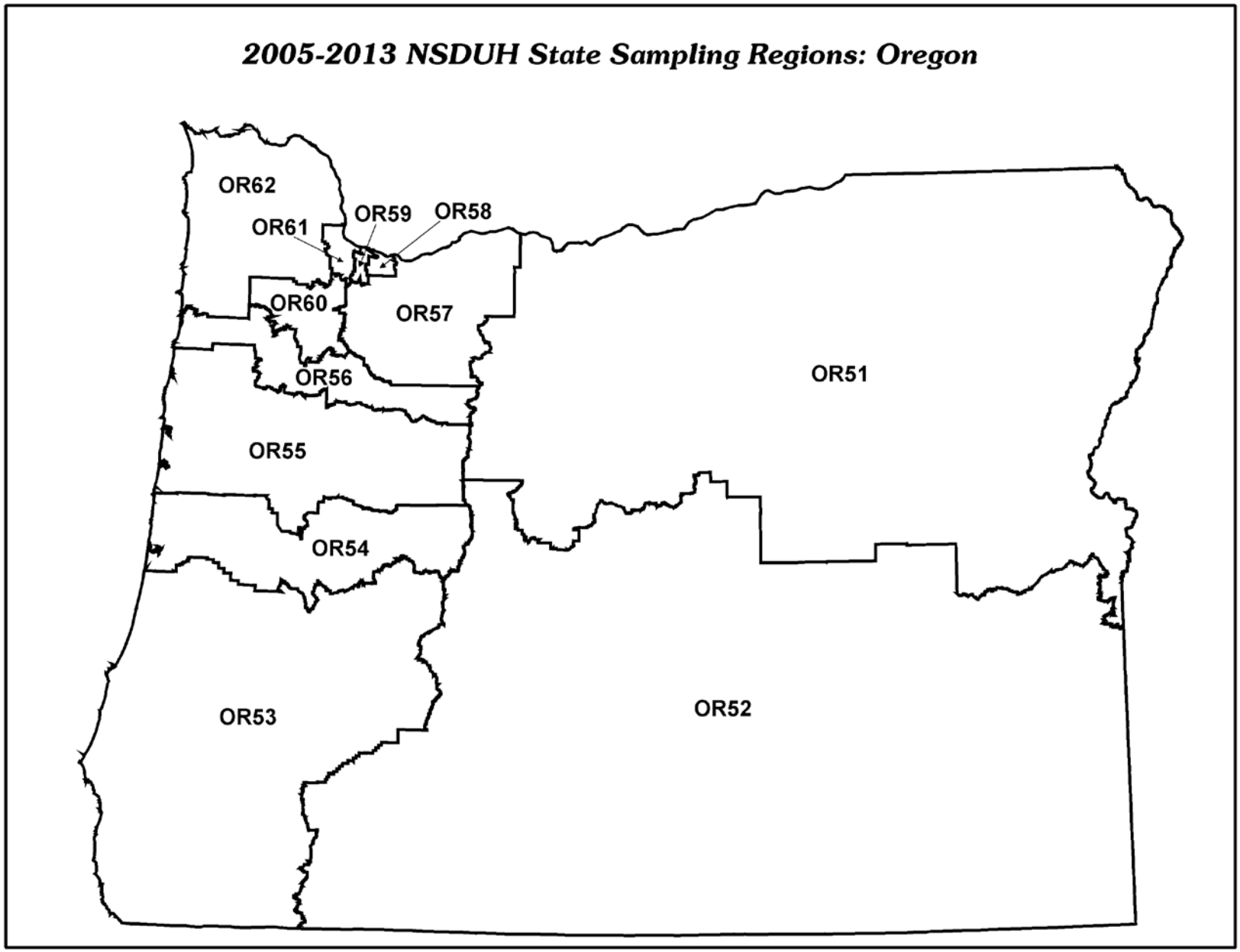


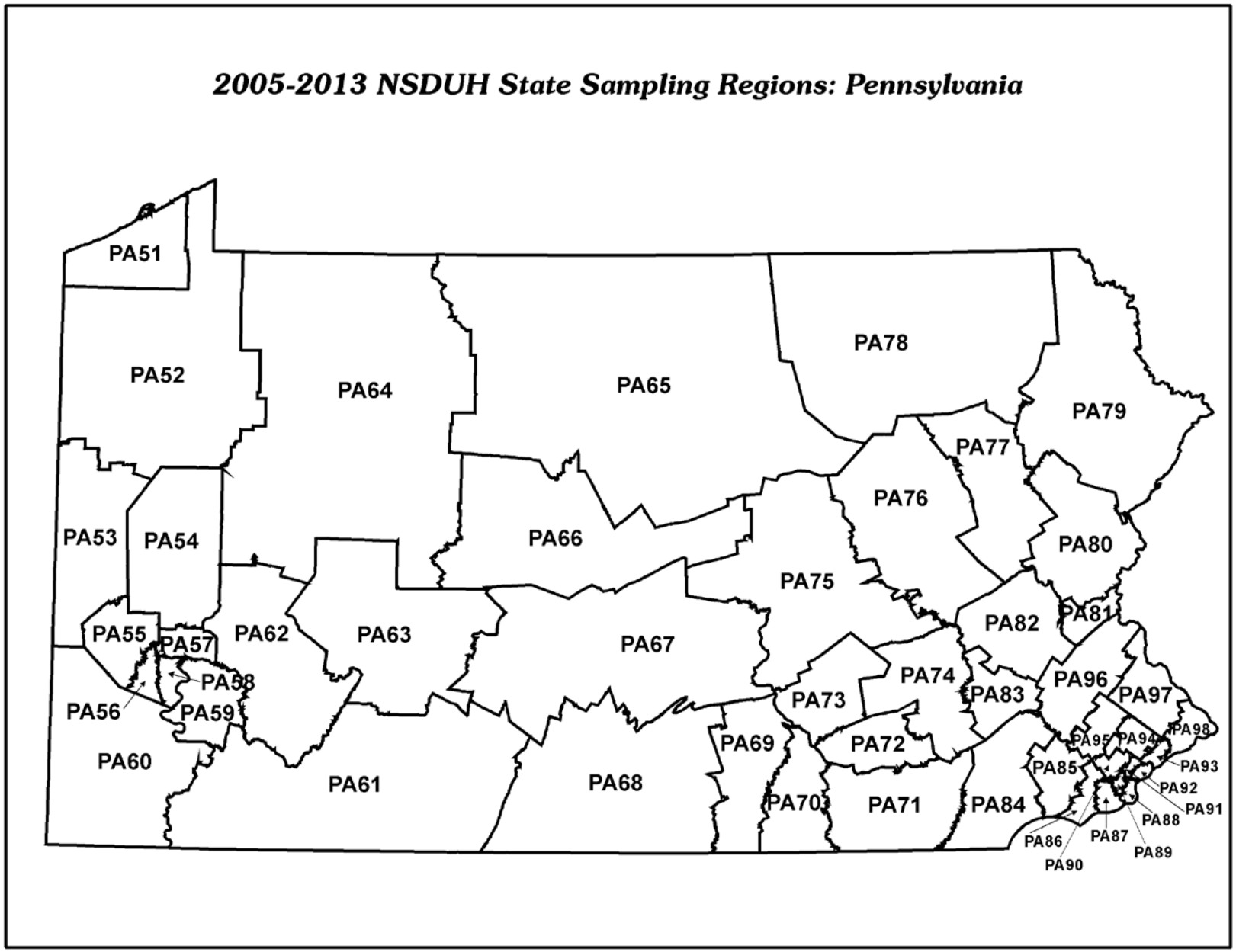

2005-2013 NSDUH State Sampling Regions: North Dakota
ND62
ND61
ND57
ND55
ND54
ND60
ND58
ND59
ND56
ND51
ND52
ND53

2005-2013 NSDUH State Sampling Regions: Ohio

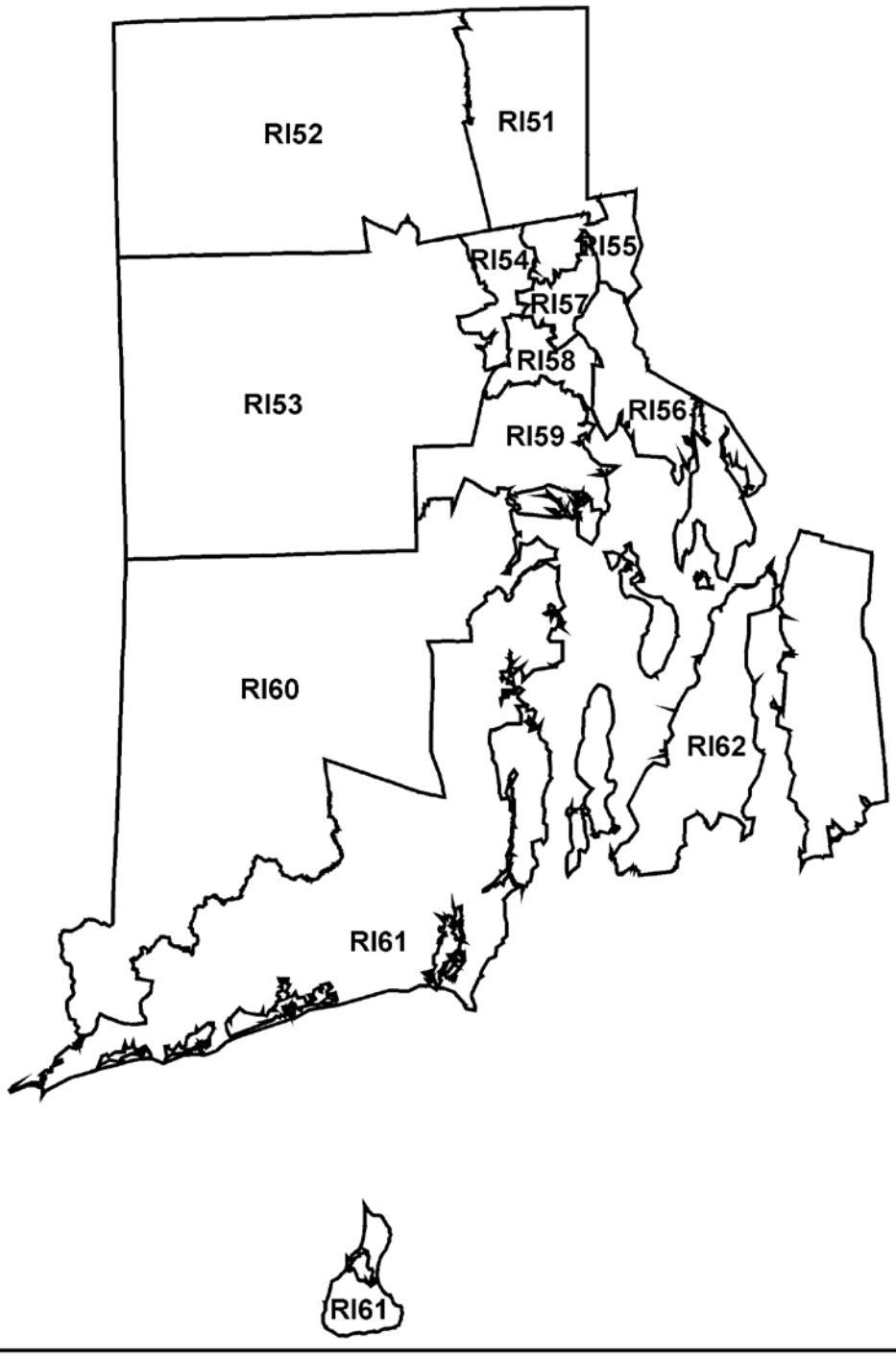



2005-2013 NSDUH State Sampling Regions: Oklahoma
OK61
OK53
OK51
OK54
OK52
OK62
OK59
OK56
OK60
OK57
OK55
OK58

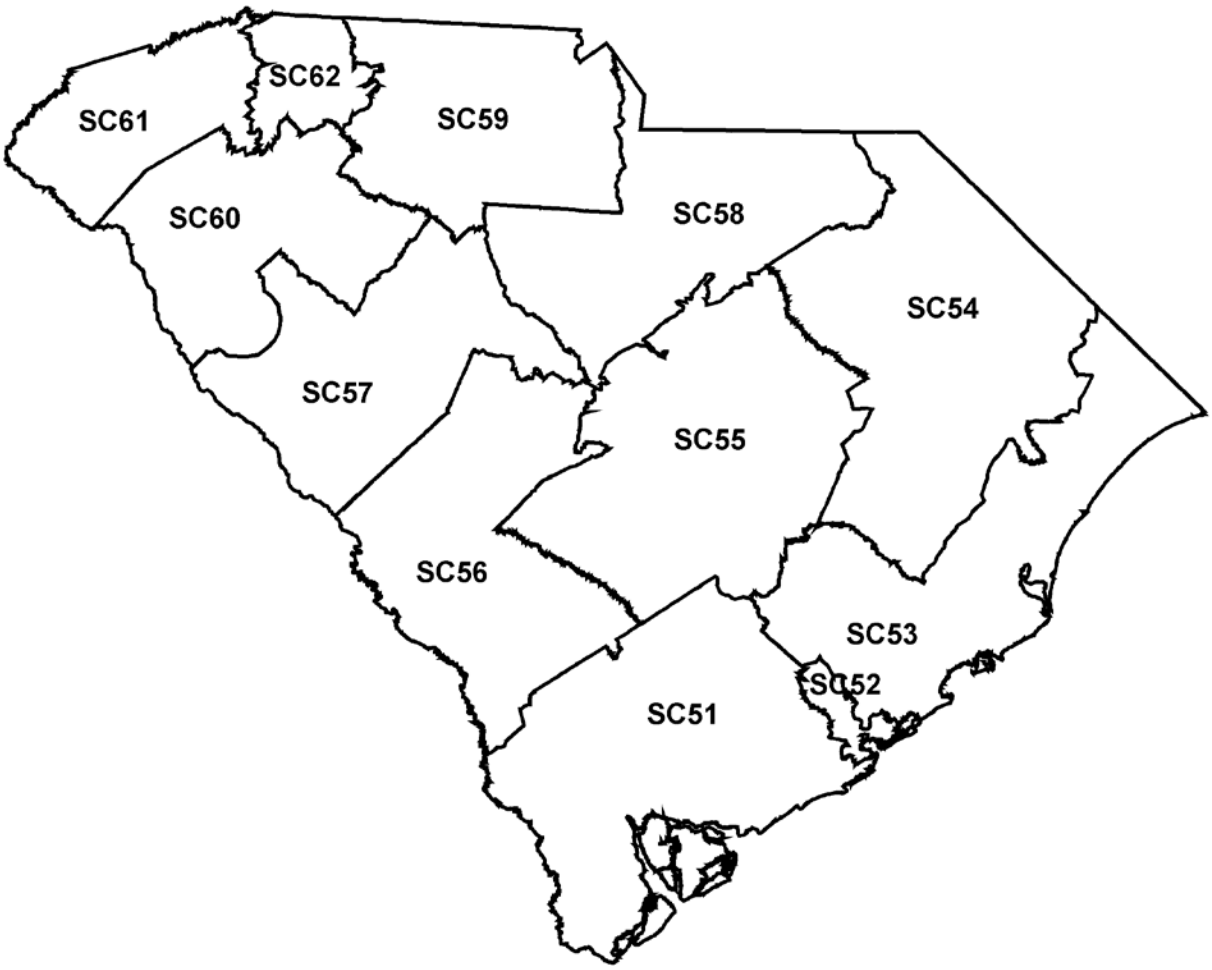

2005-2013 NSDUH State Sampling Regions: Oregon
OR62
OR61
OR59
OR58
OR60
OR57
OR56
OR51
OR55
OR54
OR53
OR52

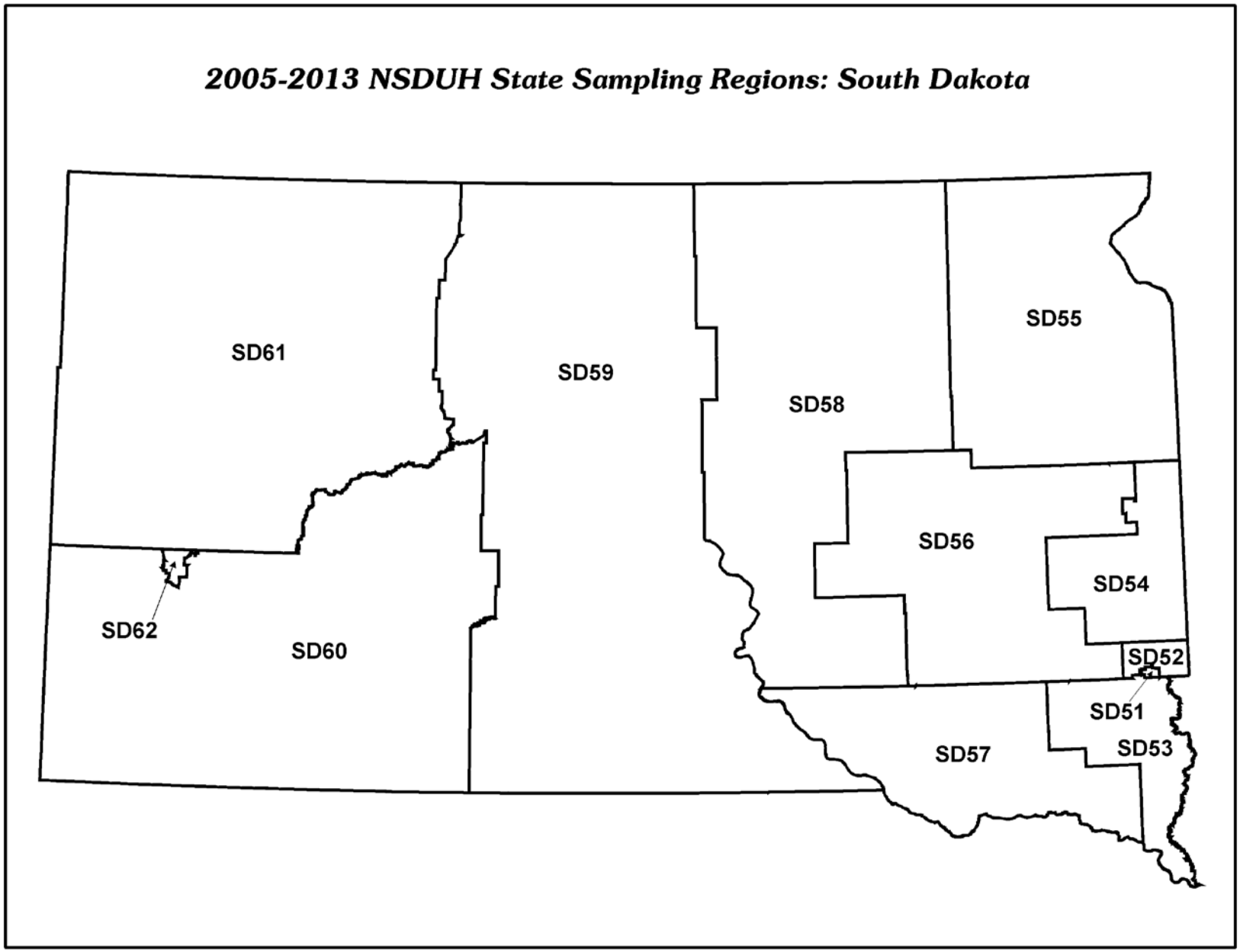

2005-2013 NSDUH State Sampling Regions: Pennsylvania
PA51
PA52
PA64
PA65
PA78
PA79
PA77
PA76
PA80
PA53
PA54
PA66
PA75
PA55
PA57
PA62
PA63
PA67
PA82
PA81
PA58
PA74
PA96
PA97
PA56
PA59
PA73
PA83
PA95
PA94
PA98
PA60
PA69
PA72
PA85
PA93
PA61
PA68
PA70
PA71
PA84
PA92
PA91
PA88
PA86
PA87
PA89
PA90

2005-2013 NSDUH State Sampling Regions: Rhode Island

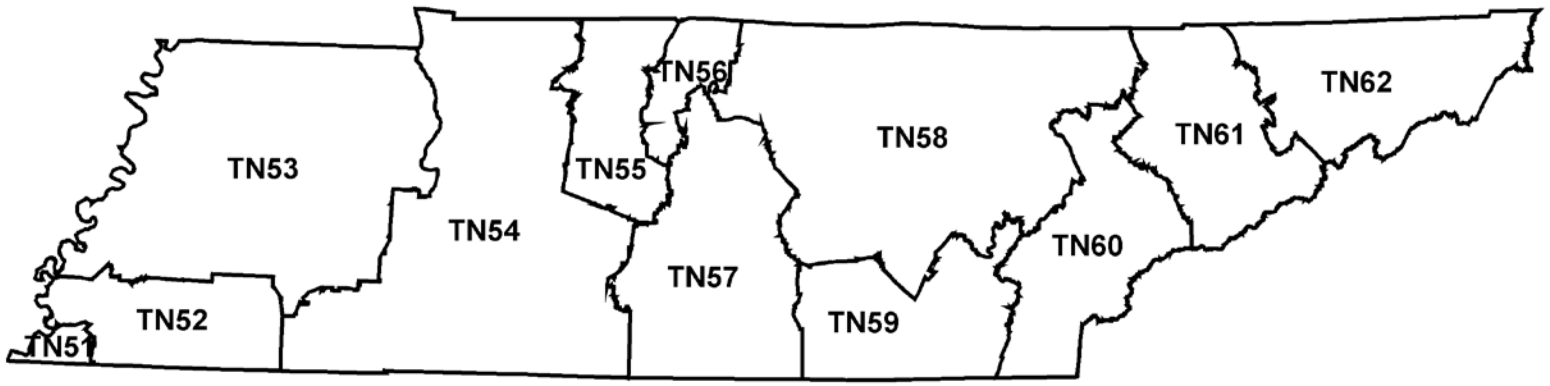


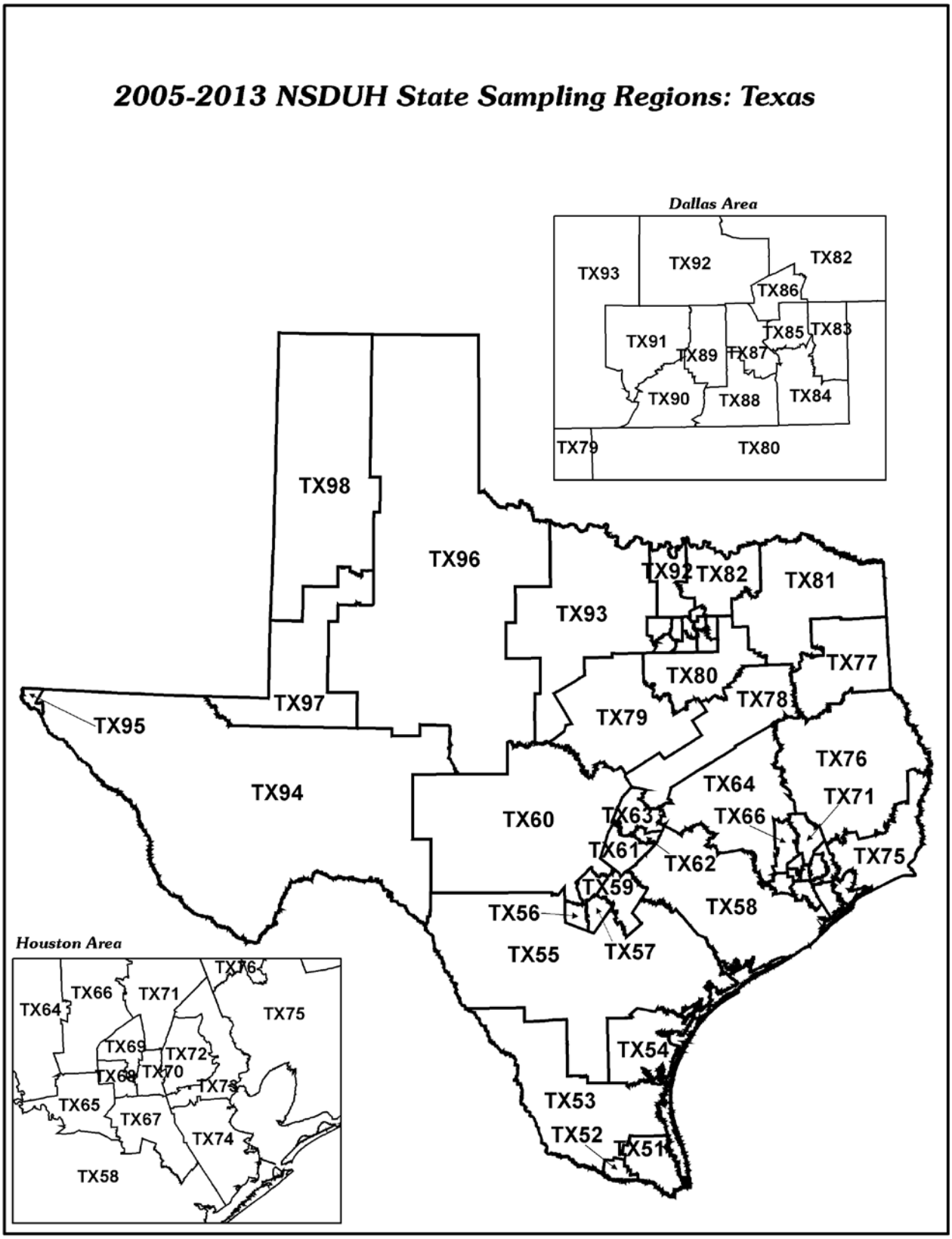
2005-2013 NSDUH State Sampling Regions: South Carolina

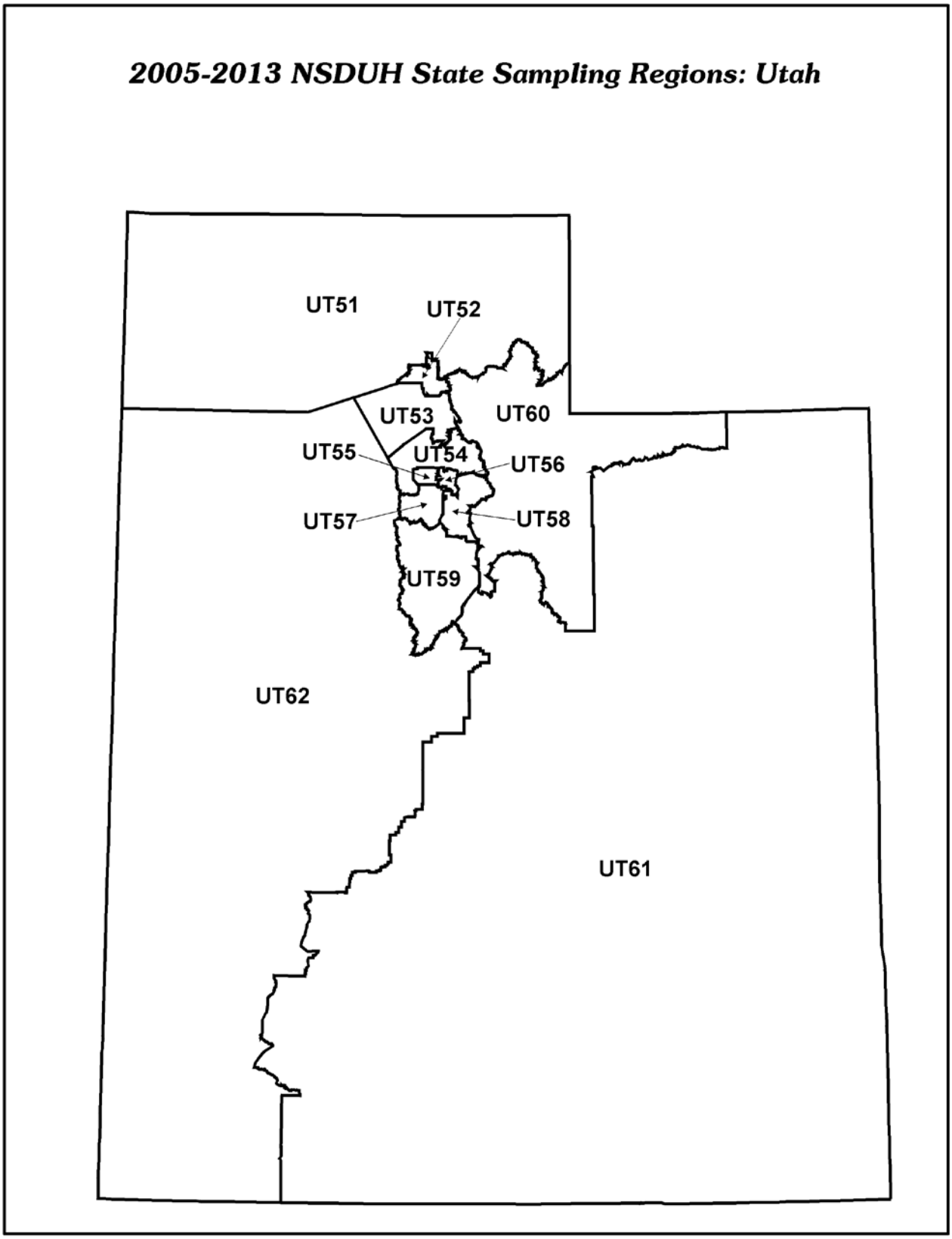



2005-2013 NSDUH State Sampling Regions: South Dakota
SD61
SD59
SD58
SD55
SD56
SD54
SD62
SD60
SD52
SD51
SD53
SD57

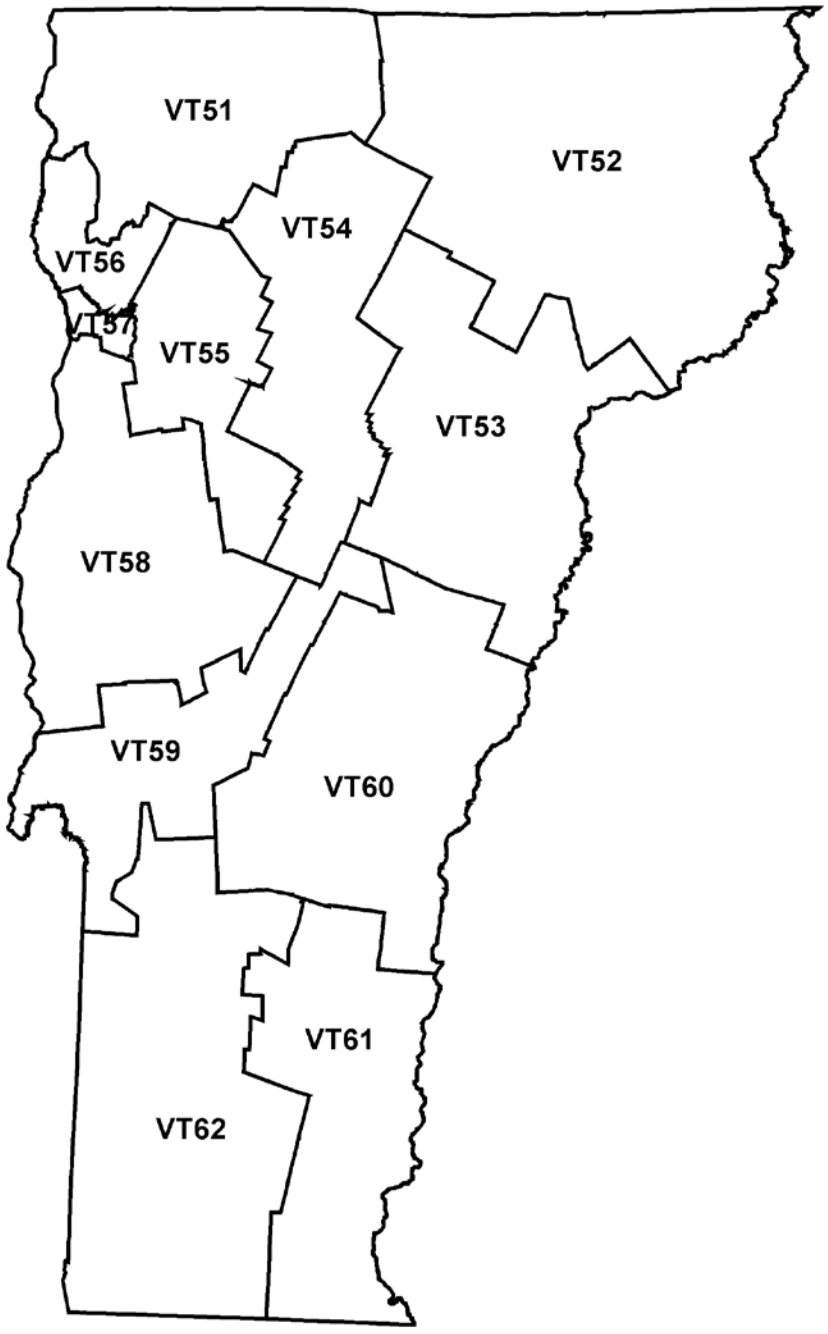
2005-2013 NSDUH State Sampling Regions: Tennessee

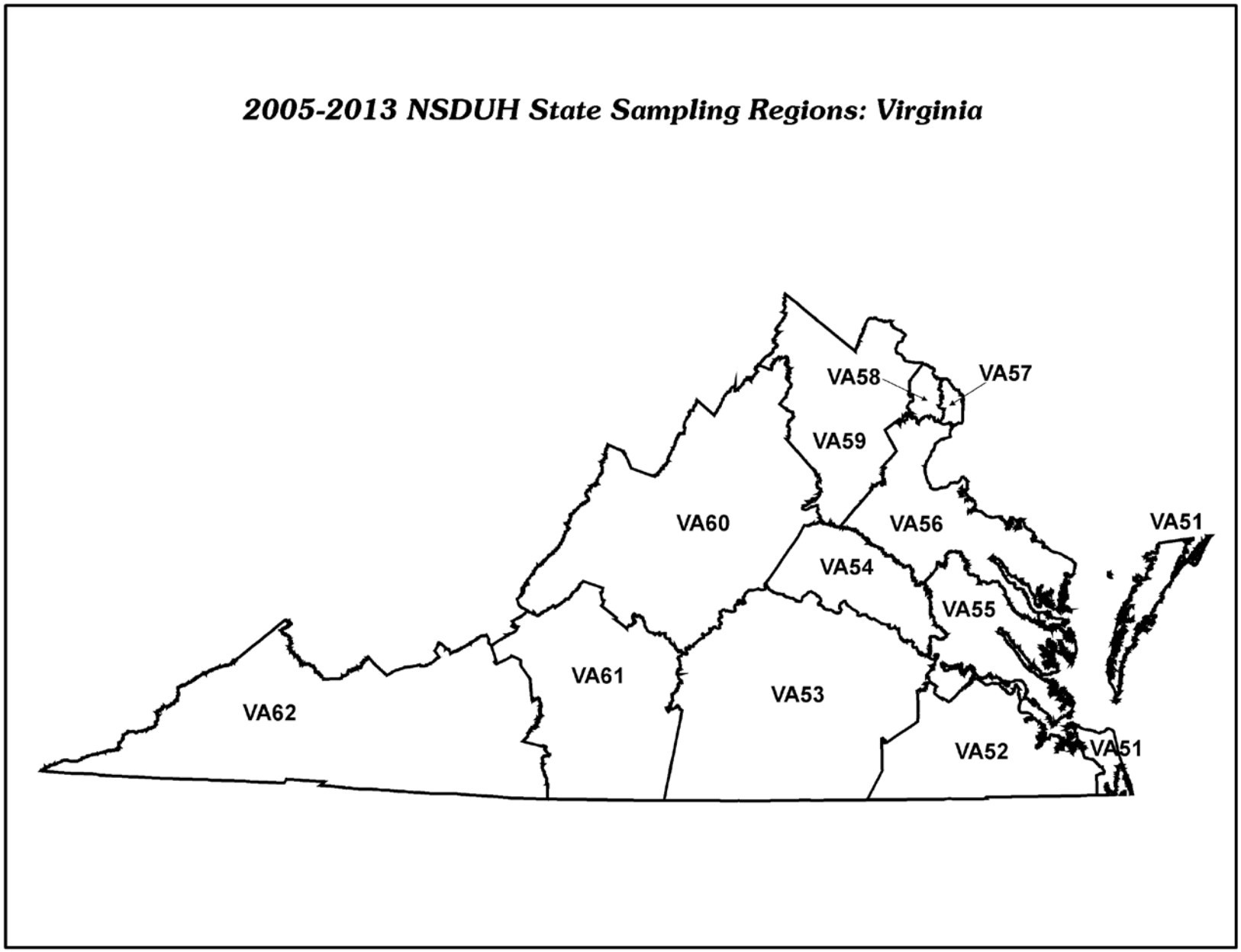


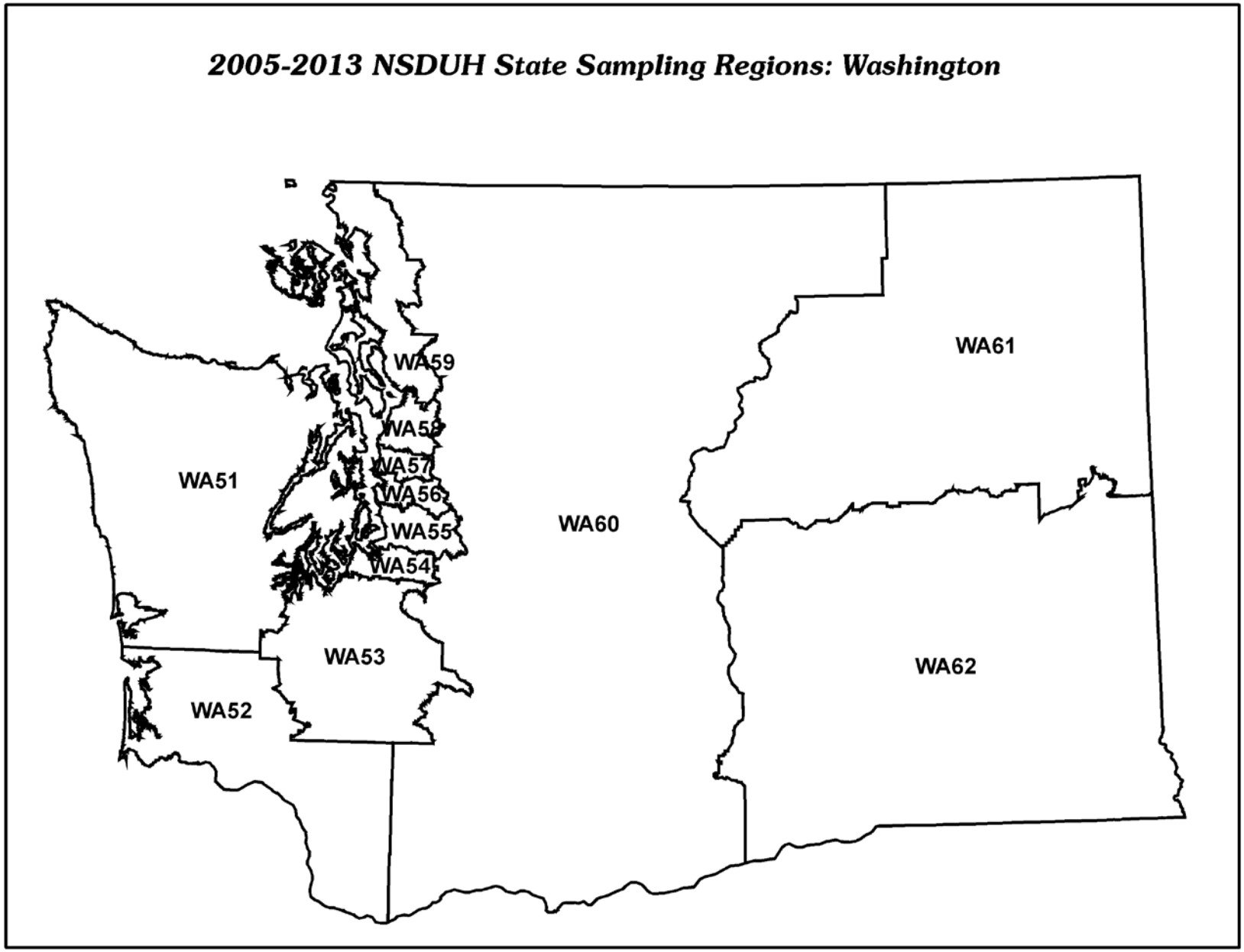

2005-2013 NSDUH State Sampling Regions: Texas
Dallas Area
TX93
TX92
TX82
TX86
TX91
TX85
TX83
TX89
TX87
TX90
TX88
TX84
TX79
TX80
TX98
TX96
TX92
TX82
TX81
TX93
TX80
TX77
TX97
TX78
TX95
TX79
TX76
TX94
TX64
TX71
TX60
TX63
TX66
TX61
TX62
TX75
TX59
TX56
TX58
TX55
TX57
Houston Area
TX76
TX66
TX71
TX64
TX75
TX69
TX72
TX68
TX70
TX73
TX65
TX67
TX74
TX58
TX54
TX53
TX52
TX51

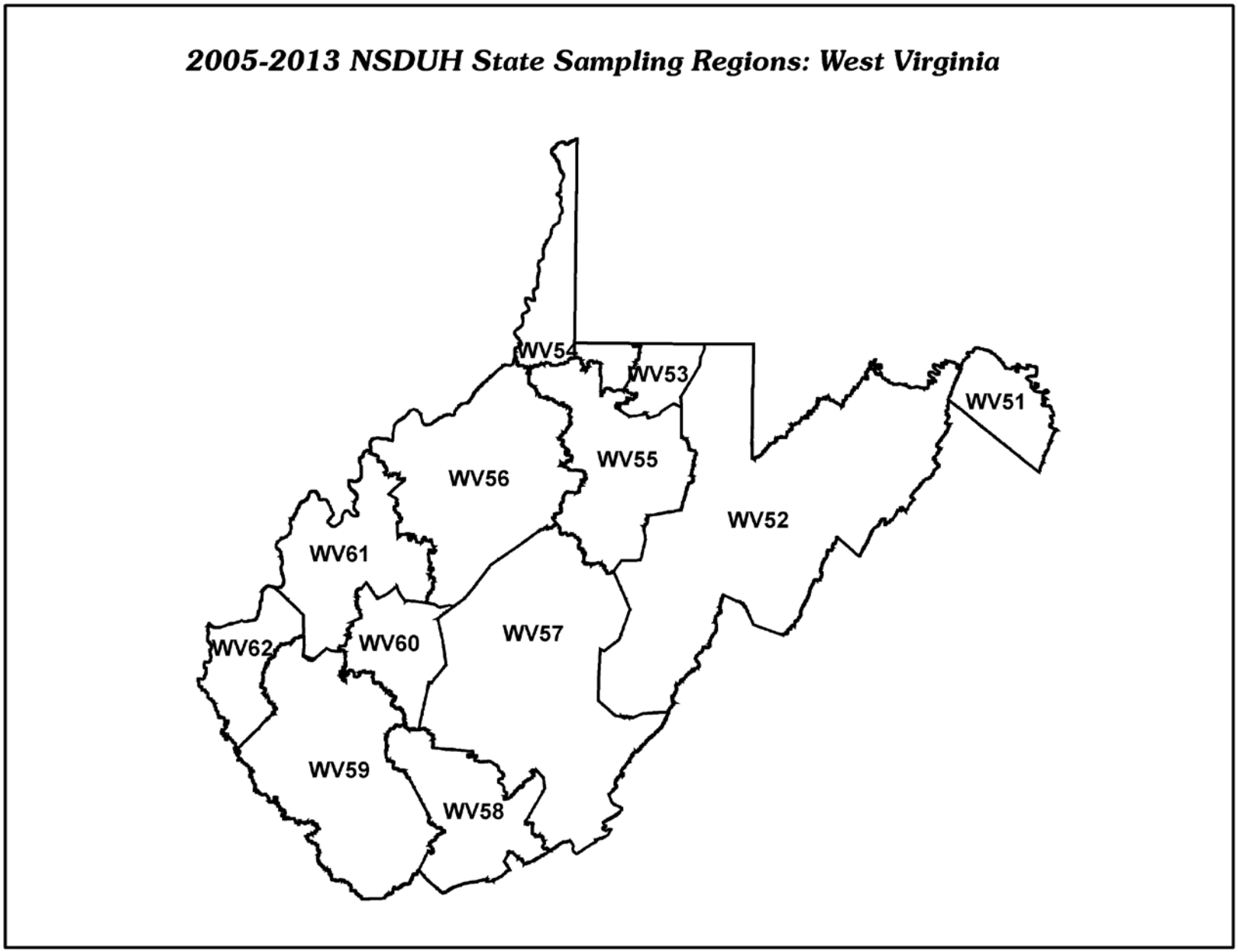

2005-2013 NSDUH State Sampling Regions: Utah
UT51
UT52
UT53
UT60
UT55
UT54
UT56
UT57
UT58
UT59
UT62
UT61

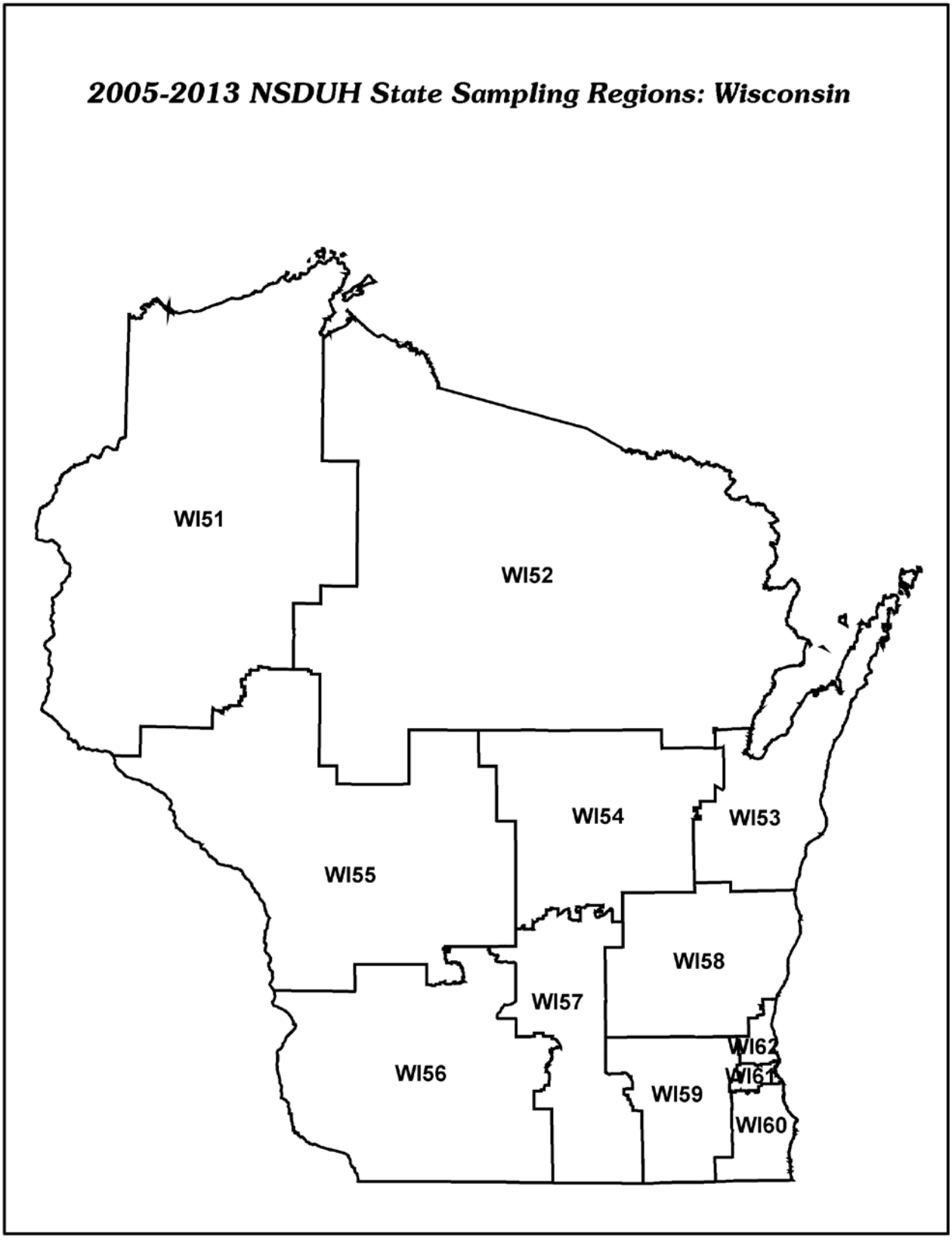
2005-2013 NSDUH State Sampling Regions: Vermont

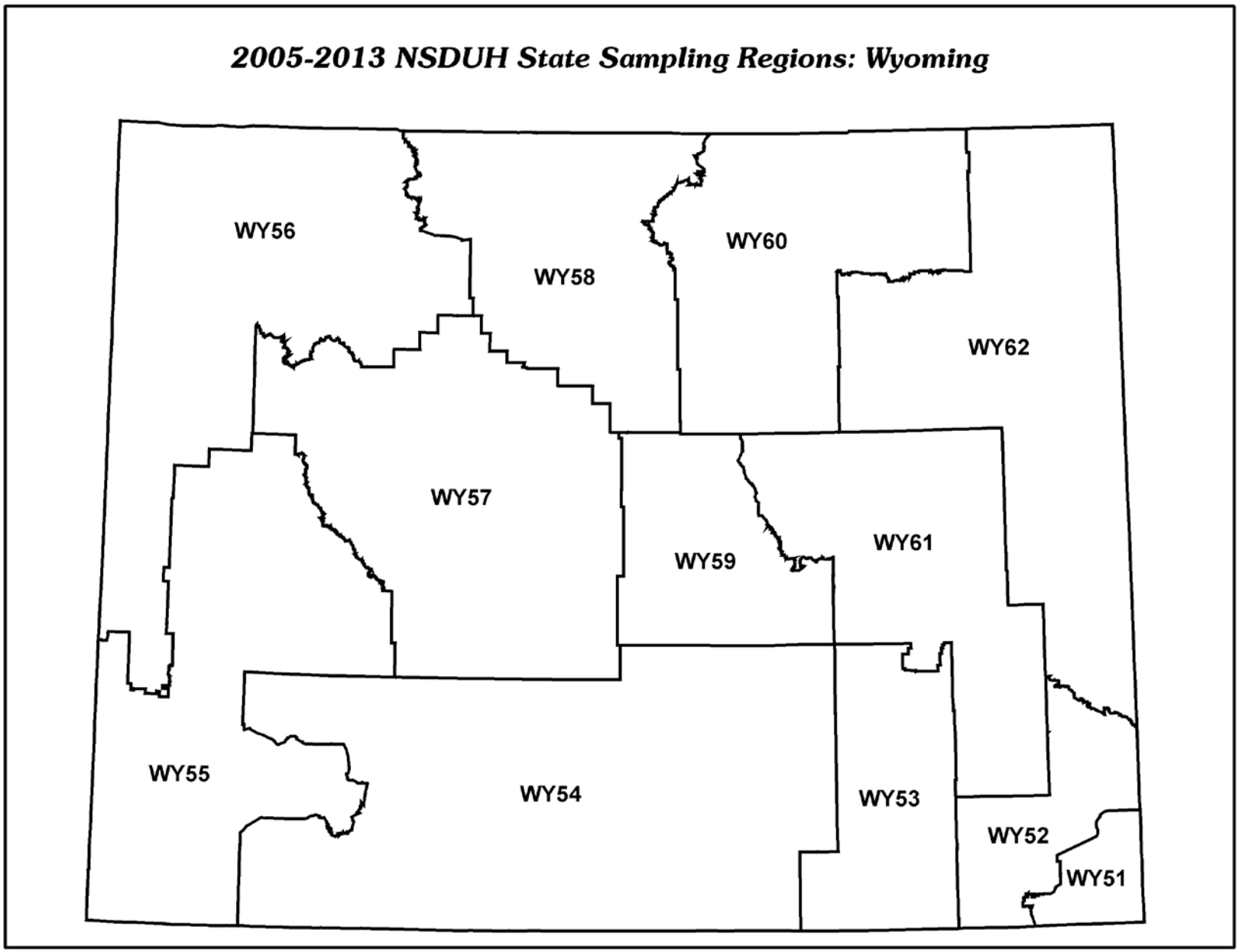



2005-2013 NSDUH State Sampling Regions: Virginia
VA58
VA57
VA59
VA60
VA56
VA51
VA54
VA55
VA61
VA53
VA62
VA52
VA51


2005-2013 NSDUH State Sampling Regions: Washington
WA59
WA58
WA57
WA56
WA55
WA54
WA51
WA60
WA61
WA53
WA52
WA62


2005-2013 NSDUH State Sampling Regions: West Virginia
WV54
WV53
WV51
WV55
WV56
WV52
WV61
WV57
WV62
WV60
WV59
WV58


2005-2013 NSDUH State Sampling Regions: Wisconsin
WI51
WI52
WI54
WI53
WI55
WI58
WI57
WI62
WI56
WI61
WI59
WI60


2005-2013 NSDUH State Sampling Regions: Wyoming
WY56
WY58
WY60
WY62
WY57
WY59
WY61
WY55
WY54
WY53
WY52
WY51

Appendix B: Expected Value of the Collapsed Stratum Estimator as Applied to the NSDUH "With Replacement" Variance Estimator

The National Survey on Drug Use and Health (NSDUH) design uses 900 State sampling regions (SSRs) as geographically defined sampling strata within States. In addition the annual sample of eight area segments in each SSR is randomly assigned to four quarters (two segments per quarter). One half of the segments (one per quarter in each SSR) are replaced each year with a fresh subset of a combined 5-year sample design.

With this structure, maximum stratification at the first stage is recognized by defining strata as the combination of SSR and quarter yielding 3,600 (900 SSRs * 4 quarters) first-stage strata with two replicates (area segments in each stratum). This approach provides 3,600 degrees of freedom (*df*) for variance estimation for national estimates, 48 *df* for small State estimates, and 192 *df* for large State estimates. The problem with applying this approach is that many segments can be anticipated to have no observations because of the combined effects of ineligibility, low sampling yields at the person level, and nonresponse at the dwelling unit (DU) or person levels. This problem was resolved in the 1999 to 2004 design by collapsing strata (and replicates) across quarters yielding 900 national variance estimation strata, 12 variance estimation strata for small sample States, and 48 variance estimation strata for large sample States. Under the NSDUH design, 1 *df* for variance estimation is associated with each variance estimation stratum.

For the 2005 through 2013 design, an alternate stratum-collapsing strategy was defined that has the combined effects of maintaining adequate degrees of freedom for national estimates (close to 900) and also obtaining higher degrees of freedom for State-level estimates (48 *df* for small sample States and 192 *df* for large sample states). This discussion is intended to show that any arbitrary grouping of sampling strata can be used to achieve variance estimators with the same expected values. This result suggests that instead of forming variance estimation strata across quarters within SSRs, it is equally feasible to form variance estimation strata across SSRs. In addition, if the SSRs that were combined to form a variance estimation stratum come from different States, they provide some additional disclosure protection because an intruder can no longer assume that all respondents in a variance estimation stratum come from the same State.

Consider a total defined in terms of the sample design structure as

$$T_Y = \sum_{h=1}^{3,600} \sum_{i=1}^{N_h} \sum_{j=1}^{N_{hi}} Y_{hij} ,$$

where Y_{hij} is a numeric characteristic of the j-th person in the i-th area segment of the h-th stratum, N_{hi} is the number of NSDUH-eligible persons in the i-th area segment of the h-th stratum, and N_h is the number of area segments defined within the h-th stratum. The NSDUH annual sample design calls for selecting two area segments from each of the 3,600 strata and a variable number of persons, n_{hi}, per area segment. The total sample of person is targeted at 67,500 for the 2013 sample or an average of 9.375 responding persons per segment.

An estimate of the population total can then be written in terms of the observed sample as

$$\hat{T}_Y = \sum_{h=1}^{3,600} \sum_{i=1}^{2} \sum_{j=1}^{n_{hi}} w_{hij} y_{hij} ,$$

where y_{hij} is the observed numeric characteristic of the j-th sample person in the i-th sample area segment of the h-th stratum, w_{hij} is the analytic weight of this person, and n_{hi} is the number of sampled and responding NSDUH-eligible persons in the i-th area segment of the h-th stratum.

Because the NSDUH first-stage sampling rate is low, the "with replacement" variance estimation option provides a nearly unbiased variance estimate for NSDUH estimates.[24] Following the notation in the SUDAAN manual (RTI International, 2012b), the variance estimate based on the 3,600 strata can be written as

$$v(\hat{T}_Y) = \sum_{h=1}^{3,600} \sum_{i=1}^{2} 2(z_{hi} - \bar{z}_h)^2,$$

where $z_{hi} = \sum_{j=1}^{n_{hi}} w_{hij} y_{hij}$, and $\bar{z}_h = \frac{\sum_{i=1}^{2} z_{hi}}{2}$.

Suppose someone wishes to collapse the 3,600 strata into K (< 3,600) strata, each containing H_k of the original strata, and such that $\sum_{k=1}^{K} H_k = 3,600.$ In addition, the replicates within these strata consist of the combined replicate 1 segments and combined replicate 2 segments from the contributing original strata. Then the variance of a total can be estimated on the collapsed strata as

$$v'(\hat{T}_Y) = \sum_{k=1}^{K} \sum_{i=1}^{2} 2(z'_{ki} - \bar{z}'_k)^2,$$

where $z'_{ki} = \sum_{h \in k}^{H_k} z_{hi}$, and $\bar{z}'_k = \frac{\sum_{i=1}^{2} z'_{ki}}{2} = \frac{\sum_{i=1}^{2} \sum_{h \in k}^{H_k} z_{hi}}{2} = \sum_{h \in k}^{H_k} \bar{z}_h$. Notice that $z'_{ki} - \bar{z}'_k = \sum_{h \in k}^{H_k} (z_{hi} - \bar{z}_h)$.

To show the equivalence of collapsed stratum variance estimate to the full stratum variance, the collapsed stratum variance can be re-expressed as

$$v'(\hat{T}_Y) = \sum_{k=1}^{K} \sum_{i=1}^{2} 2\{\sum_{h \in k}^{H_k} (z_{hi} - \bar{z}_h)\}^2 = \sum_{k=1}^{K} \sum_{i=1}^{2} 2 \sum_{h \in k}^{H_k} (z_{hi} - \bar{z}_h)^2 + \sum_{k=1}^{K} \sum_{i=1}^{2} 2 \sum_{h \neq h' \in k}^{H_k} (z_{hi} - \bar{z}_h)(z_{h'i} - \bar{z}_{h'}) .$$

[24] The assumption of "with replacement" sampling produces estimates of variance that are slightly biased on the high side because they do not take account of variance reduction due to finite population sampling at the first stage of the design.

The first term can be shown to have the same expectation over repeated sampling by rearranging the summation as

$$\sum_{k=1}^{K}\sum_{i=1}^{2}2\sum_{h\in k}^{H_k}(z_{hi}-\overline{z}_h)^2=\sum_{k=1}^{K}\sum_{h\in k}^{H_k}\sum_{i=1}^{2}2(z_{hi}-\overline{z}_h)^2=\sum_{h=1}^{3,600}\sum_{i=1}^{2}2(z_{hi}-\overline{z}_h)^2\ .$$

The second term has expectation zero because sample selection is done independently in each of the original strata. Because this second term is 0 only in expectation, the exact values of the two variance estimates are not likely to be identical.

Appendix C: 2013 NSDUH Procedures for Subsegmenting

C.1 Introduction

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

C.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).[25] The two criteria for identifying candidate segments are as follows:

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

C.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 C.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.

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

C.4 Standard Subsegmenting Procedures

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

Step 1: On the basis of the count, the segment is divided into areas (list units) containing 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 D: 2013 NSDUH Procedures for Adding Missed Dwelling Units

D.1 Introduction

The 2013 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 2012. Because the listing was done a short time before the 2013 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 in place two procedures for locating and adding missed DUs. First, during the screening interview, FIs ask the screening respondent about other units on the property of the selected DU. Next, FIs look between each selected DU and the next listed DU (i.e., the half-open interval [HOI] rule, as described in Section 3.7 of Chapter 3) to identify any unlisted units. In 2000, the rule was modified such that the HOI 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. Any missed DUs found on the property or in the geographic interval (i.e., the HOI) following a selected DU are considered "linked" to that DU. 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."

D.2 Motivation

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

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

D.3 Procedure for Adding Housing Units

This section refers to HUs that are obtained through the missed DU procedures. 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 D.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 D.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 HOI 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 missed DU procedures—not for all missed DUs found in the segment. The link number then will be entered and a line number will be assigned.[26] 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.

[26] 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.

D.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 HOI 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 D.6).

D.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.

D.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.

D.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:

Unit Number	SLN
12	**1**
13	2
14	3
15	4
16	**5**
17	6
18	7
19	8
20	9
21	**10**
22	11
23	**1**
24	2
25	3
26	4
27	**5**
28	6
29	7
30	8
31	9
32	**10**
33	11
34	**1**
35	2
36	3
37	4
38	**5**
39	6
40	7
41	8
42	9
43	**10**
44	11
45	**1**

D.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 DU) 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 D.3.

D.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.