2016 NATIONAL SURVEY ON DRUG USE AND HEALTH

METHODOLOGICAL RESOURCE BOOK SECTION 2: SAMPLE DESIGN REPORT

DISCLAIMER

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

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

January 2017

This page intentionally left blank

2016 NATIONAL SURVEY ON DRUG USE AND HEALTH: SAMPLE DESIGN REPORT

Prepared for the 2016 Methodological Resource Book (Section 2)

Contract No. HHSS283201300001C
RTI Project No. 0213985.005.103.001
Deliverable No. 18

RTI Authors:
Katherine B. Morton
Jeremy Aldworth
Erica L. Hirsch
Peilan C. Martin

SAMHSA Authors:
Matthew Williams
Arthur Hughes

Project Director:
David Hunter

SAMHSA Project Officer:
Peter Tice

For questions about this report, please e-mail 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

January 2017

Recommended Citation: Center for Behavioral Health Statistics and Quality. (2017). *2016 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 registered trademark and a trade name of Research Triangle Institute). Significant contributors besides the authors at RTI include (in alphabetical order) Bernadette Chasteen, Patrick Chen, David Chrest, Peter Frechtel, Rachel Harter, Andrew Haynes, Phillip Kott, Susan Myers, and Jesse Pegg. Also, Richard S. Straw, Debbie Bond, and Teresa F. Bass edited the report, formatted it, and produced its web version, respectively.

Table of Contents

Chapter **Page**

Table of Contents (continued)

List of Tables

This page intentionally left blank

List of Exhibits

This page intentionally left blank

1. Overview

1.1 Target Population

The respondent universe for the 2016 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 NSDUH's designs since 1991, the 2016 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 2016 universe included those with no fixed household address (e.g., homeless and/or transient persons not in shelters), the active military population, 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. Furthermore, NSDUH was redesigned in 2014 to allow for a more cost-efficient sample allocation to the largest states, while maintaining adequate sample sizes in smaller states 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. The target national sample size of 67,507 is distributed across five age groups as follows: 25 percent for youths aged 12 to 17, 25 percent for young adults aged 18 to 25, 15 percent for adults aged 26 to 34, 20 percent for adults aged 35 to 49, and 15 percent for adults aged 50 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 was required prior to 1999. This large sample is referred to as the "main sample." The achieved sample for the 2016 NSDUH was 67,942 persons.

Beginning with the 2002 NSDUH and continuing through the 2016 NSDUH, survey respondents were given a $30 incentive 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. Beginning in 2014 and continuing through 2016, this increase was offset by selecting fewer youths aged 12 to 17, requiring fewer selected households per completed interview.

Finally, a new pair sampling strategy was implemented in 2002 that increased the number of pairs selected in dwelling units (DUs) with older persons on the roster (Chromy & Penne,

[1] This report presents information from the 2016 NSDUH. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

2002). With the increase in the number of pairs came a moderate decrease in the response rate for older persons. Changes to the 2014 sample design with respect to age group and state necessitated a review of the pair sampling strategy. As a result, slightly fewer pairs were selected for the 2014, 2015, and 2016 NSDUHs.

1.3 4-Year Design

A coordinated sample design was developed for the 2014 through 2017 NSDUHs. The coordinated design facilitates 50 percent overlap in third-stage units (area segments) within each successive 2-year period from 2014 through 2017. This designed sample overlap slightly increases the precision of estimates of year-to-year trends because of the expected small but positive correlation resulting from the overlapping sample between successive survey years. The 50 percent overlap of segments significantly reduces segment listing costs because only one half of the segments need to be listed for the 2015 through 2017 surveys.

The 2014 through 2017 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. As shown in Table 1.1, the survey's sample was designed to yield the following:

- 4,560 completed interviews in California;
- 3,300 completed interviews each in Florida, New York, and Texas;
- 2,400 completed interviews each in Illinois, Michigan, Ohio, and Pennsylvania;
- 1,500 completed interviews each in Georgia, New Jersey, North Carolina, and Virginia;
- 967 completed interviews in Hawaii; and
- 960 completed interviews in each of the remaining 37 states and the District of Columbia.

To accommodate state and local policymakers' need for substate estimates in Kauai County, Hawaii, the sample was designed to yield a minimum of 200 completed interviews in this county over a 3-year period. This will allow for Kauai County to be included as a separate entity in the production of substate estimates that are produced biennially and typically based on 3 years of data. To achieve this goal while maintaining precision at the state level, Kauai County will be treated separately from the remainder of Hawaii for sample allocation and sample size management purposes. The annual sample in Hawaii will consist of 67 completed interviews in Kauai County and 900 completed interviews in the remainder of the state, for a total of 967 completed interviews each year.

Table 1.1 Annual National Sample of Area Segments and Respondents

Design Parameters	CA	FL, NY, and TX	IL, MI, OH, and PA	GA, NJ, NC, and VA	HI	Remaining 37 States and DC	Total
Total Sample							
SSRs	36	90	96	60	12	456	750
Segments	288	720	768	480	96	3,648	6,000
Respondents	4,560	9,900	9,600	6,000	967	36,480	67,507
Total per State							
SSRs	36	30	24	15	12	12	N/A
Segments	288	240	192	120	96	96	N/A
Respondents	4,560	3,300	2,400	1,500	967	960	N/A
Total per SSR							
Segments per Quarter	2	2	2	2	2	2	N/A
Segments over Four Quarters	8	8	8	8	8	8	N/A
Respondents per Segment	15.833	13.750	12.500	12.500	10.073	10.000	N/A

CA = California; DC = District of Columbia; FL = Florida; GA = Georgia; HI = Hawaii; IL = Illinois; MI = Michigan; N/A = not applicable; NC = North Carolina; NJ = New Jersey; NY = New York; OH = Ohio; PA = Pennsylvania; SSR = state sampling region; TX = Texas; VA = Virginia.

1.4 Stratification and First-, Second-, and Third-Stage Sample Selections

Within each state, state sampling regions (SSRs) were formed. Based on a composite size measure, each state was 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 within each state during each data collection period. This partitioning divided the United States into 750 SSRs. Maps for these regions can be found in Appendix A.

Similar to the 2005 through 2013 NSDUHs, the first stage of selection for the 2014 through 2017 NSDUHs was census tracts.[2] This stage was included to contain sample segments within a single census tract to the extent possible.[3] Segments that cross census tract boundaries make merging to external data sources difficult.

The first stage of selection began with the construction of an area sample frame that contained one record for each census tract in the United States. If necessary, census tracts were aggregated within SSRs until each first-stage sampling unit met the minimum size requirement. In California, Florida, Georgia, Illinois, Michigan, New Jersey, New York, North Carolina,

[2] 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).

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

Ohio, Pennsylvania, Texas, and Virginia, this minimum size requirement was 250 DUs[4] in urban areas and 200 DUs in rural areas.[5] In the remaining states and the District of Columbia, the minimum requirement was 150 DUs in urban areas and 100 DUs in rural areas.

Before selecting census tracts,[6] additional implicit stratification was achieved by sorting the first-stage sampling units by a CBSA/SES[7] (core-based statistical area/socioeconomic status) indicator[8] and by the percentage of the population that is non-Hispanic and white.[9] 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).

For the second stage of selection, adjacent census block groups were aggregated within selected census tracts as necessary to meet the minimum DU requirements (150 or 250 DUs in urban areas and 100 or 200 DUs in rural areas according to state). After the resulting second-stage sampling units were formed, they were sorted in the order they were formed (i.e., geographically), and one census block group[10] was selected per sampled census tract with probability proportionate to a composite size measure and with minimum replacement (Chromy, 1979). Compared with prior years, the selection of census block groups is an additional stage of selection that was included to facilitate possible transitioning to an address-based sampling (ABS) design.

Because census block groups generally exceed the minimum DU requirement, one smaller geographic region was selected within each sampled census block group. For this third

[4] DU counts were obtained from the 2010 census data supplemented with revised population counts from Nielsen Claritas.

[5] 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, 100 DUs are sufficient to support one field test and two main study samples in the smaller states, and 200 DUs are sufficient to support three samples in the larger sample states.

[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] CBSAs include metropolitan and micropolitan statistical areas as defined by the Office of Management and Budget (2009).

[8] 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 2006-2010 American Community Survey data 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."

[9] 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.

[10] For the remainder of the discussion, second-stage sampling units are referred to as "census block groups" even though each second-stage sampling unit contains one or more census block groups.

stage of sampling, each selected census block group was partitioned into compact clusters[11] of DUs by aggregating adjacent census blocks.[12] 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 and census block groups, segments were formed to contain a minimum of 150 or 250 DUs in urban areas and 100 or 200 DUs in rural areas according to state. 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 block group using Chromy's method of sequential random sampling (with probability proportionate to size and minimum replacement) (Chromy, 1979). The 48 selected segments then were randomly assigned to a survey year and quarter of data collection as described in Section 2.4.

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 2016, the first panel segments (panel C) were used for the 2015 and 2016 surveys, constituting the overlap sample. The second panel segments (panel D) were used for the 2016 survey and will be used again for the 2017 survey.

1.5 Sample Dwelling Units and Persons

After sample segments for the 2016 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 fourth stage of sample selection.

The primary objective of the fourth stage of sample selection (listing units) was to select 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 2010 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

[11] 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).

[12] 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).

subsegmentation inflation factor,[13] 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.[14] In addition to these factors, historical data from the 2014, 2015, and 2016 NSDUHs were used to compute predicted screening and interviewing response rate adjustments. The final adjusted sampling rate then was multiplied by the actual number of DUs found in the field during counting and listing activities. The product represents the segment's listing unit sample size.

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

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

Individuals selected in a given year are not expected to be selected in subsequent years unless they move and their new residence is also selected. Because of the new sample design, some DUs selected in 2013 may be selected in 2014 to 2017 by chance. No mechanism is currently in place for identifying duplicate persons across years within a sample design (e.g., 2014 to 2017) or across quarters within a year, but this number should be small because DUs are not sampled more than once within a design.

One advantage of using an electronic screening instrument in NSDUH is the ability to impose a more complicated person-level selection algorithm on the fifth stage of the NSDUH design. Similar to the 1999 through 2013 designs, one feature that was included in the 2014 through 2016 NSDUHs 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). The pair sampling algorithm in NSDUH is based on the Chromy and Penne (2002) adaptation of the Brewer (1963, 1975) method for selecting samples of size two. Chromy and Penne (2002) also introduced a pair sampling parameter λ that governs the number of pairs selected. Appendix B describes the simulation

[13] 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.

[14] 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.

analyses that were conducted to select the pair sampling parameter for the 2014 through 2016 NSDUHs.

As in previous years, during the data collection period, if an interviewer encountered any new or missed DU on the premises of a sampled DU (e.g., a garage apartment), the new or missed dwelling was selected into the 2016 NSDUH. However, unlike the 2005 through 2013 NSDUHs, the half-open interval (HOI) procedure[15] was not implemented beginning in 2014. 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). Excluding the HOI procedure decreases the burden on field interviewers (FIs) and simplifies the screening process. This decrease in burden outweighs the small increase in coverage resulting from implementation of the HOI procedure. To minimize bias associated with large numbers of missed DUs, interviewers were instructed to call their supervisors if they noticed large differences in the segment listing and what they encountered in the field. Then special "bust" procedures were implemented, as described in Section 3.7.

1.6 Comparison with the 2005 through 2013 Design

Similar to the sample design for the 2005 through 2013 NSDUHs, the 2014 through 2017 design is a stratified, multistage area probability sample. Both sample designs provide for estimates by state in all 50 states and the District of Columbia. However, the allocation of sample to states differs between the two designs. As shown in Table 1.2, in the 2005 through 2013 NSDUH design, the sample was divided into eight "large" states and 43 "small" states (including the District of Columbia), with the large and small sample states designed to yield 3,600 and 900 respondents per state, respectively. The 2005 through 2013 samples were designed to yield the same number of interviews from each area segment.

Table 1.2 Sample Sizes and Targeted Respondents, by State and Age Group: 2005 through 2013

State	Aged 12-17	Aged 18-25	Aged 26+	Total Aged 12+	SSRs	Average Segment Size	Number of Segments
Total Population	22,500	22,500	22,500	67,500	900	9.375	7,200
Large Sample States (CA, FL, IL, MI, NY, OH, PA, and TX)	1,200	1,200	1,200	3,600	48	9.375	384
Small Sample States (Remaining 42 States and DC)	300	300	300	900	12	9.375	96

CA = California; DC = District of Columbia; FL = Florida; GA = Georgia; HI = Hawaii; IL = Illinois; MI = Michigan; NY = New York; OH = Ohio; PA = Pennsylvania; SSR = state sampling region; TX = Texas.

[15] In summary, the HOI technique states that, if a DU is selected 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. If a large number of new or missed DUs are encountered (greater than 10), a sample of the new or missing DUs is selected, and the sample weight is adjusted accordingly.

The 2014 through 2017 sample design moves from two to essentially five state sample size groups (lumping Hawaii with the remaining states and the District of Columbia). As shown in Table 1.3, the 2014 through 2017 surveys have a sample designed to yield 4,560 completed interviews in California; 3,300 completed interviews each in Florida, New York, and Texas; 2,400 completed interviews each in Illinois, Michigan, Ohio, and Pennsylvania; 1,500 completed interviews each in Georgia, New Jersey, North Carolina, and Virginia; 967 completed interviews in Hawaii; and 960 completed interviews in each of the remaining 37 states and the District of Columbia—for a total national target sample size of 67,507. The sample is selected from 6,000 area segments that vary in size according to state.

Table 1.3 Sample Sizes and Projected Respondents, by State and Age Group: 2014 through 2017

State	Aged 12-17	Aged 18-25	Aged 26-34	Aged 35-49	Aged 50+	Total Aged 12+	SSRs	Average Segment Size	Number of Segments
Total Population	16,877	16,877	10,126	13,501	10,126	67,507	750	11.251	6,000
CA	1,140	1,140	684	912	684	4,560	36	15.833	288
FL, NY, and TX	825	825	495	660	495	3,300	30	13.750	240
IL, MI, OH, and PA	600	600	360	480	360	2,400	24	12.500	192
GA, NJ, NC, and VA	375	375	225	300	225	1,500	15	12.500	120
HI	242	242	145	193	145	967	12	10.073	96
Remaining 37 States and DC	240	240	144	192	144	960	12	10.000	96

CA = California; DC = District of Columbia; FL = Florida; GA = Georgia; HI = Hawaii; IL = Illinois; MI = Michigan; NC = North Carolina; NJ = New Jersey; NY = New York; OH = Ohio; PA = Pennsylvania; SSR = state sampling region; TX = Texas; VA = Virginia.

The change in the state sample allocation was driven by the need to increase sample in the original 43 small states (to improve the precision of state and substate estimates in these states) while moving closer to a proportional allocation in the larger states. Table 1.4 displays population percentages by state from the 2010 census and sample sizes and percentages for the 2013 survey and each of the 2014 through 2017 surveys. The five state groups are grouped in separate blocks of rows.

In addition to having a different sample allocation by state, the 2014 through 2017 design places more sample in the 26 or older age groups to estimate drug use and related mental health measures more accurately among the aging drug use population. For the 2014 through 2017 NSDUHs, each state sample will be allocated to age groups as follows: 25 percent for youths aged 12 to 17, 25 percent for young adults aged 18 to 25, 15 percent for adults aged 26 to 34, 20 percent for adults aged 35 to 49, and 15 percent for adults aged 50 or older. In the 2005 through 2013 NSDUHs, the sample was allocated equally across the 12 to 17, 18 to 25, and 26 or older age groups.

Table 1.4 Population and Sample Percentages, by State: 2013 and 2014-2017

Variable	2010 CNI Population	Percent of Population	2014-2017 Annual Sample Size	Percent of 2014-2017 Annual Sample	2013 Sample Size	Percent of 2013 Sample
Total U.S.	253,619,107	100.0%	67,507	100.0%	67,500	100.0%
California	30,322,142	12.0%	4,560	6.8%	3,600	5.3%
Texas	19,847,501	7.8%	3,300	4.9%	3,600	5.3%
New York	16,410,083	6.5%	3,300	4.9%	3,600	5.3%
Florida	15,611,774	6.2%	3,300	4.9%	3,600	5.3%
Illinois	10,629,517	4.2%	2,400	3.6%	3,600	5.3%
Pennsylvania	10,607,311	4.2%	2,400	3.6%	3,600	5.3%
Ohio	9,580,362	3.8%	2,400	3.6%	3,600	5.3%
Michigan	8,313,433	3.3%	2,400	3.6%	3,600	5.3%
Georgia	7,940,651	3.1%	1,500	2.2%	900	1.3%
North Carolina	7,679,126	3.0%	1,500	2.2%	900	1.3%
New Jersey	7,269,834	2.9%	1,500	2.2%	900	1.3%
Virginia	6,471,190	2.6%	1,500	2.2%	900	1.3%
Massachusetts	5,605,641	2.2%	960	1.4%	900	1.3%
Washington	5,585,609	2.2%	960	1.4%	900	1.3%
Arizona	5,386,782	2.1%	960	1.4%	900	1.3%
Indiana	5,286,018	2.1%	960	1.4%	900	1.3%
Tennessee	5,238,574	2.1%	960	1.4%	900	1.3%
Missouri	4,952,896	2.0%	960	1.4%	900	1.3%
Wisconsin	4,726,785	1.9%	960	1.4%	900	1.3%
Maryland	4,737,806	1.9%	960	1.4%	900	1.3%
Minnesota	4,382,130	1.7%	960	1.4%	900	1.3%
Colorado	4,151,930	1.6%	960	1.4%	900	1.3%
Alabama	3,893,688	1.5%	960	1.4%	900	1.3%
South Carolina	3,760,624	1.5%	960	1.4%	900	1.3%
Louisiana	3,661,821	1.4%	960	1.4%	900	1.3%
Kentucky	3,574,784	1.4%	960	1.4%	900	1.3%
Oregon	3,229,211	1.3%	960	1.4%	900	1.3%
Oklahoma	2,995,565	1.2%	960	1.4%	900	1.3%
Connecticut	2,951,217	1.2%	960	1.4%	900	1.3%
Iowa	2,502,115	1.0%	960	1.4%	900	1.3%
Mississippi	2,373,593	0.9%	960	1.4%	900	1.3%
Arkansas	2,375,992	0.9%	960	1.4%	900	1.3%
Kansas	2,296,286	0.9%	960	1.4%	900	1.3%
Nevada	2,155,405	0.8%	960	1.4%	900	1.3%

(continued)

Table 1.4 Population and Sample Percentages, by State: 2013 and 2014-2017 (continued)

Variable	2010 CNI Population	Percent of Population	2014-2017 Annual Sample Size	Percent of 2014-2017 Annual Sample	2013 Sample Size	Percent of 2013 Sample
Utah	2,180,889	0.9%	960	1.4%	900	1.3%
New Mexico	1,641,892	0.6%	960	1.4%	900	1.3%
West Virginia	1,543,694	0.6%	960	1.4%	900	1.3%
Nebraska	1,469,129	0.6%	960	1.4%	900	1.3%
Idaho	1,250,238	0.5%	960	1.4%	900	1.3%
Maine	1,127,285	0.4%	960	1.4%	900	1.3%
New Hampshire	1,128,997	0.4%	960	1.4%	900	1.3%
Hawaii	1,047,745	0.4%	967	1.4%	900	1.3%
Rhode Island	896,384	0.4%	960	1.4%	900	1.3%
Montana	820,115	0.3%	960	1.4%	900	1.3%
Delaware	737,571	0.3%	960	1.4%	900	1.3%
South Dakota	666,589	0.3%	960	1.4%	900	1.3%
Alaska	555,964	0.2%	960	1.4%	900	1.3%
Vermont	538,568	0.2%	960	1.4%	900	1.3%
North Dakota	540,202	0.2%	960	1.4%	900	1.3%
District of Columbia	517,942	0.2%	960	1.4%	900	1.3%
Wyoming	448,513	0.2%	960	1.4%	900	1.3%

CNI = civilian, noninstitutionalized population.

The 2014 through 2017 design includes the selection of census block groups at the second stage of selection. As mentioned in Section 1.4, this additional stage of selection was included to facilitate moving to an ABS design, if desired. Compared with geocoding at the census block level, geocoding accuracy improves significantly at the census block group level in both rural and urban areas. Thus, in an ABS design, census block groups would serve as geographic clusters in areas with sufficient mailing address coverage. The selection of census tracts at the first stage of selection and census block groups at the second stage has the potential to reduce sampling variance by controlling the distribution of selected areas and reducing the chance of selecting neighboring and possibly similar areas within tracts and block groups. In addition, the merging of NSDUH data to external data sources for future analysis purposes is simplified when sampled areas are contained within tract and block group boundaries to the extent possible.

Finally, as mentioned in Section 1.5, the 2014 through 2017 NSDUH fourth-stage sampling frames are supplemented with new DUs on the premises of sampled DUs that were missed during the original counting and listing activities (e.g., garage apartments). However, the HOI procedure is no longer being implemented.

2. Coordinated 2014 through 2017 Sample

As was mentioned previously, the sample design was developed simultaneously for each of the 2014 through 2017 National Surveys on Drug Use and Health (NSDUHs). Starting with a census block-level frame, first-, second-, and third-stage sampling units (census tracts, census block groups, and area segments, respectively) were formed by aggregating the block-level data to the appropriate level. A sufficient number of segments then were selected within sampled census tracts to support the 4-year design and any supplemental studies that the Substance Abuse and Mental Health Services Administration (SAMHSA) may choose to field.

2.1 Formation of and Objectives for Using the Composite Size Measures

The composite size measure procedure is used to obtain self-weighting[16] 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.

[16] Self-weighting implies equal weights within domains defined by state and age group.

Using the 2010 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

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

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 the expected sample size per region, which varies by state. 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 within a state.

A total of 750 SSRs were formed for the coordinated 2014 through 2017 design: 36 SSRs in California; 30 SSRs each in Florida, New York, and Texas; 24 SSRs each in Illinois, Michigan, Ohio, and Pennsylvania; 15 SSRs each in Georgia, New Jersey, North Carolina, and Virginia; and 12 SSRs in each of the remaining 38 states and the District of Columbia. To facilitate sample allocation and sample size management in Kauai County, Hawaii, this county was assigned to 1 SSR, and the remainder of Hawaii was divided into 11 approximately equal-sized SSRs.

2.3 First-, Second-, and Third-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 a sufficient number of DUs, some had to be collapsed in order to meet the minimum requirement. In California, Florida, Georgia, Illinois, Michigan, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Texas, and Virginia, this minimum size requirement was 250 DUs in urban areas and 200 DUs in rural areas. In the remaining states and the District of Columbia, the minimum requirement was 150 DUs in urban areas and 100 DUs in rural areas. Once the 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.

For the second stage of selection, adjacent census block groups were collapsed as needed within selected census tracts. The block groups were required to have the same minimum number of DUs as the census tracts from which they were selected (150 or 250 in urban areas and 100 or 200 in rural areas, according to state). The resulting block groups were then sorted in the order in which they were formed, and one block group was selected per selected census tract with probability proportionate to a composite size measure.

To form segments within sampled census block groups, adjacent census blocks were collapsed until the total number of DUs within the area was at least 150 or 250 in urban areas and 100 or 200 in rural areas, according to state. In order to obtain geographic ordering of the blocks within block groups, block centroids were serpentine-sorted by latitude and longitude.[17] 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 block group 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 36,000 segments. Although only 20 segments per SSR were needed to support the 4-year study from 2014 through 2017, an additional 28 segments were selected to serve as replacements when segment DUs are depleted and/or to support any supplemental studies embedded within NSDUH. These 28 segments constitute the "reserve" sample and are also available to extend the sample to the next decennial census, if desired.

[17] The latitude and longitude for each census block were obtained from the Census 2010 Summary File 1, which is available at http://www2.census.gov/census_2010/04-Summary_File_1/.

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

State	State Abbrevia-tion	State FIPS Code	Number of Census Tracts on Sampling Frame	Number of Block Groups on Sampling Frame	Number of Segments on Sampling Frame	Total Number of Census Tracts/ Block Groups/ Segments Selected	Number Selected for 4-Year Sample	Unique Segments in 4-Year Sample
Total U.S.			72,006	96,898	99,137	36,000	15,000	14,605
Northeast								
Connecticut	CT	09	824	1,837	1,646	576	240	239
Maine	ME	23	351	1,037	1,960	576	240	230
Massachusetts	MA	25	1,458	2,100	1,690	576	240	240
New Hampshire	NH	33	292	907	1,883	576	240	219
New Jersey	NJ	34	1,988	2,417	1,321	720	300	300
New York	NY	36	4,798	4,956	2,260	1,440	600	600
Pennsylvania	PA	42	3,184	3,849	2,202	1,152	480	480
Rhode Island	RI	44	240	796	1,486	576	240	215
Vermont	VT	50	183	515	1,363	576	240	210
Midwest								
Illinois	IL	17	3,107	3,706	2,311	1,152	480	479
Indiana	IN	18	1,503	1,968	2,114	576	240	240
Iowa	IA	19	823	1,781	1,830	576	240	239
Kansas	KS	20	758	1,728	1,840	576	240	237
Michigan	MI	26	2,736	3,476	2,078	1,152	480	480
Minnesota	MN	27	1,332	1,891	1,996	576	240	240
Missouri	MO	29	1,386	2,014	2,035	576	240	240
Nebraska	NE	31	529	1,471	1,602	576	240	226
North Dakota	ND	38	205	569	1,452	576	240	217
Ohio	OH	39	2,931	3,729	2,253	1,152	480	480
South Dakota	SD	46	221	644	1,555	576	240	214
Wisconsin	WI	55	1,390	1,975	1,925	576	240	240
South								
Alabama	AL	01	1,174	1,801	2,224	576	240	240
Arkansas	AR	05	683	1,708	2,146	576	240	239
Delaware	DE	10	214	546	1,480	576	240	215
District of Columbia	DC	11	178	448	1,057	576	240	193
Florida	FL	12	4,148	4,009	3,984	1,440	600	600
Georgia	GA	13	1,949	2,142	1,820	720	300	300
Kentucky	KY	21	1,104	1,875	1,876	576	240	240
Louisiana	LA	22	1,121	1,869	1,912	576	240	240
Maryland	MD	24	1,384	1,713	1,662	576	240	240
Mississippi	MS	28	654	1,727	2,199	576	240	239
North Carolina	NC	37	2,158	2,171	1,739	720	300	300
Oklahoma	OK	40	1,036	1,781	1,961	576	240	239
South Carolina	SC	45	1,082	1,711	2,270	576	240	240
Tennessee	TN	47	1,475	1,708	2,310	576	240	240
Texas	TX	48	5,191	4,493	3,426	1,440	600	599
Virginia	VA	51	1,864	2,174	1,412	720	300	300
West Virginia	WV	54	484	1,485	1,801	576	240	232

(continued)

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

State	State Abbreviation	State FIPS Code	Number of Census Tracts on Sampling Frame	Number of Block Groups on Sampling Frame	Number of Segments on Sampling Frame	Total Number of Census Tracts/ Block Groups/ Segments Selected	Number Selected for 4-Year Sample	Unique Segments in 4-Year Sample
West								
Alaska	AK	02	165	516	1,212	576	240	208
Arizona	AZ	04	1,508	1,721	1,951	576	240	240
California	CA	06	7,935	5,155	3,570	1,728	720	720
Colorado	CO	08	1,234	1,760	2,166	576	240	239
Hawaii	HI	15	309	789	1,380	576	240	216
Idaho	ID	16	296	939	2,166	576	240	228
Montana	MT	30	268	814	1,765	576	240	215
Nevada	NV	32	678	1,448	1,718	576	240	236
New Mexico	NM	35	497	1,301	2,000	576	240	231
Oregon	OR	41	825	1,816	2,063	576	240	240
Utah	UT	49	582	1,448	1,944	576	240	234
Washington	WA	53	1,440	2,055	1,816	576	240	240
Wyoming	WY	56	131	409	1,305	576	240	197

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 repeated 11 times until the 48 sampled segments were assigned to 12 subsamples of 4 segments each—5 subsamples to field the 2014 through 2017 surveys and 7 "reserve" subsamples.

The first subsample of segments was assigned to the 2014 NSDUH and constituted the panel of segments to be used for 2014 only. The second subsample of segments was assigned to the 2014 NSDUH and was used again in the 2015 survey; the third subsample was assigned to the 2015 and 2016 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 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," with the next-to-last digit being the panel identifier and the last digit being the original quarter assignment. A small number of segments are switched with another segment in the same SSR and panel due to difficult conditions during the winter months. In general, quarter 1 segments are switched with quarter 2 segments, and quarter 4 segments are switched with quarter 3 segments. The 2016 main survey corresponds to segment suffixes C1 through C4 and D1 through D4.

Table 2.2 Segment Identification Number Suffixes and Quarter Assignment

Segment Suffix	2014 NSDUH	2015 NSDUH	2016 NSDUH	2017 NSDUH	Variance Replicate
A1	x (Q1)				1
A2	x (Q2)				1
A3	x (Q3)				1
A4	x (Q4)				1
B1	x (Q1)	x (Q1)			2
B2	x (Q2)	x (Q2)			2
B3	x (Q3)	x (Q3)			2
B4	x (Q4)	x (Q4)			2
C1		x (Q1)	x (Q1)		1
C2		x (Q2)	x (Q2)		1
C3		x (Q3)	x (Q3)		1
C4		x (Q4)	x (Q4)		1
D1			x (Q1)	x (Q1)	2
D2			x (Q2)	x (Q2)	2
D3			x (Q3)	x (Q3)	2
D4			x (Q4)	x (Q4)	2
E1				x (Q1)	1
E2				x (Q2)	1
E3				x (Q3)	1
E4				x (Q4)	1

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

2.5 Creation of Variance Estimation Strata and Replicates for Person-Level Estimates

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 representing the variance estimation strata and replicates were created to capture explicit stratification and to identify clustering. For the 2014 through 2017 NSDUHs, variance estimation strata are defined at the SSR level, and each SSR appears in a different stratum every quarter. Because census tracts, block groups, and segments are nested within variance replicates, the variance contributions of all three sampling units are covered by the nesting variables. Also, because one segment is selected per sampled census tract and block group, the selection of census tracts and block groups at the first stages of selection may reduce variance by minimizing the chance of selecting adjacent and possibly similar segments within the same census tract or block group.

To define the variance estimation strata for the 2014 through 2017 NSDUHs, the 750 SSRs were placed in random order (states were randomly sorted, and regions were randomly sorted within states). This list, numbered 1 to 750, defined the quarter 1 variance estimation strata (VESTRQ1). For quarter 2, the variance estimation strata, VESTRQ2, were defined as VESTRQ1 + 150 (or VESTRQ1 + 150 - 750 if VESTRQ1 is ≥ 601). Similarly, VESTRQ3 = VESTRQ2 + 150 (- 750 if VESTRQ2 ≥ 601), and VESTRQ4 = VESTRQ3 + 150 (- 750 if VESTRQ3 ≥ 601). As an example, an SSR that was assigned to stratum 451 in quarter 1 was assigned to stratum 601 (= 451 + 150) in quarter 2, stratum 1 (= 601 + 150 - 750) in quarter 3, and stratum 151 (= 1 + 150) in quarter 4. Finally, to make the values unique from previous years, 40,000 was added to the quarterly variance estimation strata to create the final variance estimation strata (e.g., VESTR =

VESTRQ1 + 40,000). The resulting 750 strata values are the same for the 2014 through 2017 NSDUHs, but have different values from variance estimation strata defined for previous years because the 2014 through 2017 sample was selected independently from previous samples.

The 2014 through 2017 method of defining variance estimation strata had the effect of assigning SSRs to strata in a pseudo-random fashion while ensuring that each stratum consists of four SSRs from four different states. Combining SSRs across states to form strata reduces the risk of disclosure because an intruder cannot assume that respondents in the same variance estimation stratum reside in the same state.

Similar to the 2005 through 2013 definition of variance estimation strata, the 2014 through 2017 definition also 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 (750). Each of the smallest sample states is in 48 different strata (12 SSRs × 4 quarters); therefore, 48 degrees of freedom are available for state estimates in these states. At the other extreme, the largest sample state, California, is in 144 strata (36 SSRs × 4 quarters) and therefore has 144 degrees of freedom for estimation. As demonstrated in Appendix C, the 2014 through 2017 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 shown earlier describes the assignment of segments to variance estimation replicates that are designed to account for positive covariance among consecutive year change estimates. As shown in the table, the replicate values alternate between 1 and 2 for each panel of the 2014 through 2017 design. As a result, when combining data from multiple years, the pooling of the samples within replicates provides increased precision of estimates.

In addition to variance estimation strata and replicates, a sample weight is computed for each final respondent (see Section 3.9.1 in Chapter 3). 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.[18] Using data from the 2007 and 2008 NSDUHs and examining multiple measures,

[18] 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.

the average relative change in the standard error (SE) after accounting for covariance was about 1 percent.

2.6 Creation of Variance Estimation Strata and Replicates for Pair Analyses

Alternate versions of the variance estimation strata and replicates are created for use in analyses involving responding pairs (i.e., two unit respondents in the same household). In most years, these alternate versions are simply copies of the person-level variance estimation strata and replicates described in Section 2.5. Occasionally, however, alternate versions must be made because not all variance estimation strata have at least one responding pair in both variance replicates. In this scenario, strata with only one variance replicate represented by responding pairs were collapsed with a neighboring stratum.

The algorithm for selecting the neighboring stratum is as follows:

- Identify the person-level variance estimation stratum that precedes the problematic one and the person-level variance estimation stratum that follows the problematic one. The numbering is described in Section 2.5.
- If one of the neighboring strata is in a different state, then collapse the problematic one with the neighboring stratum in the same state. Otherwise, choose between the two neighbors randomly.

2.7 Other Sampling-Related Variables

Because area segments consist of one or more census blocks, a number of demographic and geographic variables are available for sampled areas. The demographic data include the following: population counts by age, race, and ethnicity; estimated civilian, noninstitutional population aged 12 or older; DU counts; estimated group quarters units; and group quarters population by type of group quarter.[19] 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 2014 through 2017 NSDUH sample segments. These are state, county and county name, 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.[20] 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). However, the definition slightly overestimates the urban population.

[19] Data were obtained or derived from the Census 2010 Summary File 1 and adjusted using revised population counts from Nielsen Claritas.

[20] All variables were obtained or derived from the Census 2010 Summary File 1.

Similar to the 2005 through 2013 NSDUH samples, the 2014 through 2017 samples were 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 in Chapter 3), only the group of blocks in which a NSDUH respondent resides is known. Beginning in 2014 and continuing through 2017, the specific census block associated with each NSDUH respondent's listing unit will be assigned. Using desktop computer mapping software, census block information is recorded by manually comparing electronic segment maps to field listings as is described in Appendix D. As a result, block-level data can be associated with NSDUH respondents, improving the accuracy of geographic variables used for data analyses. The block-level geographic variables include place name, rural or urban indicator, and American Indian, Alaska Native, and Native Hawaiian area codes and names.

At the end of quarters 2 and 4, preliminary census block assignments were prepared so that block-level covariates could be defined and used in item imputation models. To accommodate the 6- and 12-month imputation schedules, all respondent DUs as of June 1, 2016, and December 1, 2016, for quarters 2 and 4, respectively, were identified. Next, the specific census block associated with each of these DUs was manually assigned using the procedures described in Appendix D. Addresses for all pending DUs were geocoded. Finally, block-level covariates were defined using the manually assigned or geocoded census block. If geocoding did not result in a valid census block (e.g., the address was geocoded outside the segment boundary, or the DU address was a description), the segment-level covariate was used.

Prior to preparing preliminary census block assignments for the first time in 2015, the process was simulated using data from quarter 2 of the 2014 NSDUH. First, all interviews that were completed before June 1, 2014, were identified. The respondent DUs associated with these interviews received the manually assigned census block. Addresses for all other DUs were geocoded. The block-level covariates were then merged onto the data using the simulated census block. If no census block was assigned (i.e., the geocoding was not successful), the respondent received segment-level covariates. Finally, the simulated census block was compared with the final assigned census block. Because the majority of the quarter 2 cases are completed by June 1, the simulation found high agreement between the preliminary and final census blocks. Furthermore, the preliminary and final imputation model covariates had even higher match rates because of the correlation between the segment- and block-level covariates.

This page intentionally left blank

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 2016 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) design requirements while simultaneously minimizing the cost of data collection. Costs were minimized by determining the fewest 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 achieve acceptable precision for several outcome measures was determined. Using the results of several optimization models and other related analyses conducted in advance of the 2014 through 2017 NSDUHs, SAMHSA specified the 255 m_{ha} values for each state h (51) and age group a (5). This is described in further detail in Section 3.2.
2. Using the m_{ha} from Step 1, the next step was to determine the optimal number of selected dwelling (D_{hj}) units (i.e., fourth-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.[21] 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.

3.1 Notation

h = state indicator for 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.

[21] Direct application of Brewer's method would require a fixed sample size.

j = Individual segment indicator (total of 6,000; 1,500 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 2016 population is estimated using the 2010 census data adjusted to the 2013 Nielsen 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 2013, A = the population in 2010, and $x = 3$.[22] Then the 2016 population (y^*_{ha}) is computed using the original formula and this time allowing x to be 6. Finally, the 2016 population is adjusted by the ratio of estimated eligible listed DUs to the Nielsen Claritas DU counts (U_{hj}). This adjustment factor considers the number of added DUs expected to be found on the premises of sampled DUs (1.005) and the probability of a DU being eligible (ε_h), both determined via historical data. The coefficient adjustment of 1.005 is estimated using 2014 NSDUH 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 and census block group selection probabilities). 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

[22] A comparison of B to annual growth rates computed using weighted population estimates from the 2011 through 2014 NSDUHs found that this approach may slightly underestimate the annual growth rate.

subdivisions or subsegments (using a best guess estimate of the number of DUs in each subsegment: G^*_{1hj}) 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: G^*_{2hj}). Then one subsegment was selected for regular counting and listing by sampling staff. For the subsegment to represent the entire segment, the weights were adjusted up to reflect the unused portion of the segment:

= (G^*_{1hj} / L^*_{1hj}), if one round of subsegmenting was done;

= $(G^*_{1hj} / L^*_{1hj}) * (G^*_{2hj} / L^*_{2hj})$, if two rounds of subsegmenting were required; and

= 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 2014 NSDUH quarter 4 and 2015 NSDUH quarters 1 through 3 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 quarter 4 of the 2014 NSDUH and quarters 1 through 3 of the 2015 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 2014 NSDUH quarter 4 and 2015 NSDUH quarters 1 through 3 data by dividing the weighted total number of rostered persons in an age group by the weighted total number of complete screened DUs by state.

δ_{ha} = State- and age group-specific maximum-of-two rule adjustment. The survey design restricts the number of interviews per DU to a total of two. This is achieved through a modified Brewer's Method of Selection, 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 2014 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 fifth 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, while ensuring adequate precision for key outcomes of interest. In preparation for the 2014 through 2017 NSDUH sample redesign, several optimization models and other related analyses were conducted (RTI International, 2012a). SAMHSA used the results from these analyses to inform the 2014 through 2017 design. Compared with the sample allocation in prior years, the 2014 through 2017 design allows for a more cost-efficient sample allocation to the largest states, while maintaining a sufficient sample size in each of the smaller states to support small area estimation (SAE) at the state and substate levels. Furthermore, the 2014 through 2017 design increases the 26 or older sample size to more accurately estimate drug use and related mental health measures among this age group.

Sample requirements for the 2016 NSDUH were defined in terms of the following:

- minimum sample sizes of 4,560 completed interviews in California; 3,300 completed interviews each in Florida, New York, and Texas; 2,400 completed interviews each in Illinois, Michigan, Ohio, and Pennsylvania; 1,500 completed interviews each in Georgia, New Jersey, North Carolina, and Virginia; 967 completed interviews in Hawaii; and 960 completed interviews in each of the remaining 37 states and the District of Columbia; and
- allocation to age groups as follows: 25 percent for youths aged 12 to 17, 25 percent for young adults aged 18 to 25, 15 percent for adults aged 26 to 34, 20 percent for adults aged 35 to 49, and 15 percent for adults aged 50 or older.

Using the initial state and age group sample requirements and 2013 NSDUH data, estimates and relative standard errors (RSEs) for 11 outcome measures and several domains of interest were modeled and are presented in Table 3.1. Specifically, the 11 key NSDUH outcome measures are as follows:

1. *Past month alcohol use (ALCMON)*. Had at least one drink in the past 30 days.
2. *Past month binge alcohol use (BINGEDRK)*. Drinking five or more drinks on the same occasion (i.e., at the same time or within a couple hours of each other) on at least 1 day in the past 30 days.
3. *Past month marijuana use (MRJMON)*.
4. *Past month cigarette use (CIGMON)*.

5. *Past month nonmedical use of pain relievers (ANLMON = 1).*
6. *Past year alcohol disorder (ABODALC).* Dependence or abuse of alcohol during the past 12 months.
7. *Past year illicit drug disorder (ABODILL).* Dependence or abuse of illicit drugs during the past 12 months.
8. *Substance use disorder (ABODILAL).* Dependence or abuse of illicit drugs or alcohol during the past 12 months.
9. *Past year specialty substance use treatment (SPILLALC).* Received treatment in the past year for an illicit drug or alcohol problem at a specialty facility. A specialty facility is a hospital (inpatient) or rehabilitation center (inpatient or outpatient).
10. *Past year serious mental illness (SMIYR).* Serious mental illness based on the predicted probability of SMI (SMIPP) and based on both distress and impairment.
11. *Past year major depressive episode (AMDEYR).*

Table 3.1 Relative Standard Errors and Sample Sizes for Key Outcome Measures, by Demographic Domain

Data File Variable Name	Measure	Domain	2013 Prevalence	Projected RSE (2016)	Expected Sample Size (2016)
ALCMON	Past Month Alcohol Use	12+	0.5216	0.0069	67,507
ALCMON	Past Month Alcohol Use	12-20	0.2272	0.0238	23,261
ALCMON	Past Month Alcohol Use	50+	0.4985	0.0127	10,126
ALCMON	Past Month Alcohol Use	API, 12+	0.3475	0.0472	3,506
ALCMON	Past Month Alcohol Use	AIAN, 12+	0.3728	0.0752	736
ALCMON	Past Month Alcohol Use	Pregnant, 12-44	0.0982	0.1341	814
BINGEDRK	Past Month Binge Alcohol Use	18-25	0.3791	0.0135	16,877
BINGEDRK	Past Month Binge Alcohol Use	12+	0.2289	0.0116	67,507
MRJMON	Past Month Marijuana Use	12+	0.0755	0.0205	67,507
MRJMON	Past Month Marijuana Use	12-17	0.0708	0.0384	16,877
MRJMON	Past Month Marijuana Use	18-25	0.1908	0.0221	16,877
MRJMON	Past Month Marijuana Use	50+	0.0296	0.0722	10,126
MRJMON	Past Month Marijuana Use	API, 12+	0.0293	0.1458	3,506
MRJMON	Past Month Marijuana Use	AIAN, 12+	0.1076	0.1633	736
MRJMON	Past Month Marijuana Use	Pregnant, 12-44	0.0467	0.1453	814
CIGMON	Past Month Cigarette Use	12-17	0.0561	0.0446	16,877
CIGMON	Past Month Cigarette Use	12+	0.2126	0.0136	67,507
ANLMON	Past Month Pain Reliever Use	18-25	0.0326	0.0559	16,877
ANLMON	Past Month Pain Reliever Use	12+	0.0172	0.0423	67,507
ABODALC	Past Year Alcohol Disorder	12+	0.0659	0.0219	67,507
ABODILL	Past Year Illicit Drug Disorder	12+	0.0261	0.0305	67,507
ABODILAL	Past Year Substance Use Disorder	50+	0.0398	0.0646	10,126

See notes at end of table.

(continued)

Table 3.1 Relative Standard Errors and Sample Sizes for Key Outcome Measures, by Demographic Domain (continued)

Data File Variable Name	Measure	Domain	2013 Prevalence	Projected RSE (2016)	Expected Sample Size (2016)
SPILLALC	Past Year Specialty Substance Use Treatment	12+	0.0094	0.0616	67,507
SMIYR	Past Year SMI	18+	0.0422	0.0305	50,630
AMDEYR	Past Year MDE	18+	0.0667	0.0244	50,630

AIAN = American Indian or Alaska Native (NEWRACE2 = 3); API = Asian or Other Pacific Islander (NEWRACE2 = 4 or 5); MDE = major depressive episode; Pregnant, 12-44 (PREG2 = 1); RSE = relative standard error; SMI = serious mental illness.

NOTE Projected RSEs were determined using 2014 through 2017 state and age sample allocations in a variance component model. Unequal weighting effects by state and age groups were updated using 2013 NSDUH data. Further, prevalence estimates by domain and associated simple random sample variances were also updated. Model components capturing the impact of the clustering of the variability of cluster sizes were not updated, however.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2013.

SAMHSA based the selection of the above outcome measures on an assessment of how the data are used and what estimates are important for policymakers. Domains such as pregnant women, American Indian or Alaska Native, and aged 50 or older were chosen based on the importance of generally maintaining precision of estimates in those subgroups compared with the 2005 through 2013 design. Among the 51 states, a required total sample size of 67,507 respondents was necessary to meet all sample size requirements while ensuring sufficient precision for key outcome measures. The larger overall sample makes it possible to get adequate precision for Hispanic and non-Hispanic black or African-American populations without any targeted oversampling of these populations. Consistent with previous surveys, the 2016 NSDUH was designed to oversample the younger age groups. Table 3.2 shows the 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, 102, and 98 percent, respectively. Only minor increases in unequal weighting resulted from not distributing the sample equally across quarters.

Table 3.2 Sample Sizes and Projected Respondents, by State and Age Group

State	State Sampling Regions (SSRs)	Total Segments	Total Selected Dwelling Units	Total Selected Persons	Age Groups for Total Respondents					
					12-17	18-25	26-34	35-49	50+	Total
Total Population	750	6,000	185,884	90,894	16,877	16,877	10,126	13,501	10,126	67,507
Northeast										
Connecticut	12	96	2,643	1,293	240	240	144	192	144	960
Maine	12	96	2,643	1,293	240	240	144	192	144	960
Massachusetts	12	96	2,643	1,293	240	240	144	192	144	960
New Hampshire	12	96	2,643	1,293	240	240	144	192	144	960
New Jersey	15	120	4,130	2,020	375	375	225	300	225	1,500
New York	30	240	9,087	4,443	825	825	495	660	495	3,300
Pennsylvania	24	192	6,609	3,231	600	600	360	480	360	2,400
Rhode Island	12	96	2,643	1,293	240	240	144	192	144	960
Vermont	12	96	2,643	1,293	240	240	144	192	144	960

(continued)

Table 3.2 Sample Sizes and Projected Respondents, by State and Age Group (continued)

State	State Sampling Regions (SSRs)	Total Segments	Total Selected Dwelling Units	Total Selected Persons	Age Groups for Total Respondents					
					12-17	18-25	26-34	35-49	50+	Total
Midwest										
Illinois	24	192	6,609	3,231	600	600	360	480	360	2,400
Indiana	12	96	2,643	1,293	240	240	144	192	144	960
Iowa	12	96	2,643	1,293	240	240	144	192	144	960
Kansas	12	96	2,643	1,293	240	240	144	192	144	960
Michigan	24	192	6,609	3,231	600	600	360	480	360	2,400
Minnesota	12	96	2,643	1,293	240	240	144	192	144	960
Missouri	12	96	2,643	1,293	240	240	144	192	144	960
Nebraska	12	96	2,643	1,293	240	240	144	192	144	960
North Dakota	12	96	2,643	1,293	240	240	144	192	144	960
Ohio	24	192	6,609	3,231	600	600	360	480	360	2,400
South Dakota	12	96	2,643	1,293	240	240	144	192	144	960
Wisconsin	12	96	2,643	1,293	240	240	144	192	144	960
South										
Alabama	12	96	2,643	1,293	240	240	144	192	144	960
Arkansas	12	96	2,643	1,293	240	240	144	192	144	960
Delaware	12	96	2,643	1,293	240	240	144	192	144	960
District of Columbia	12	96	2,643	1,293	240	240	144	192	144	960
Florida	30	240	9,087	4,443	825	825	495	660	495	3,300
Georgia	15	120	4,130	2,020	375	375	225	300	225	1,500
Kentucky	12	96	2,643	1,293	240	240	144	192	144	960
Louisiana	12	96	2,643	1,293	240	240	144	192	144	960
Maryland	12	96	2,643	1,293	240	240	144	192	144	960
Mississippi	12	96	2,643	1,293	240	240	144	192	144	960
North Carolina	15	120	4,130	2,020	375	375	225	300	225	1,500
Oklahoma	12	96	2,643	1,293	240	240	144	192	144	960
South Carolina	12	96	2,643	1,293	240	240	144	192	144	960
Tennessee	12	96	2,643	1,293	240	240	144	192	144	960
Texas	30	240	9,087	4,443	825	825	495	660	495	3,300
Virginia	15	120	4,130	2,020	375	375	225	300	225	1,500
West Virginia	12	96	2,643	1,293	240	240	144	192	144	960
West										
Alaska	12	96	2,643	1,293	240	240	144	192	144	960
Arizona	12	96	2,643	1,293	240	240	144	192	144	960
California	36	288	12,556	6,140	1,140	1,140	684	912	684	4,560
Colorado	12	96	2,643	1,293	240	240	144	192	144	960
Hawaii	12	96	2,663	1,302	242	242	145	193	145	967
Idaho	12	96	2,643	1,293	240	240	144	192	144	960
Montana	12	96	2,643	1,293	240	240	144	192	144	960
Nevada	12	96	2,643	1,293	240	240	144	192	144	960
New Mexico	12	96	2,643	1,293	240	240	144	192	144	960
Oregon	12	96	2,643	1,293	240	240	144	192	144	960
Utah	12	96	2,643	1,293	240	240	144	192	144	960
Washington	12	96	2,643	1,293	240	240	144	192	144	960
Wyoming	12	96	2,643	1,293	240	240	144	192	144	960

3.3 Fourth-Stage Sample Allocation for Each Segment

Once the desired respondent sample size for each state and age group (m_{ha}) was 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 fourth 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 fourth-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) not enough information was 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 2016 NSDUH sample are summarized in Table 3.4. For more information on the subsegmenting procedures, see Appendix E.

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	6,000	33,244	5.5
Alabama	96	569	5.9
Alaska	96	591	6.2
Arizona	96	413	4.3
Arkansas	96	489	5.1
California	288	1,505	5.2
Colorado	96	497	5.2
Connecticut	96	352	3.7
Delaware	96	454	4.7
District of Columbia	96	249	2.6
Florida	240	1,307	5.4
Georgia	120	627	5.2
Hawaii	96	330	3.4
Idaho	96	667	6.9
Illinois	192	1,126	5.9
Indiana	96	531	5.5
Iowa	96	656	6.8
Kansas	96	586	6.1
Kentucky	96	462	4.8
Louisiana	96	440	4.6
Maine	96	462	4.8
Maryland	96	354	3.7
Massachusetts	96	418	4.4
Michigan	192	1,256	6.5
Minnesota	96	424	4.4
Mississippi	96	517	5.4
Missouri	96	541	5.6
Montana	96	598	6.2
Nebraska	96	647	6.7
Nevada	96	393	4.1
New Hampshire	96	454	4.7
New Jersey	120	659	5.5
New Mexico	96	607	6.3
New York	240	1,015	4.2
North Carolina	120	745	6.2
North Dakota	96	907	9.4
Ohio	192	1,335	7.0
Oklahoma	96	562	5.9
Oregon	96	425	4.4
Pennsylvania	192	1,443	7.5
Rhode Island	96	393	4.1
South Carolina	96	553	5.8
South Dakota	96	823	8.6
Tennessee	96	347	3.6
Texas	240	1,533	6.4

(continued)

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

State	Total Segments	Cumulative Number of Map Pages per State	Average Number of Map Pages per Segment
Utah	96	450	4.7
Vermont	96	467	4.9
Virginia	120	788	6.6
Washington	96	433	4.5
West Virginia	96	588	6.1
Wisconsin	96	509	5.3
Wyoming	96	747	7.8

Table 3.4 Segment and Dwelling Unit Summary

State	Total Segments	Total Subsegmented Segments	Second-Level Subsegmented Segments	Listed Dwelling Units
Total Population	6,000	1,286	3	1,508,299
Alabama	96	3	0	21,757
Alaska	96	23	0	23,351
Arizona	96	6	0	23,088
Arkansas	96	2	0	19,889
California	288	113	1	81,538
Colorado	96	11	0	23,966
Connecticut	96	7	0	21,350
Delaware	96	5	0	24,349
District of Columbia	96	12	0	26,472
Florida	240	124	0	70,064
Georgia	120	68	1	31,782
Hawaii	96	10	0	25,569
Idaho	96	9	0	20,951
Illinois	192	61	0	58,150
Indiana	96	7	0	21,871
Iowa	96	6	0	21,736
Kansas	96	4	0	20,717
Kentucky	96	5	0	22,586
Louisiana	96	10	0	23,211
Maine	96	10	0	21,143
Maryland	96	6	0	25,809
Massachusetts	96	8	0	21,000
Michigan	192	91	0	53,721
Minnesota	96	6	0	21,204
Mississippi	96	2	0	20,667
Missouri	96	3	0	22,922
Montana	96	8	0	20,634
Nebraska	96	4	0	20,075

(continued)

Table 3.4 Segment and Dwelling Unit Summary (continued)

State	Total Segments	Total Subsegmented Segments	Second-Level Subsegmented Segments	Listed Dwelling Units
Nevada	96	5	0	25,070
New Hampshire	96	3	0	23,726
New Jersey	120	53	0	36,013
New Mexico	96	11	0	21,117
New York	240	123	1	69,141
North Carolina	120	61	0	34,704
North Dakota	96	9	0	20,566
Ohio	192	78	0	53,832
Oklahoma	96	9	0	21,440
Oregon	96	6	0	20,587
Pennsylvania	192	85	0	52,762
Rhode Island	96	8	0	22,181
South Carolina	96	5	0	23,256
South Dakota	96	10	0	19,663
Tennessee	96	10	0	21,417
Texas	240	107	0	70,068
Utah	96	5	0	20,987
Vermont	96	3	0	20,683
Virginia	120	45	0	36,459
Washington	96	8	0	20,202
West Virginia	96	1	0	22,414
Wisconsin	96	9	0	22,551
Wyoming	96	8	0	19,888

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. A detailed description of the counting and listing procedures is provided in the 2016 counting and listing general manual (RTI International, 2015).

[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 $[\ f_{ha} / (\phi_h * \lambda_{ha} * \delta_{ha})] = F_h$, known now as the driving age group (see Section 1.5 in Chapter 1), 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 Fifth-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 remains 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 fourth-stage sample, which could in turn reduce the expected fifth-stage sample size. Therefore, the reduction in the fourth-stage sample is reallocated back to the segments by applying a marginal adjustment to the fifth-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 quarterly sample (see Section 3.2), to allow for supplemental releases based on state experiences within each quarter. 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 #8) 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 will be provided in the forthcoming 2016 NSDUH sample experiences report.

3.7 Procedures for Adding Dwelling Units

To ensure that most DUs had a chance of selection and to minimize bias associated with incomplete frames, a check for missed DUs was implemented at each sampled DU. 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). When found on the property of a sampled DU, 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, and 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 F.

In addition to checking for missed DUs at each sampled DU, interviewers were instructed to call their supervisors if they noticed large differences in the segment listing and what they encountered in the field. If the interviewer identified 150 or more missed DUs in a segment or 50 or more missed DUs following any DU, special "bust" procedures were implemented to minimize bias associated with large numbers of missed DUs. The bust procedures involve selecting a subsample of the missed DUs and adding them to the interviewer's assignment and are described in more detail in Appendix F. The total number of added DUs identified during the screening interview or added through the bust procedures will be summarized in a forthcoming report on the 2016 NSDUH sample experiences.

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.
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 2015 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 3 102 percent sample then was adjusted by this amount, and an additional 20 percent sample was included.
Procedural:	Data from quarters 2 through 4 of the 2015 NSDUH and quarter 1 of the 2016 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 98 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.

Procedural: Data from quarters 3 and 4 of the 2015 NSDUH and quarters 1 and 2 of the 2016 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 person-level weights for the 2016 NSDUH had not yet been computed. However, the planned procedures are described in this section. For details on how pair weights are computed, see the report on the 2016 NSDUH questionnaire dwelling-unit level and person pair-level sampling weight calibration (Center for Behavioral Health Statistics and Quality [CBHSQ], in press b).

3.9.1 Main Study Sampling Weights

The calculation of the sampling weights will be based on the stratified, five-stage design of the study. Specifically, the person-level sampling weights will be the product of the five 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 census block group.

Stage 3: Selection of segment. Two possible adjustments exist with this stage of selection:

(1) quarter segment weight: adjusts for the number of quarters being examined (equal to one if all four quarters are included);
(2) subsegmentation inflation: by-product of counting and listing (includes up to two levels of subsegmenting).

Stage 4: Selection of DU. Two possible adjustments exist with this stage of selection:

(1) added DU: results from subsampling missed DUs; and
(2) release adjustment.

Stage 5: 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 3.5 (shown on a page following this discussion).
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 3.6.

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 fifth-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 3.7.
5. *Person-Level Nonresponse Adjustment.* This adjustment allows for the correction of weights resulting from the failure of selected sample persons to complete the interview. Respondent sample weights 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 3.7.
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 3.6.
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 fifth stage of selection.

All weight adjustments for the 2016 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 3.8. 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 16, 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 2016 NSDUH (CBHSQ, in press a).

Table 3.5 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 3.6 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 3.7 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 3.7 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 3.8 Model Group Definitions (Census Divisions)

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

Exhibit 1 Flowchart of Sample Weighting Steps

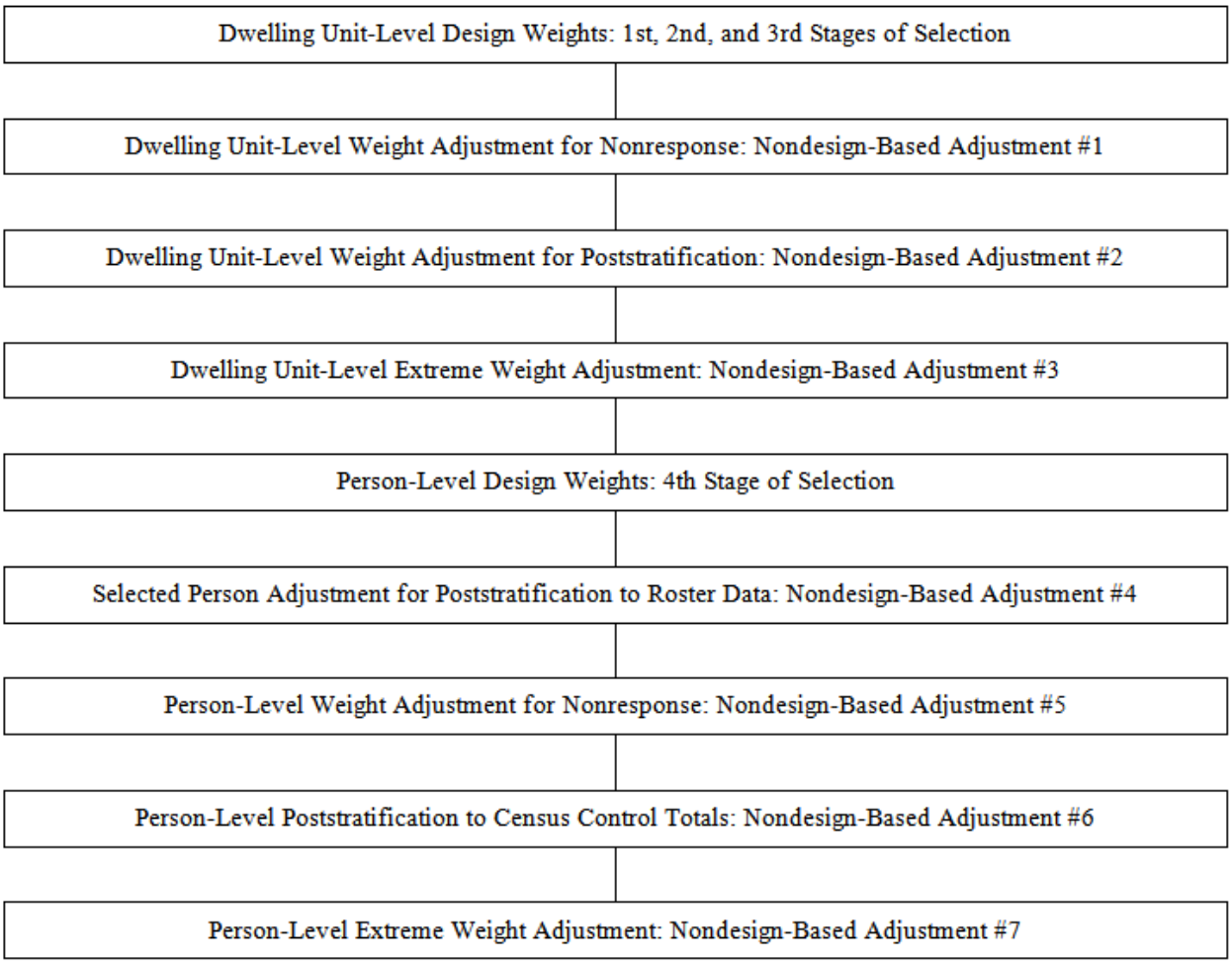


Exhibit 2 Sample Weight Components

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

3.9.2 Quality Control Measures in Design-Based Weighting Procedures

Quality control (QC) measures are applied to every component of the DU-level and person-level design weights. In addition to the QC measures outlined below, SAS® programs are examined for errors, warnings, and variable 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 2.
- Check the person-level weight to ensure that there are no missing values and the sum is close to the expected value.

This page intentionally left blank

References

Brewer, K. R. W. (1963). A model of systematic sampling with unequal probabilities. *Australian Journal of Statistics, 5*(1), 5-13. https://doi.org/10.1111/j.1467-842x.1963.tb00132.x

Brewer, K. R. W. (1975). A simple procedure for sampling πpswor. *Australian Journal of Statistics*, 17(3), 166-172. https://doi.org/10.1111/j.1467-842x.1975.tb00954.x

Center for Behavioral Health Statistics and Quality. (in press a). *2016 National Survey on Drug Use and Health: Methodological resource book (Section 11, Person-level sampling weight calibration report).* Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (in press b). *2016 National Survey on Drug Use and Health: Methodological resource book (Section 12, Questionnaire dwelling unit-level and person pair-level weight calibration).* Rockville, MD: Substance Abuse and Mental Health Services Administration.

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., & 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.

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. (2009, December 1). *OMB Bulletin No. 10-02: Update of statistical area definitions and guidance on their uses*. Washington, DC: The White House.

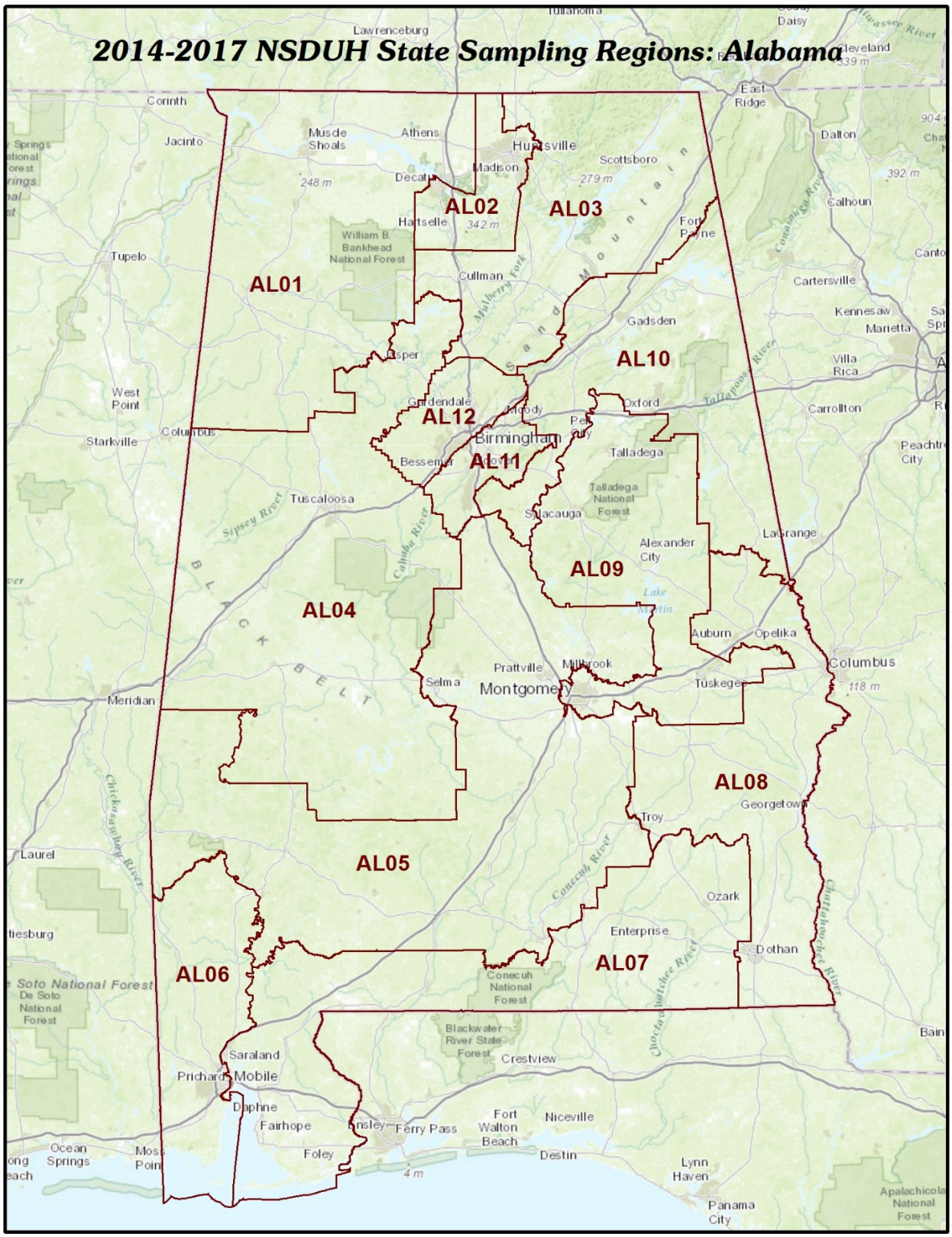
RTI International. (2012a). *National Survey on Drug Use and Health: Sample redesign issues and methodological studies* (RTI/0209009.486.001 and 0211838.108.006.005, prepared for the Substance Abuse and Mental Health Services Administration under Contract Nos. 283-2004-00022 and HHSS283200800004C). Research Triangle Park, NC: Author.

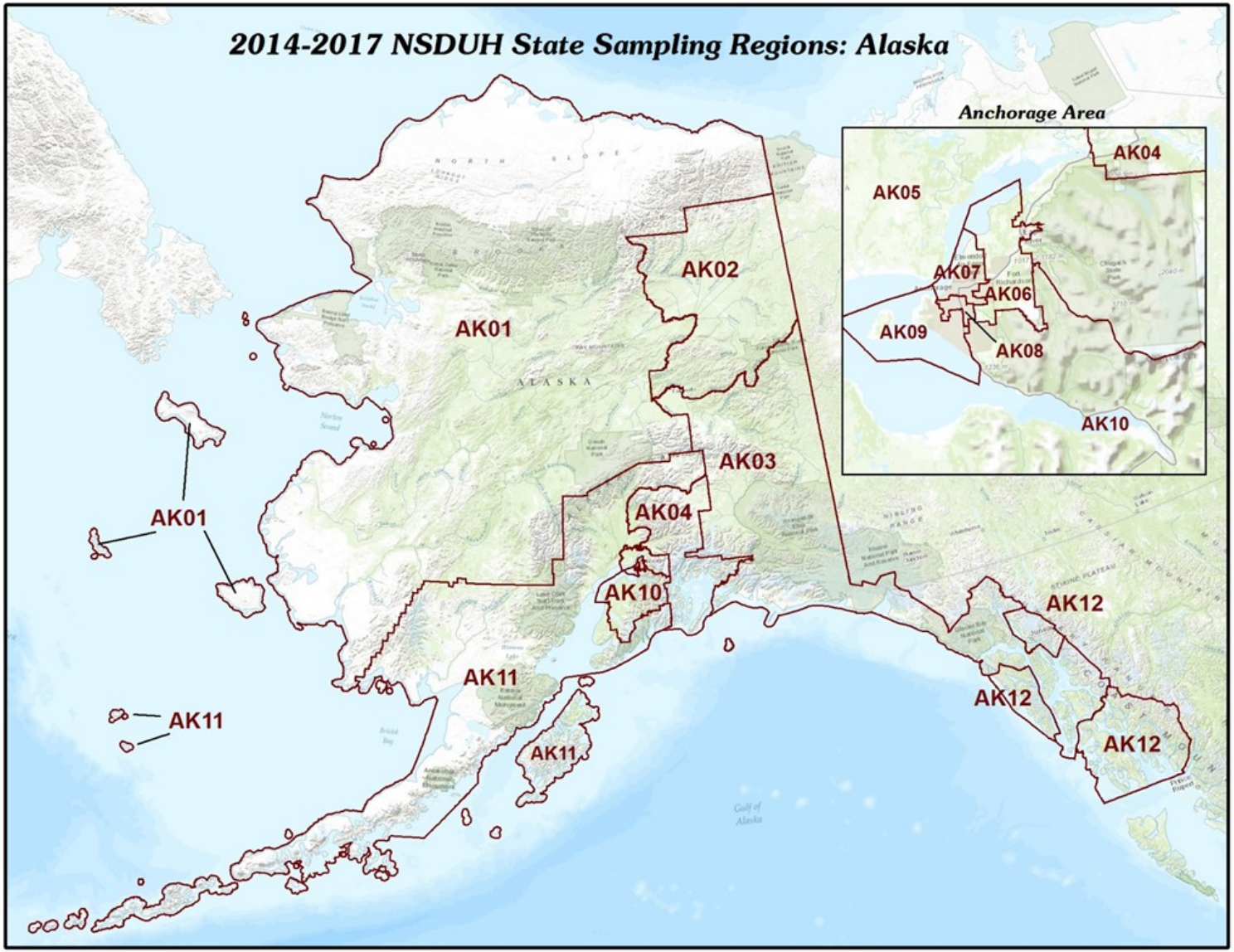
RTI International. (2012b). *SUDAAN®, Release 11.0* [computer software]. Research Triangle Park, NC: Author.

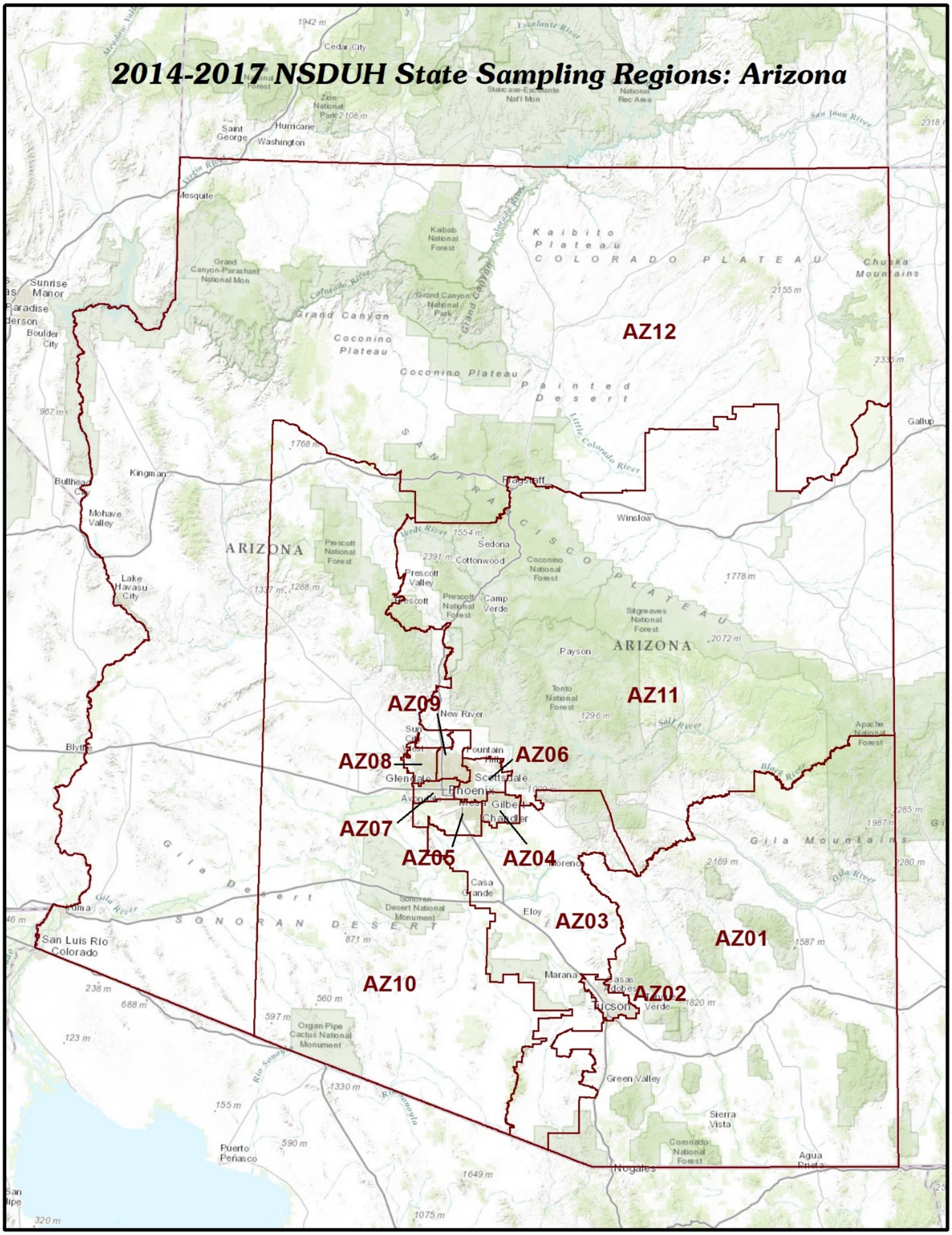
RTI International. (2015). *2016 National Survey on Drug Use and Health: Counting and listing general manual* (RTI/0213985, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283201300001C). Research Triangle Park, NC: Author.

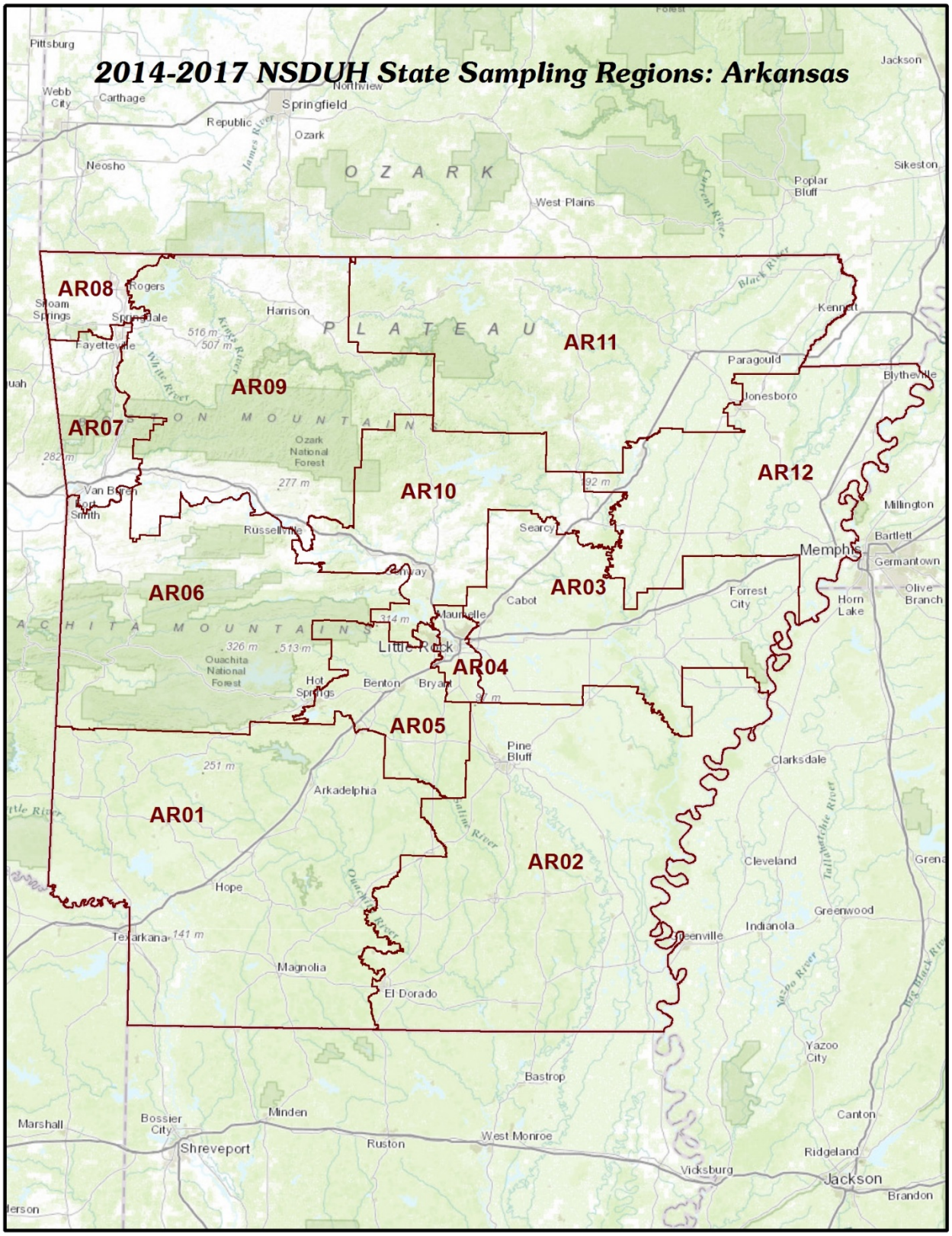
U.S. Census Bureau, Redistricting Data Office. (2009). Appendix A: Geographic terms and concepts. In *2008 Redistricting data prototype (Public Law 94-171) summary file*. Retrieved from http://www.census.gov/rdo/pdf/pl94-171.pdf

Appendix A:
2014 through 2017 NSDUH State Sampling Regions


2014-2017 NSDUH State Sampling Regions: Alabama
AL01
AL02
AL03
AL04
AL05
AL06
AL07
AL08
AL09
AL10
AL11
AL12
Lawrenceburg
Corinth
Jacinto
Muscle Shoals
Athens
Huntsville
Madison
Decatur
Scottsboro
Hartselle
William B. Bankhead National Forest
Tupelo
Cullman
Sand Mountain
Gadsden
Fort Payne
Cartersville
Kennesaw
Marietta
Villa Rica
Carrollton
Oxford
Jasper
Gardendale
Birmingham
Bessemer
Talladega
Talladega National Forest
Sylacauga
West Point
Starkville
Columbus
Tuscaloosa
Sipsey River
Black Belt
Meridian
Alexander City
Lake Martin
Auburn
Opelika
Lagrange
Peachtree City
Prattville
Selma
Montgomery
Millbrook
Tuskegee
Columbus
Georgetown
Troy
Laurel
Ozark
Enterprise
Dothan
Conecuh National Forest
Blackwater River State Forest
Crestview
Saraland
Prichard
Mobile
Daphne
Fairhope
Foley
Ferry Pass
Fort Walton Beach
Niceville
Destin
Lynn Haven
Panama City
Ocean Springs
Apalachicola National Forest

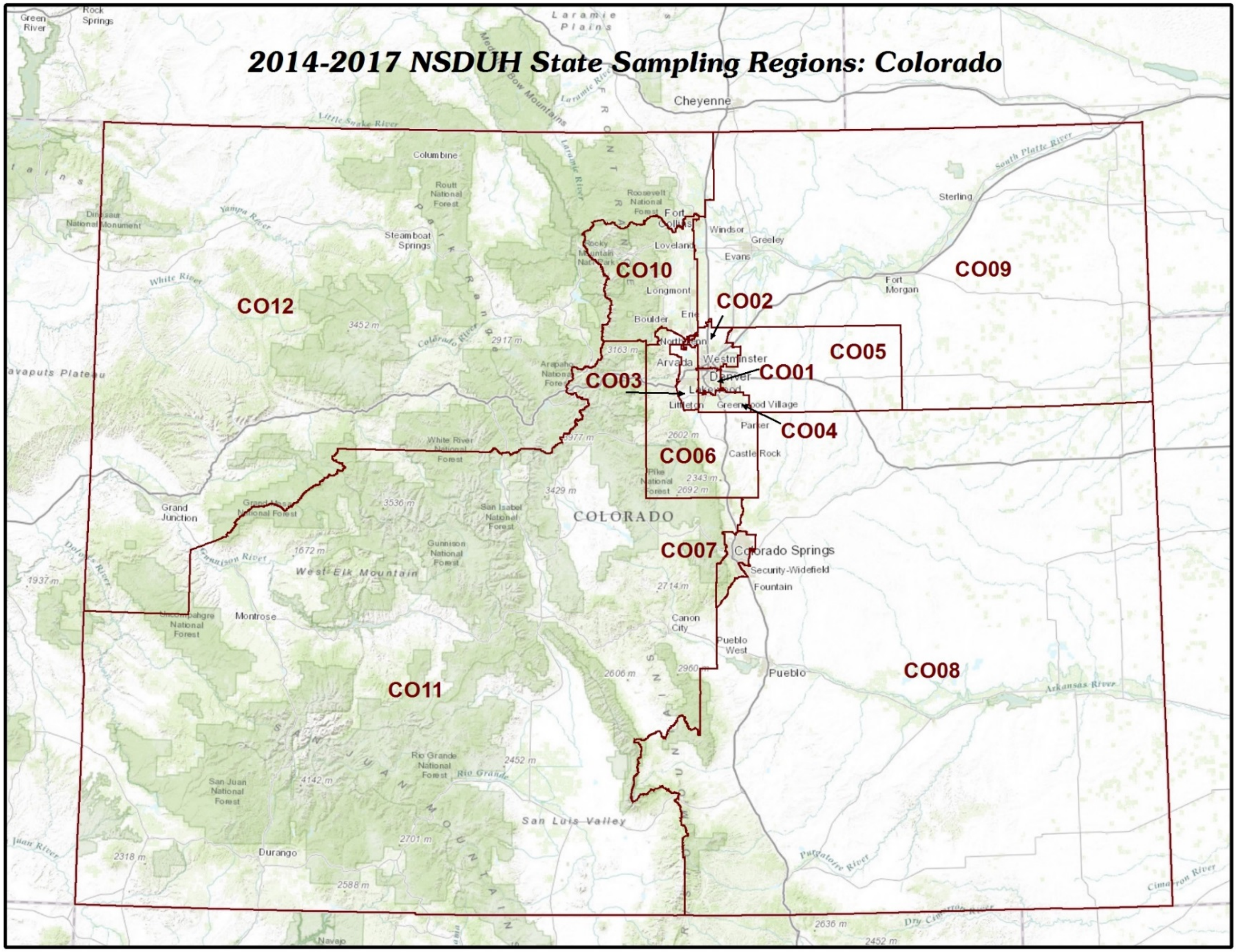

2014-2017 NSDUH State Sampling Regions: Alaska
Anchorage Area
AK01
AK02
AK03
AK04
AK05
AK06
AK07
AK08
AK09
AK10
AK11
AK12
AK01
AK11
AK11
AK10
AK12
AK12
AK12

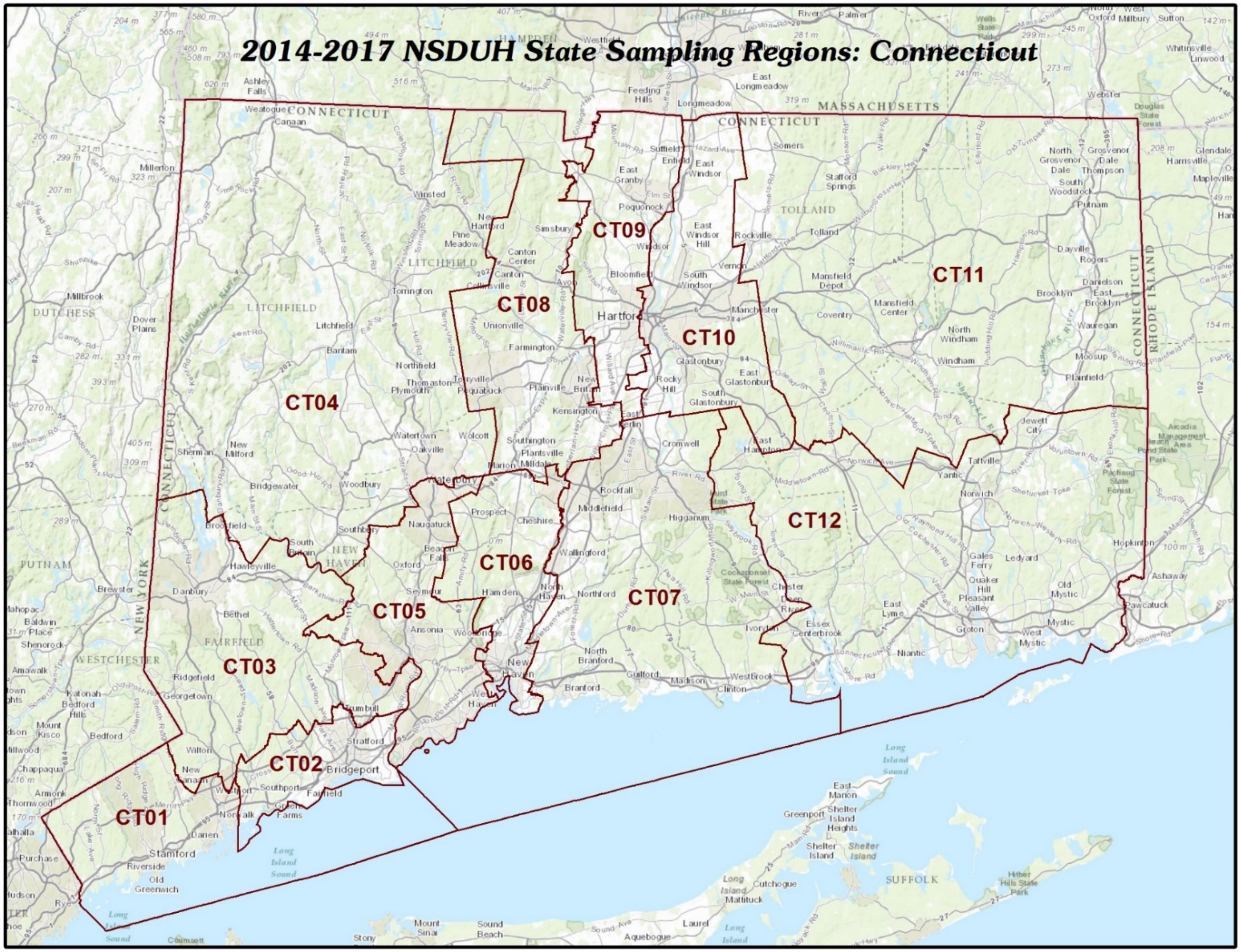

2014-2017 NSDUH State Sampling Regions: Arizona
AZ01
AZ02
AZ03
AZ04
AZ05
AZ06
AZ07
AZ08
AZ09
AZ10
AZ11
AZ12
Flagstaff
Kingman
Prescott
Sedona
Phoenix
Glendale
Scottsdale
Tucson
Yuma
Casa Grande
Green Valley
Nogales
ARIZONA

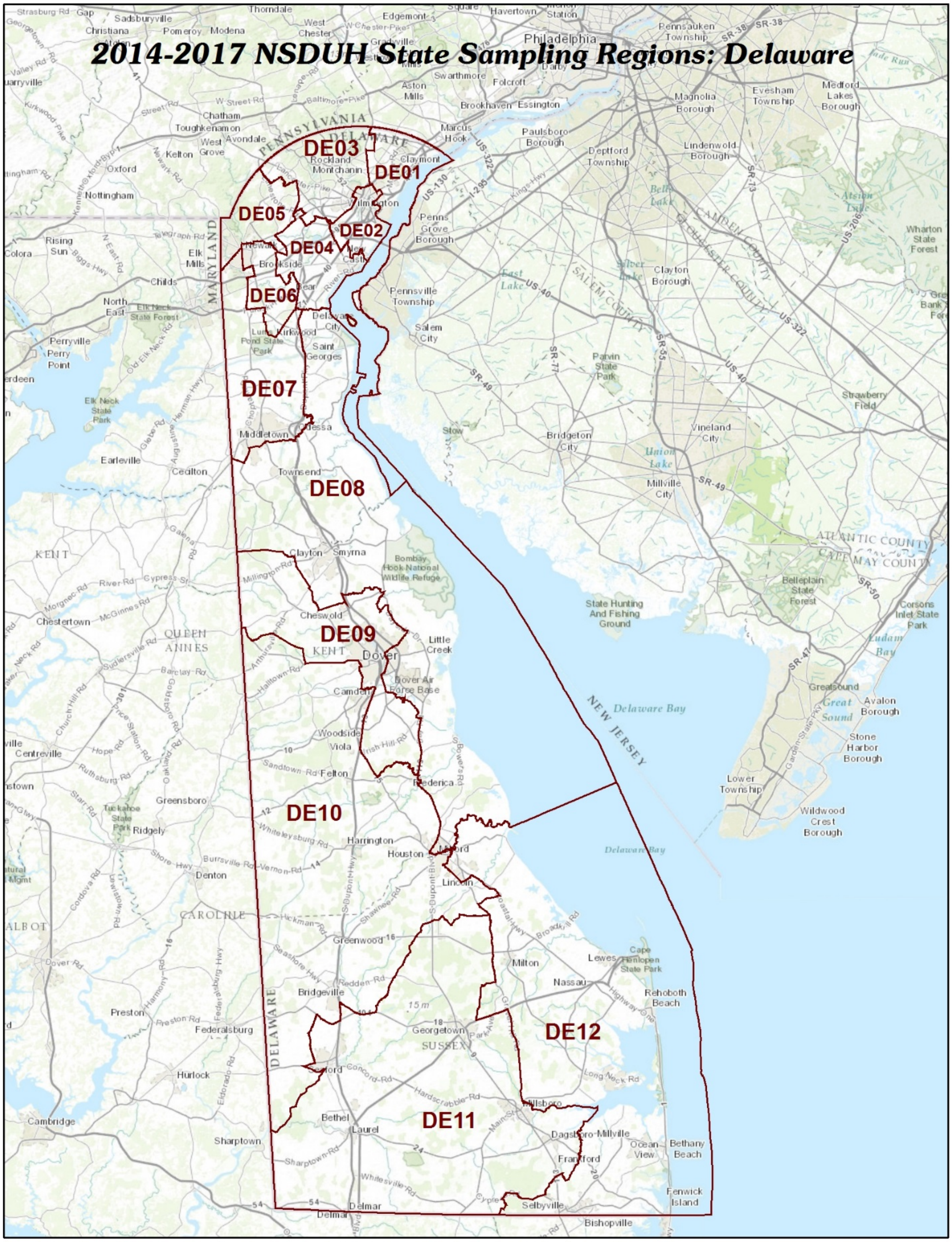

2014-2017 NSDUH State Sampling Regions: Arkansas
AR01
AR02
AR03
AR04
AR05
AR06
AR07
AR08
AR09
AR10
AR11
AR12
OZARK PLATEAU
OUACHITA MOUNTAINS
Pittsburg
Webb City
Carthage
Springfield
Republic
Ozark
Neosho
West Plains
Jackson
Sikeston
Poplar Bluff
Rogers
Harrison
Fayetteville
Ozark National Forest
Van Buren
Fort Smith
Russellville
Searcy
Paragould
Blytheville
Jonesboro
Kennett
Millington
Bartlett
Memphis
Germantown
Olive Branch
Horn Lake
Forest City
Cabot
Maumelle
Little Rock
Benton
Bryant
Hot Springs
Ouachita National Forest
Pine Bluff
Arkadelphia
Clarksdale
Cleveland
Greenwood
Indianola
Greenville
Hope
Texarkana
Magnolia
El Dorado
Bastrop
Yazoo City
Canton
Ridgeland
Jackson
Brandon
Vicksburg
Marshall
Bossier City
Shreveport
Minden
Ruston
West Monroe

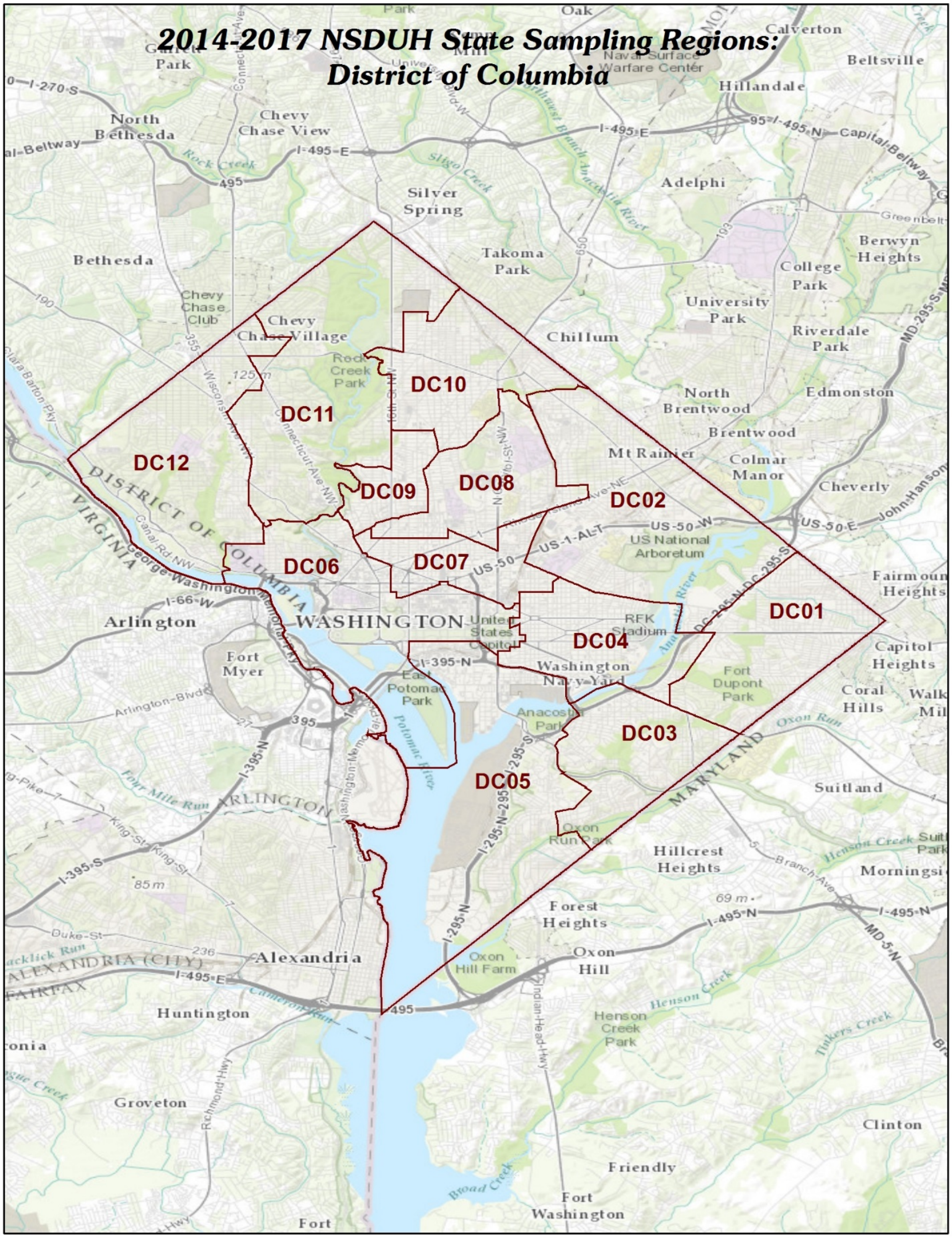

2014-2017 NSDUH State Sampling Regions: California
Los Angeles Area
San Francisco Area
CA01
CA02
CA03
CA04
CA05
CA06
CA07
CA08
CA09
CA10
CA11
CA12
CA13
CA14
CA15
CA16
CA17
CA18
CA19
CA20
CA21
CA22
CA23
CA24
CA25
CA26
CA27
CA28
CA29
CA30
CA31
CA32
CA33
CA34
CA35
CA36
OREGON
CALIFORNIA
Sacramento
San Francisco
San Jose
Los Angeles
Fresno
Bakersfield
Las Vegas
Mojave Desert
Ensenada
Tijuana
Mexicali

2014-2017 NSDUH State Sampling Regions: Colorado

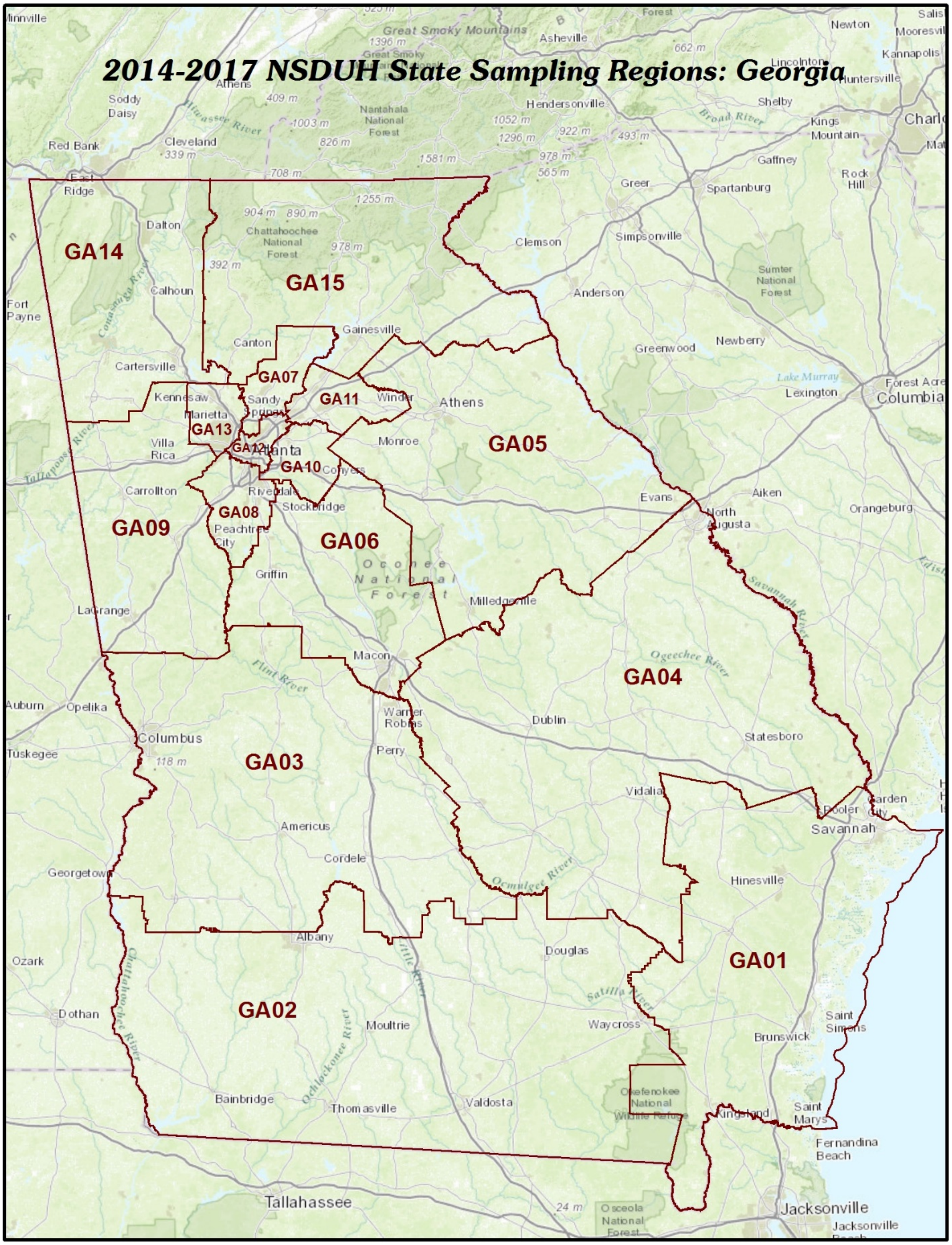


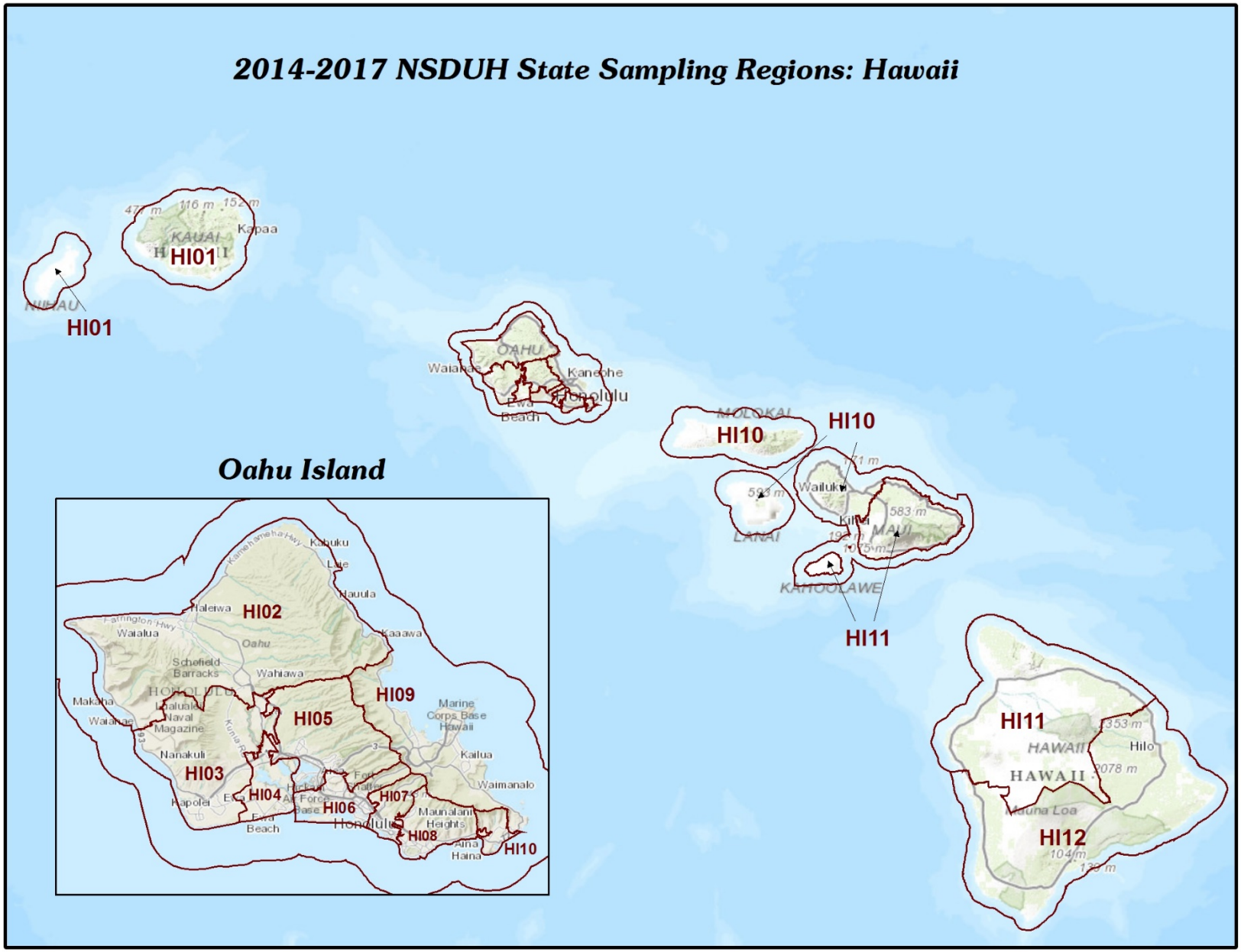

2014-2017 NSDUH State Sampling Regions: Connecticut
CT01
CT02
CT03
CT04
CT05
CT06
CT07
CT08
CT09
CT10
CT11
CT12
MASSACHUSETTS
CONNECTICUT
NEW YORK
RHODE ISLAND
LITCHFIELD
TOLLAND
FAIRFIELD
NEW HAVEN
WESTCHESTER
PUTNAM
DUTCHESS
SUFFOLK
Hartford
Bridgeport
Stamford
Danbury
Long Island Sound

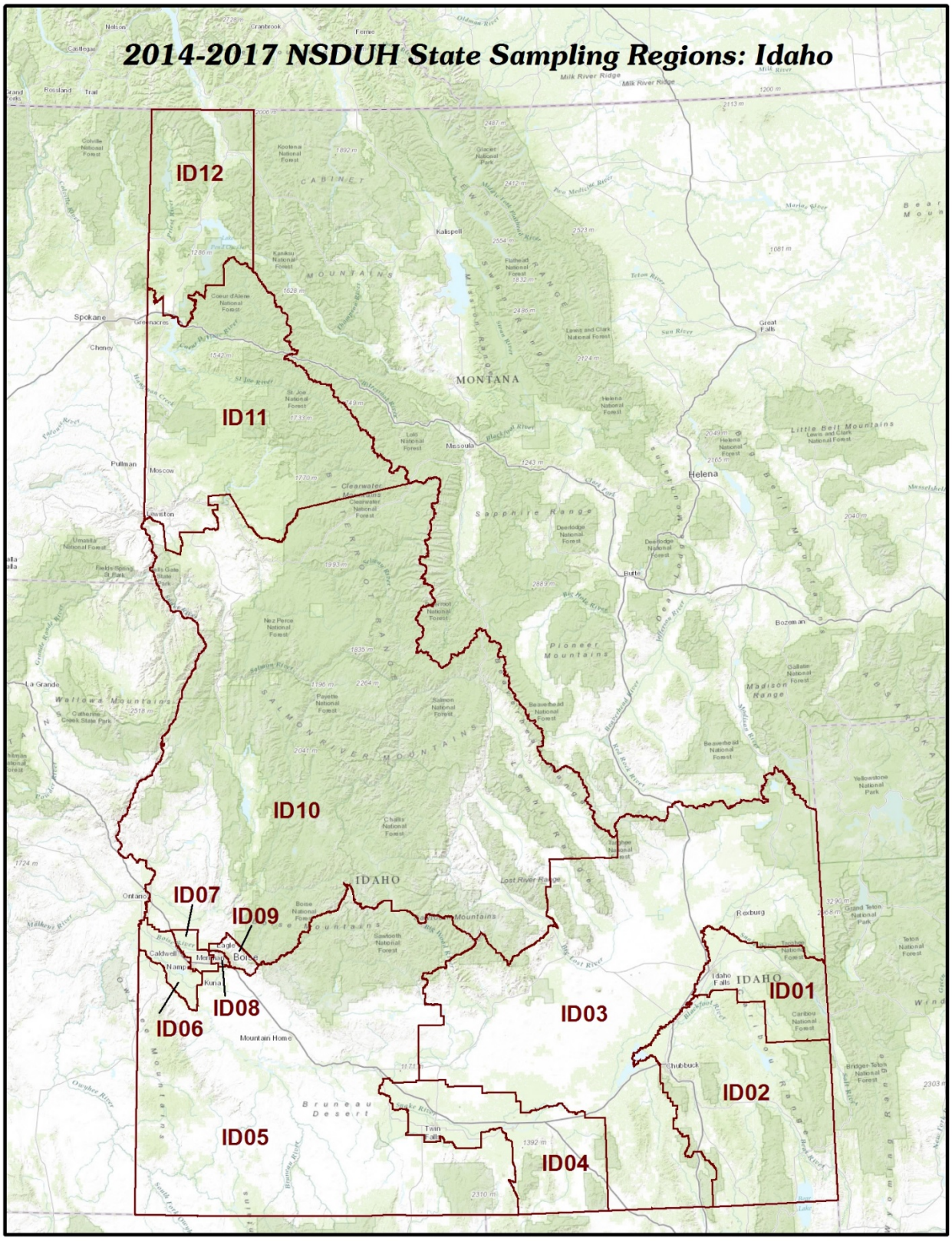

2014-2017 NSDUH State Sampling Regions: Delaware
DE01
DE02
DE03
DE04
DE05
DE06
DE07
DE08
DE09
DE10
DE11
DE12

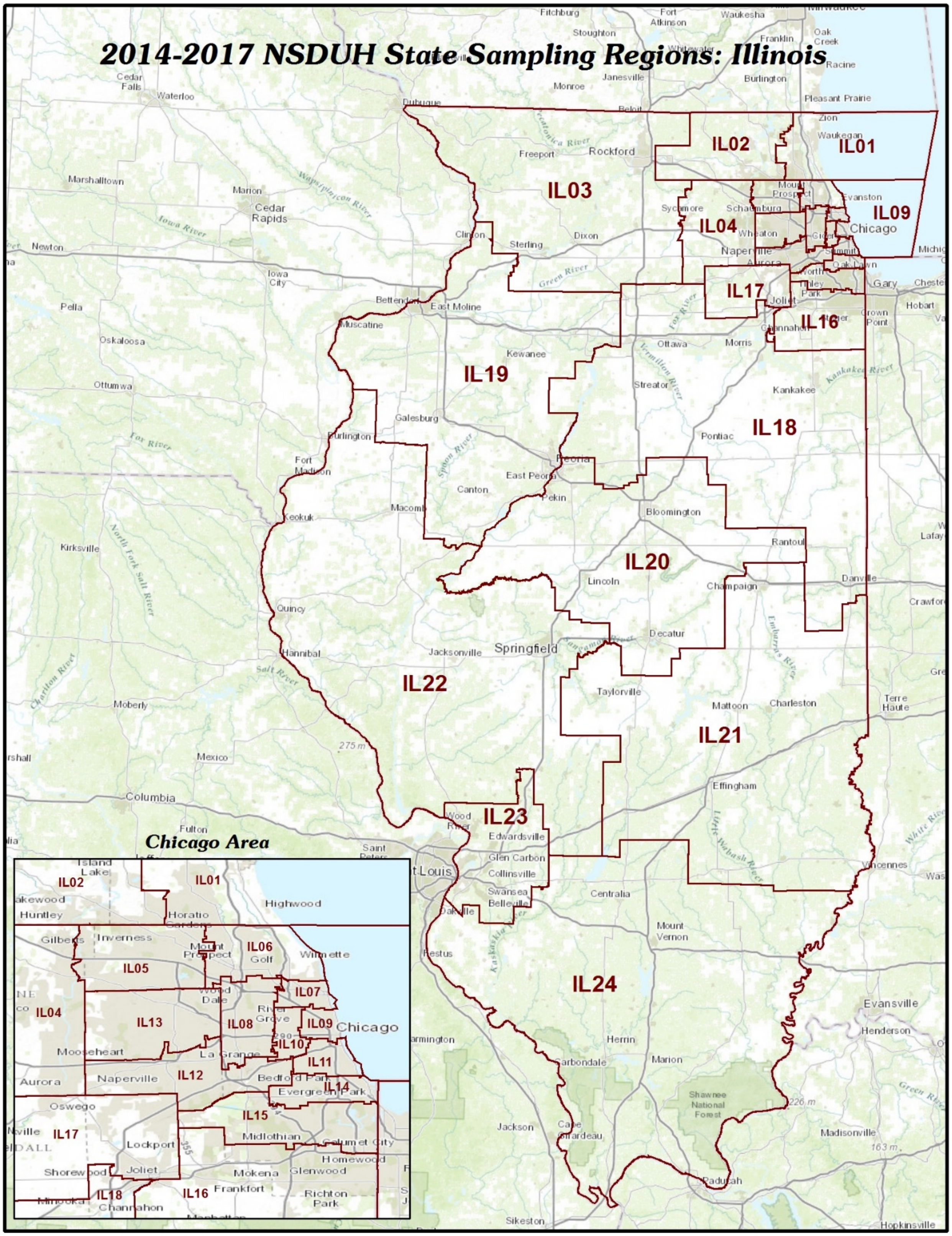

2014-2017 NSDUH State Sampling Regions:
District of Columbia
DC01
DC02
DC03
DC04
DC05
DC06
DC07
DC08
DC09
DC10
DC11
DC12
WASHINGTON
DISTRICT OF COLUMBIA
VIRGINIA
MARYLAND
ARLINGTON
Arlington
Alexandria
Bethesda
North Bethesda
Silver Spring
Takoma Park
Chillum
Adelphi
College Park
Berwyn Heights
University Park
Riverdale Park
Edmonston
North Brentwood
Brentwood
Mt Rainier
Colmar Manor
Cheverly
Fairmount Heights
Capitol Heights
Coral Hills
Suitland
Hillcrest Heights
Forest Heights
Oxon Hill
Huntington
Groveton
Friendly
Fort Washington
Clinton
Fort Myer
Chevy Chase Village
Chevy Chase View
Rock Creek Park
US National Arboretum
RFK Stadium
United States Capitol
Washington Navy Yard
East Potomac Park
Anacostia Park
Fort Dupont Park
Oxon Run Park
Oxon Hill Farm
Henson Creek Park
Potomac River

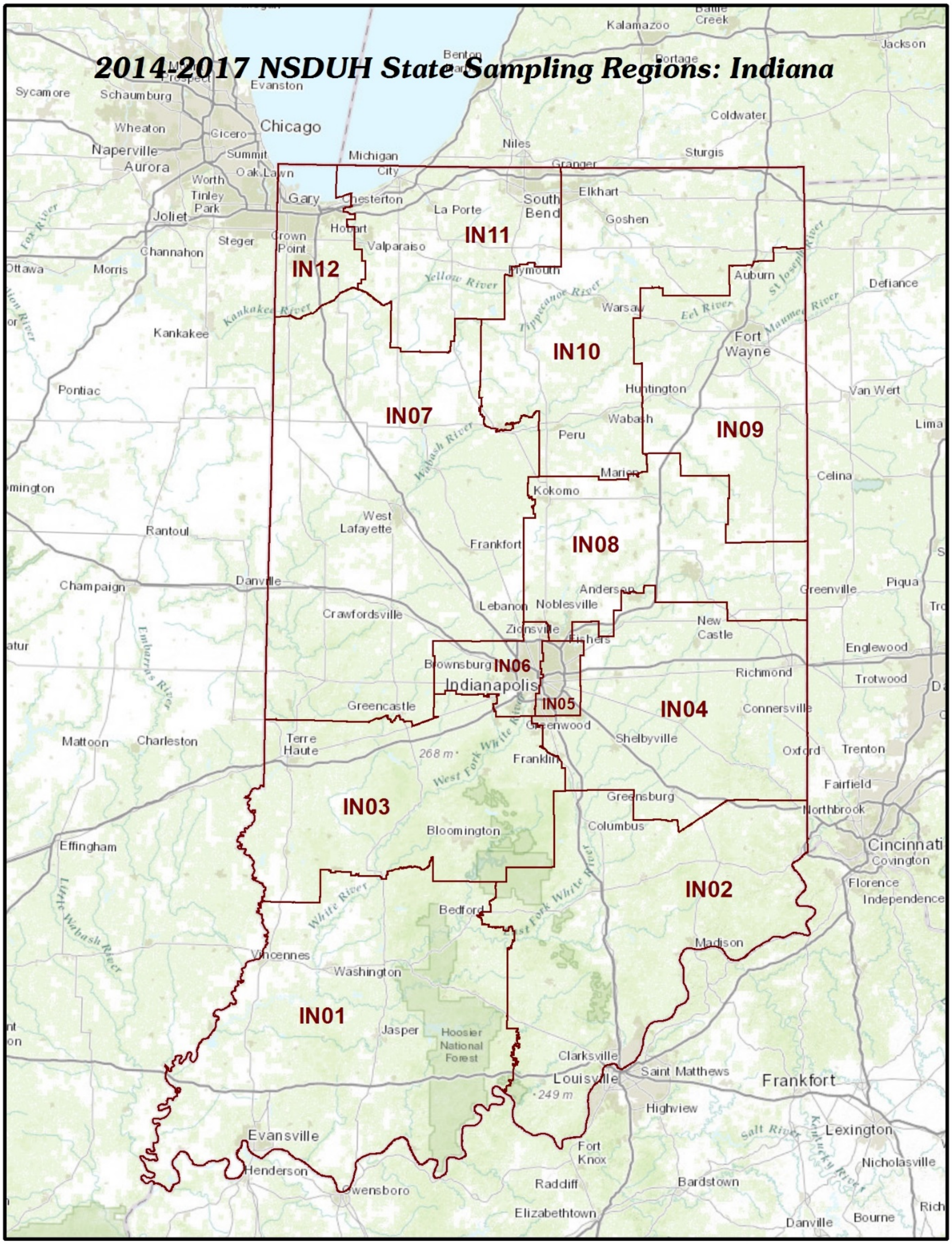

2014-2017 NSDUH State Sampling Regions: Florida
FL01
FL02
FL03
FL04
FL05
FL06
FL07
FL08
FL09
FL10
FL11
FL12
FL15
FL16
FL17
FL18
FL19
FL20
FL21
FL22
FL23
FL27
FL28
FL30
Tampa Area
Miami Area
Orlando Area
Montgomery
Columbus
Warner Robins
Savannah
Dothan
Valdosta
Pensacola
Tallahassee
Jacksonville
Gainesville
Palm Coast
Ocala National Forest
FLORIDA
Orlando
Melbourne
Palm Bay
Saint Petersburg
Tampa
Port Saint Lucie
Cape Coral
West Palm Beach
Pompano Beach
Boca Raton
Coral Springs
Everglades National Park
Miami
Okeechobee
Palm Beach Gardens
Belle Glade
Lake Worth
Delray Beach
Lauderhill
Fort Lauderdale
Hollywood
Big Cypress National Preserve
Carol City
North Miami
Hialeah
South Miami Heights
Daytona Beach
DeLand
Edgewater
Lady Lake
DeBary
Mount Dora
Lake Mary
Titusville
Clermont
Kissimmee
Saint Cloud
Cocoa
Cape Canaveral
Cocoa Beach
Santa Clara
Haines City
Lake Wales
Sancti Spiritus
Hudson
New Port Richey
Temple Terrace
Clearwater
Pinellas Park
Hill
Palmetto
Peace River

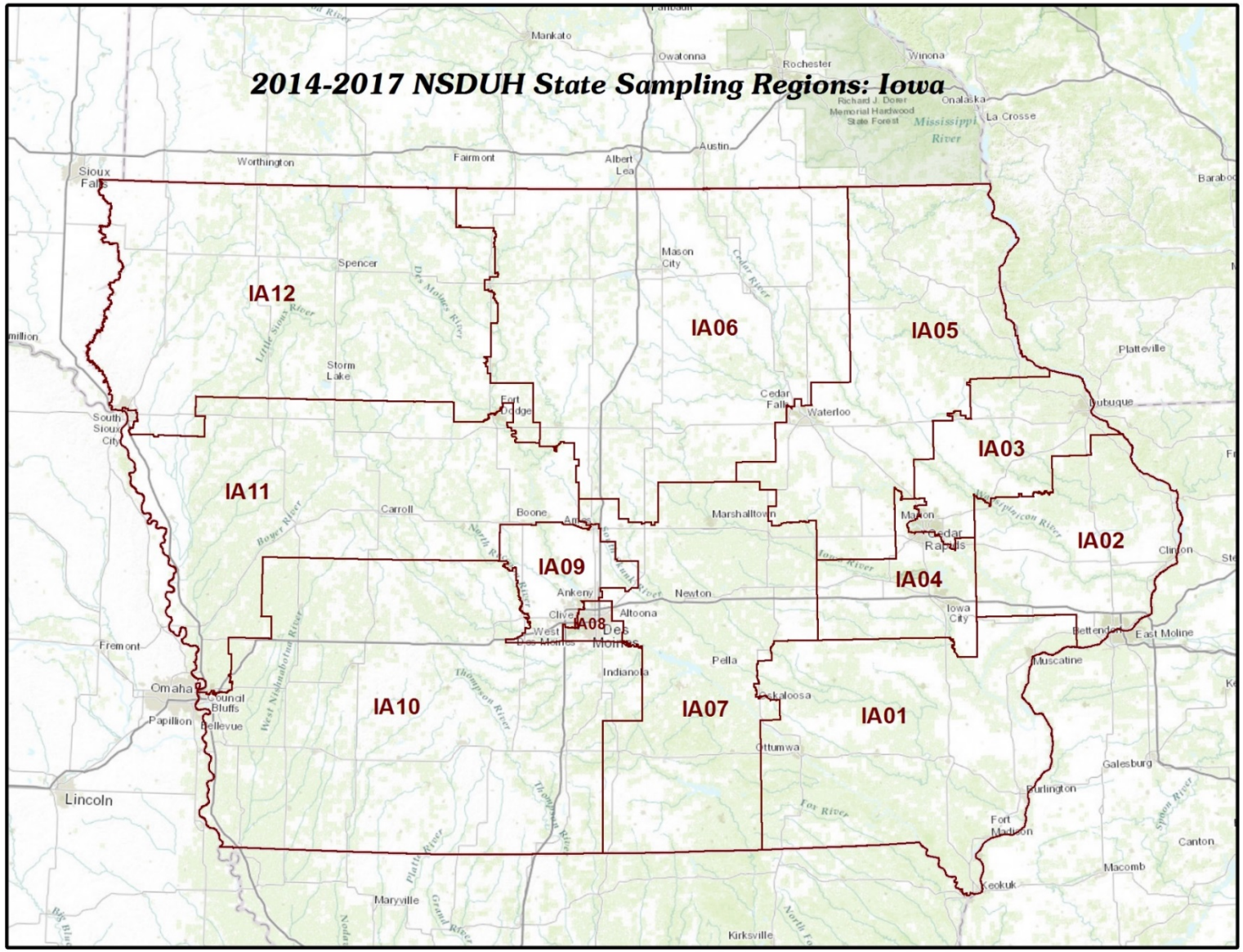

2014-2017 NSDUH State Sampling Regions: Georgia
GA01
GA02
GA03
GA04
GA05
GA06
GA07
GA08
GA09
GA10
GA11
GA12
GA13
GA14
GA15

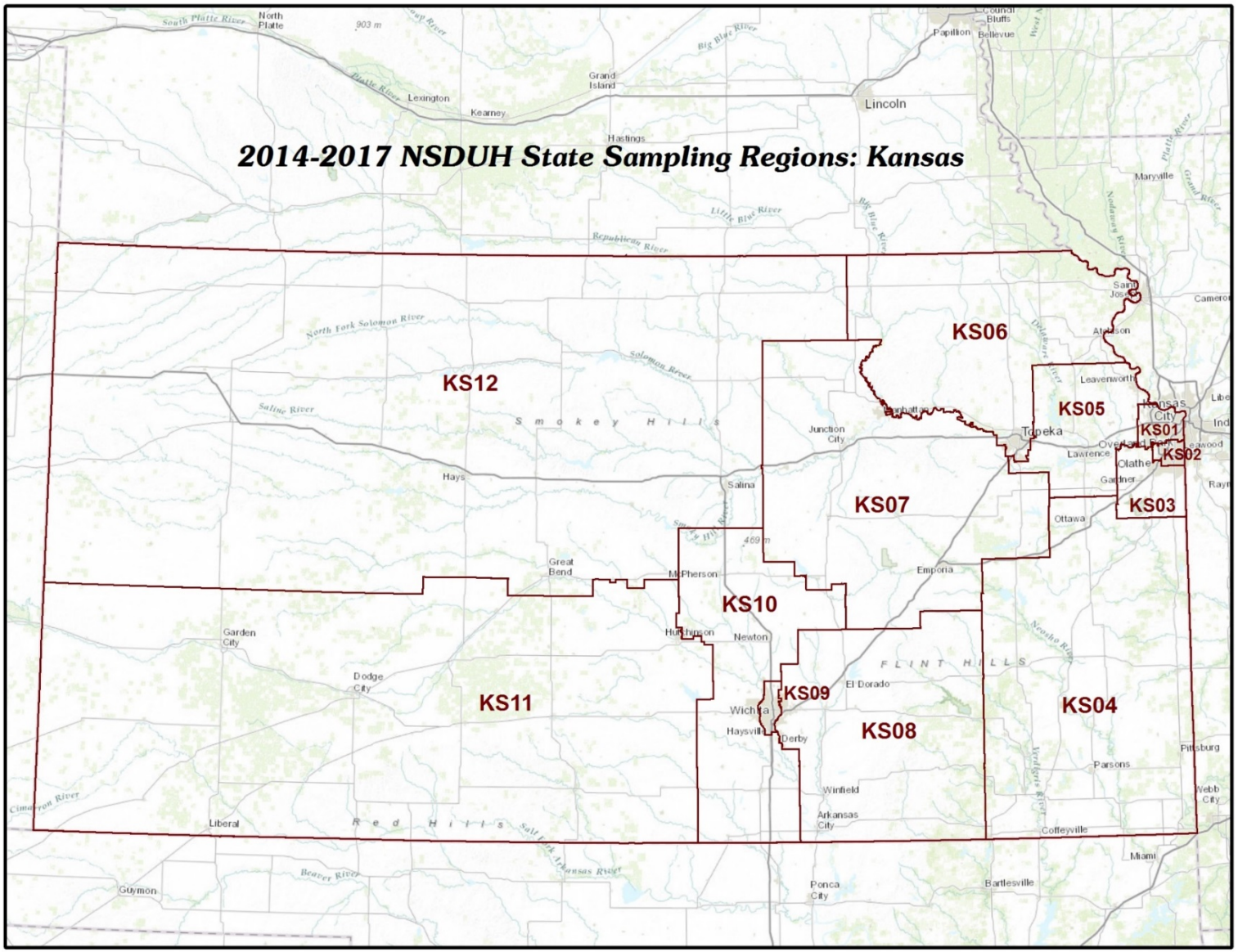

2014-2017 NSDUH State Sampling Regions: Hawaii
HI01
HI01
HI10
HI10
HI11
HI11
HI12
Oahu Island
HI02
HI03
HI04
HI05
HI06
HI07
HI08
HI09
HI10
KAUAI
NIIHAU
OAHU
MOLOKAI
LANAI
KAHOOLAWE
MAUI
HAWAII
Kapaa
Waianae
Kaneohe
Honolulu
Ewa Beach
Wailuku
Kihei
Hilo
Mauna Loa
Haleiwa
Waialua
Wahiawa
Schofield Barracks
Kaaawa
Kahuku
Laie
Hauula
Makaha
Waianae
Nanakuli
Kailua
Waimanalo
Maunalani Heights
Aina Haina
Hapolei
Ewa Beach
Marine Corps Base Hawaii
Kamehameha Hwy
Farrington Hwy
Kunia Rd

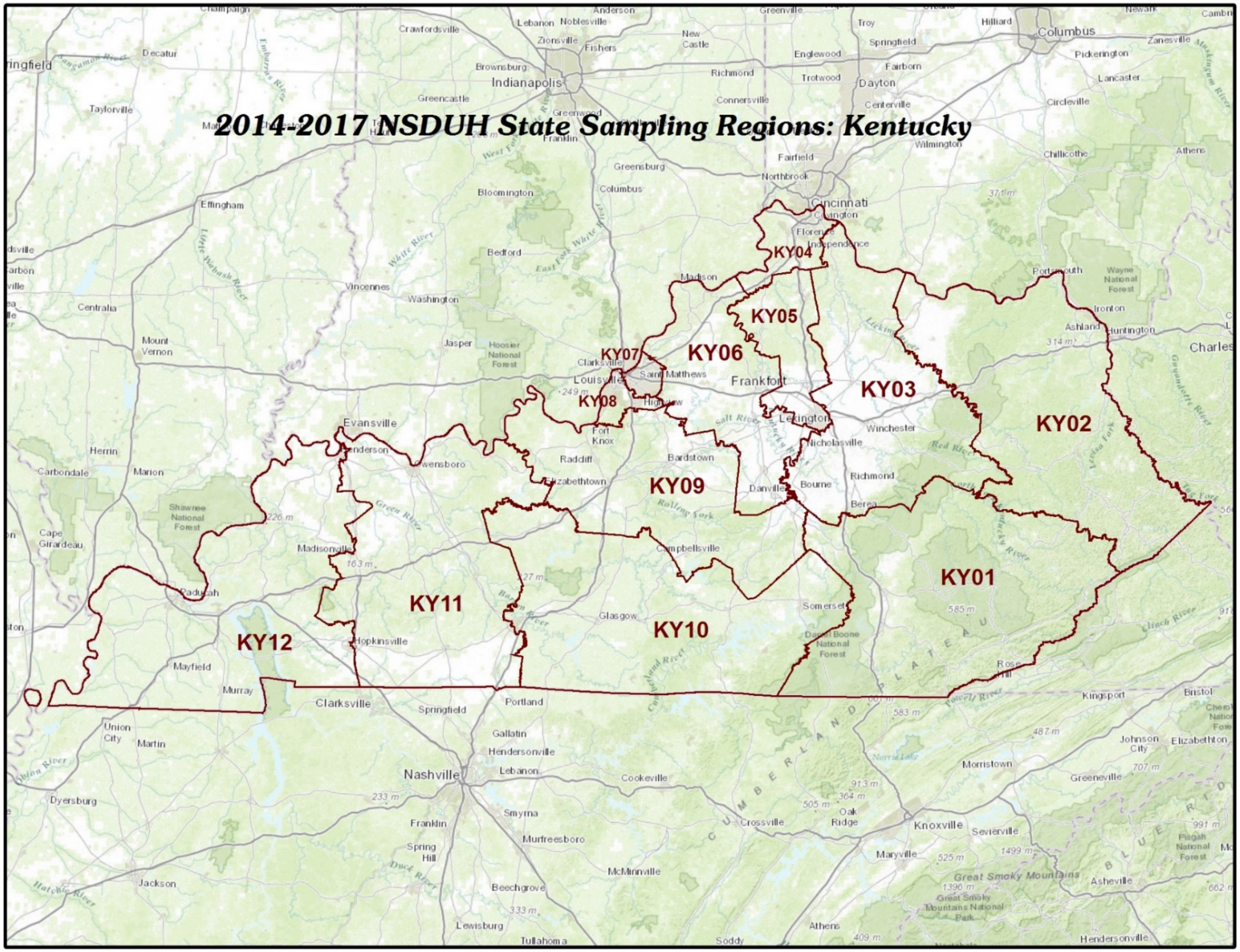

2014-2017 NSDUH State Sampling Regions: Idaho
ID12
ID11
ID10
ID07
ID09
ID08
ID06
ID05
ID03
ID01
ID02
ID04
MONTANA
IDAHO
Spokane
Cheney
Pullman
Moscow
Lewiston
Kalispell
Missoula
Helena
Great Falls
Butte
Bozeman
Boise
Nampa
Caldwell
Meridian
Eagle
Kuna
Mountain Home
Idaho Falls
Rexburg
Chubbuck
Twin Falls
La Grande
Ontario
Sapphire Range
Pioneer Mountains
Lemhi Range
Lost River Range
Bitterroot Range
Salmon River Mountains
Owyhee Mountains
Bruneau Desert
Snake River

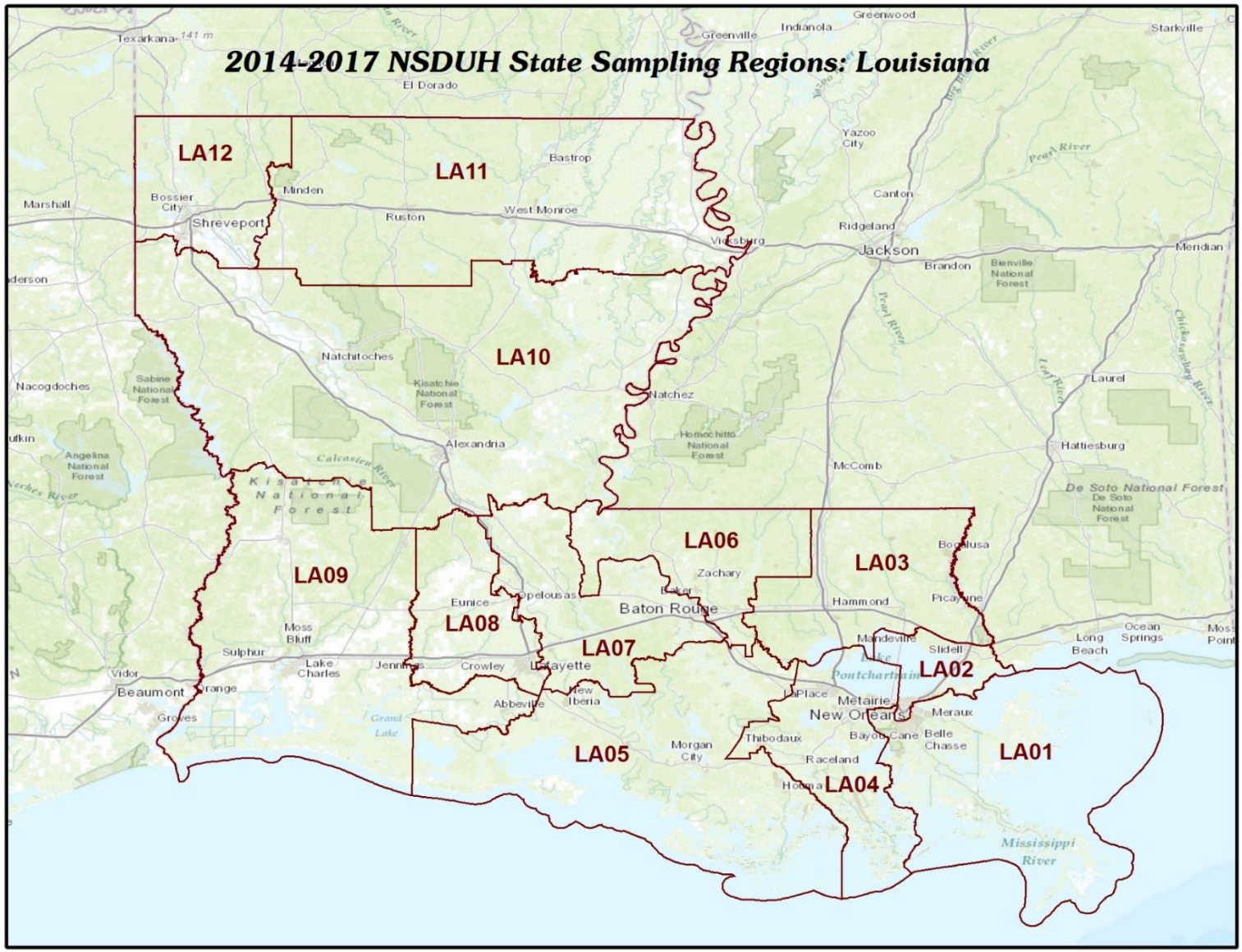

2014-2017 NSDUH State Sampling Regions: Illinois
IL01
IL02
IL03
IL04
IL09
IL16
IL17
IL18
IL19
IL20
IL21
IL22
IL23
IL24
Chicago Area
IL01
IL02
IL04
IL05
IL06
IL07
IL08
IL09
IL10
IL11
IL12
IL13
IL14
IL15
IL16
IL17
IL18

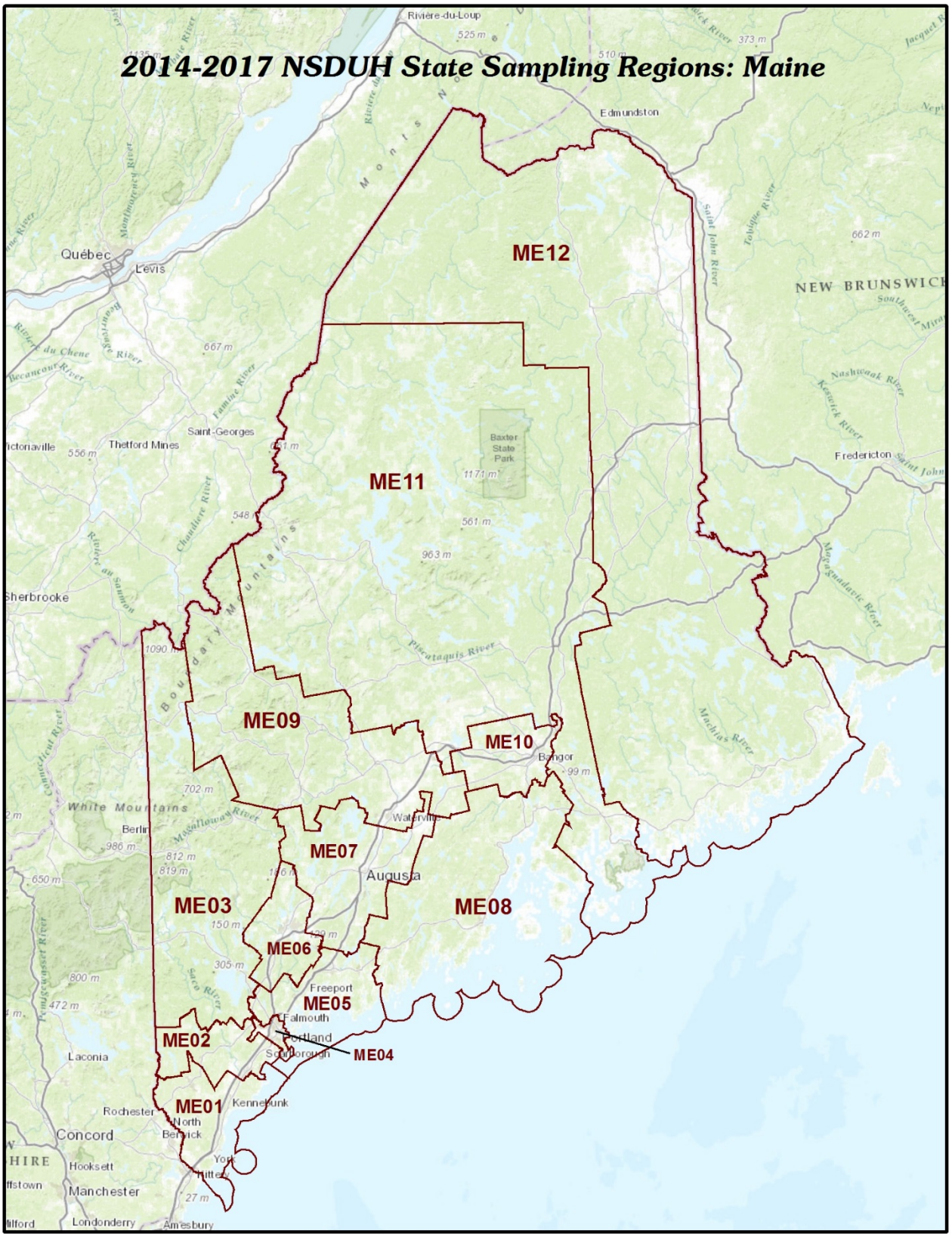

2014-2017 NSDUH State Sampling Regions: Indiana
IN01
IN02
IN03
IN04
IN05
IN06
IN07
IN08
IN09
IN10
IN11
IN12
Chicago
Gary
South Bend
Fort Wayne
Kokomo
West Lafayette
Indianapolis
Terre Haute
Bloomington
Evansville
Louisville
Cincinnati
Lexington
Frankfort

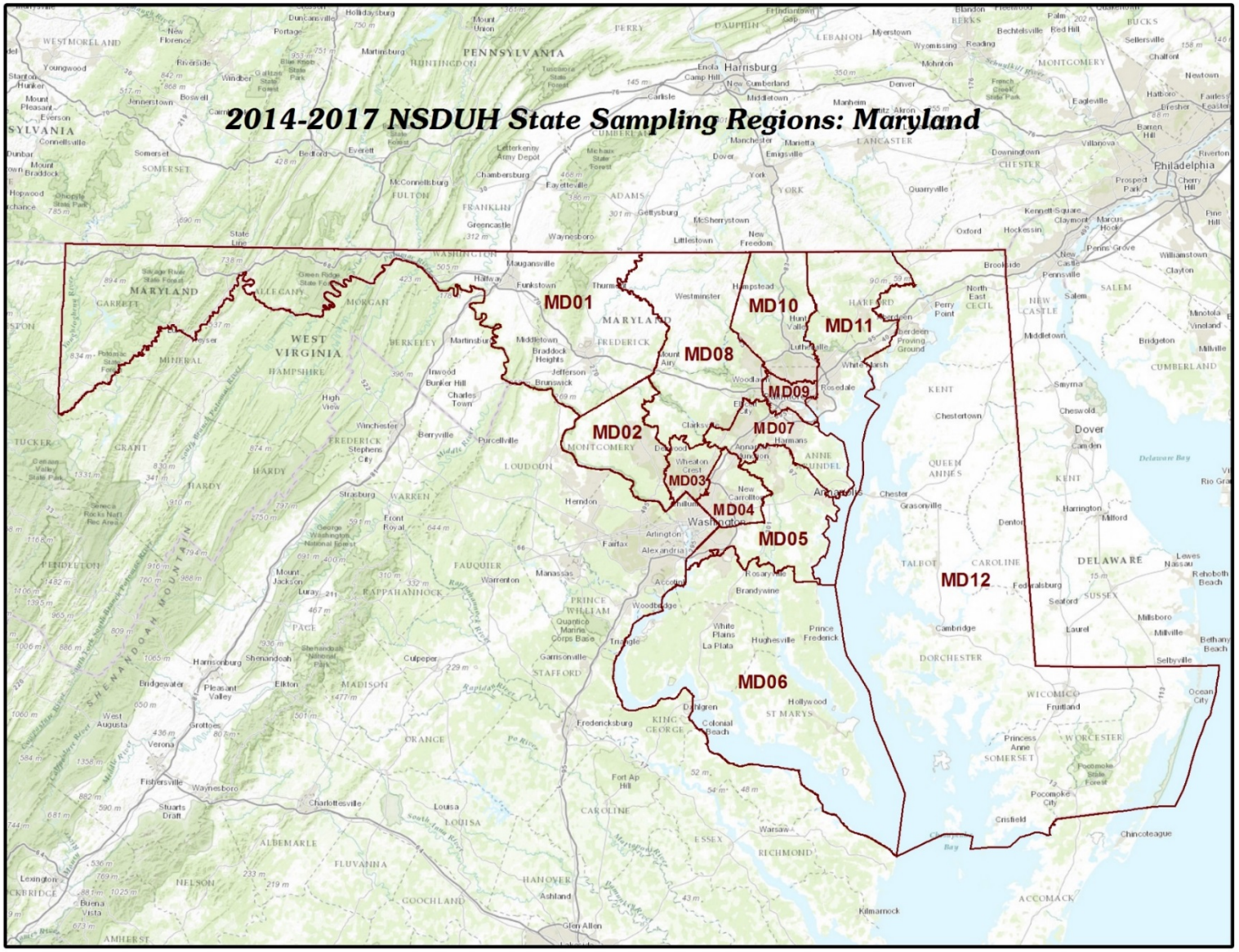

2014-2017 NSDUH State Sampling Regions: Iowa
IA12
IA06
IA05
IA11
IA03
IA02
IA09
IA04
IA08
IA10
IA07
IA01
Mankato
Owatonna
Rochester
Winona
Worthington
Fairmont
Albert Lea
Austin
La Crosse
Sioux Falls
Spencer
Mason City
Storm Lake
Fort Dodge
Cedar Falls
Waterloo
Dubuque
Platteville
South Sioux City
Carroll
Boone
Ames
Marshalltown
Marion
Cedar Rapids
Clinton
Ankeny
Newton
Iowa City
Altoona
Clive
West Des Moines
Des Moines
Bettendorf
East Moline
Muscatine
Fremont
Omaha
Council Bluffs
Papillion
Bellevue
Indianola
Pella
Oskaloosa
Ottumwa
Galesburg
Burlington
Fort Madison
Keokuk
Canton
Macomb
Lincoln
Maryville
Kirksville

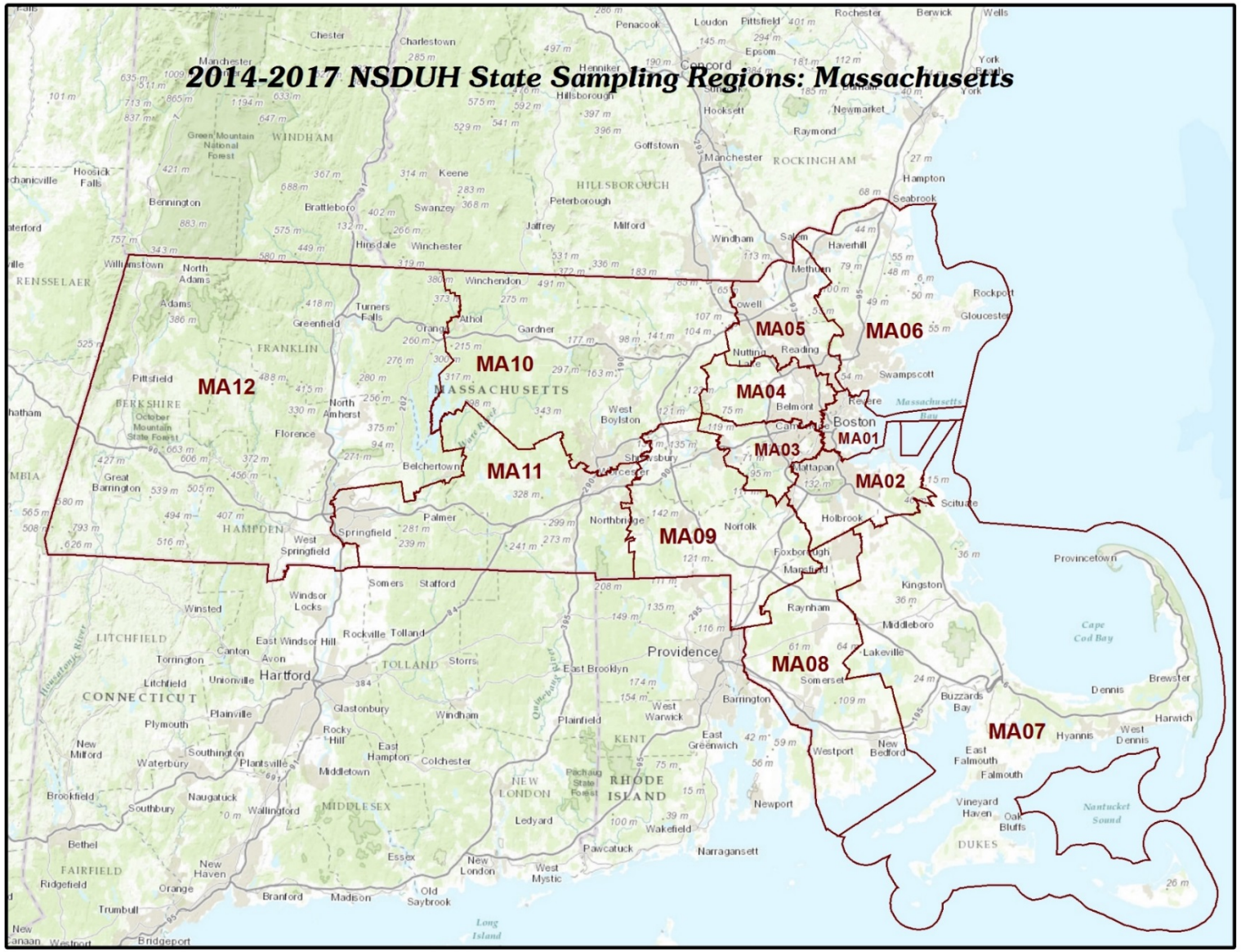

2014-2017 NSDUH State Sampling Regions: Kansas
KS01
KS02
KS03
KS04
KS05
KS06
KS07
KS08
KS09
KS10
KS11
KS12
Lincoln
Topeka
Kansas City
Wichita
Salina
Hays
Dodge City
Garden City
Liberal
Great Bend
McPherson
Hutchinson
Newton
El Dorado
Emporia
Ottawa
Olathe
Lawrence
Leavenworth
Atchison
Junction City
Manhattan
Arkansas City
Winfield
Parsons
Coffeyville
Bartlesville
Ponca City
Guymon
Smokey Hills
Red Hills
Flint Hills

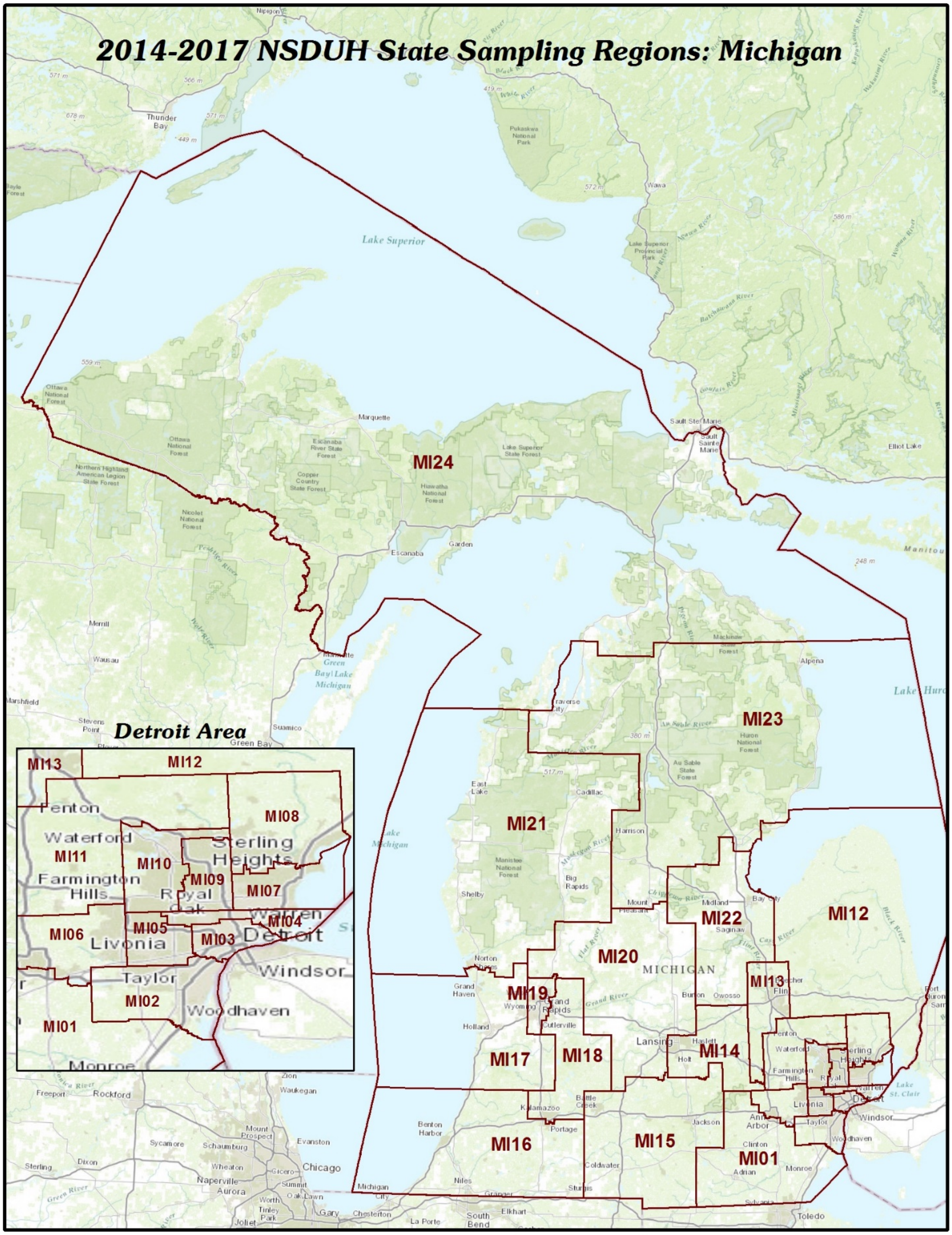

2014-2017 NSDUH State Sampling Regions: Kentucky
KY01
KY02
KY03
KY04
KY05
KY06
KY07
KY08
KY09
KY10
KY11
KY12

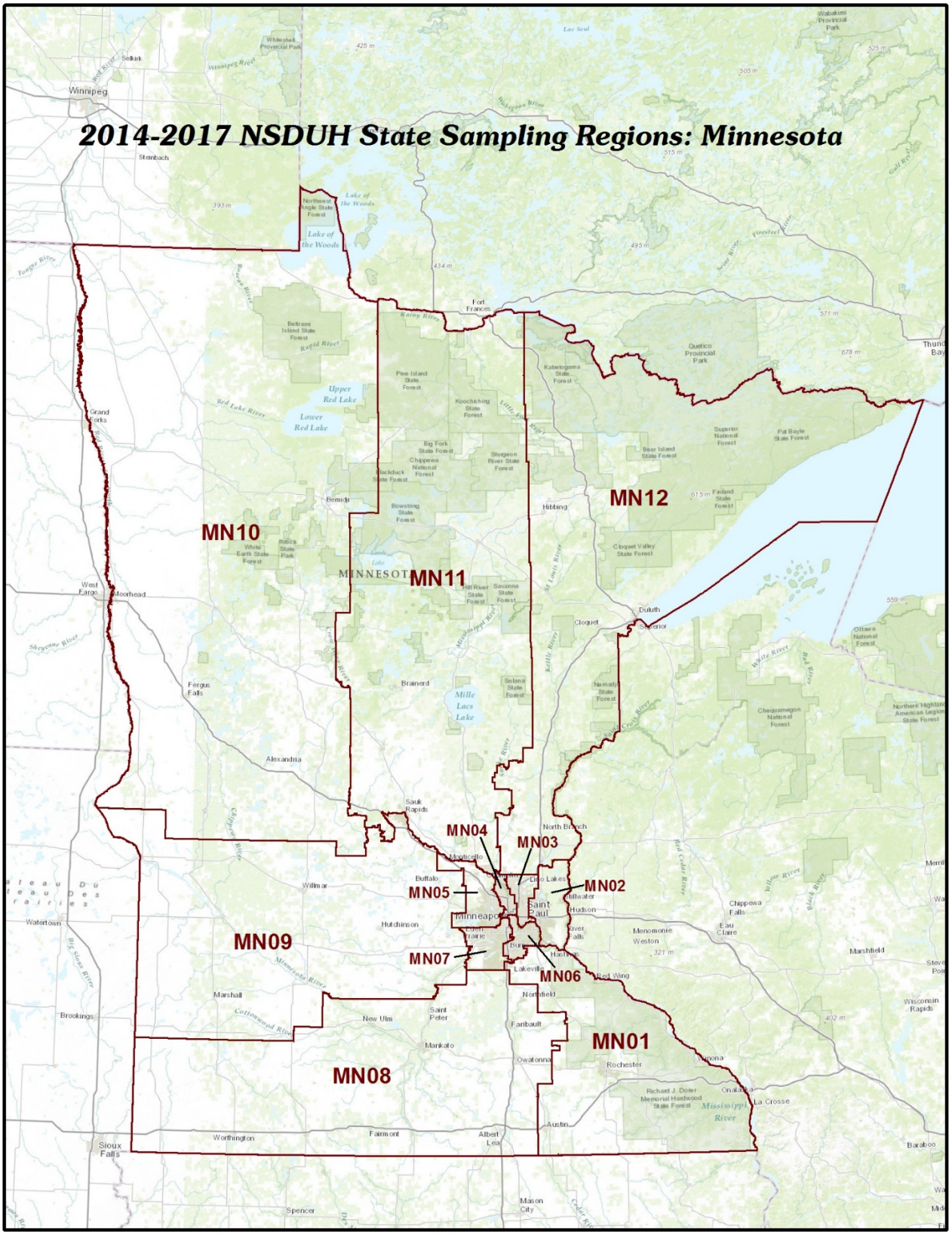

2014-2017 NSDUH State Sampling Regions: Louisiana
LA12
LA11
LA10
LA09
LA08
LA07
LA06
LA05
LA04
LA03
LA02
LA01
Texarkana
El Dorado
Bastrop
Minden
Bossier City
Shreveport
Marshall
Ruston
West Monroe
Vicksburg
Greenville
Indianola
Greenwood
Starkville
Yazoo City
Canton
Ridgeland
Jackson
Brandon
Meridian
Natchitoches
Nacogdoches
Sabine National Forest
Kisatchie National Forest
Natchez
Homochitto National Forest
Alexandria
Angelina National Forest
Calcasieu River
Laurel
Hattiesburg
McComb
De Soto National Forest
Zachary
Baker
Baton Rouge
Opelousas
Eunice
Hammond
Picayune
Bogalusa
Moss Bluff
Sulphur
Lake Charles
Jennings
Crowley
Lafayette
New Iberia
Abbeville
Vidor
Beaumont
Orange
Groves
Grand Lake
Mandeville
Slidell
Long Beach
Ocean Springs
Pontchartrain
LaPlace
Metairie
New Orleans
Meraux
Thibodaux
Bayou Cane
Belle Chasse
Morgan City
Raceland
Houma
Mississippi River

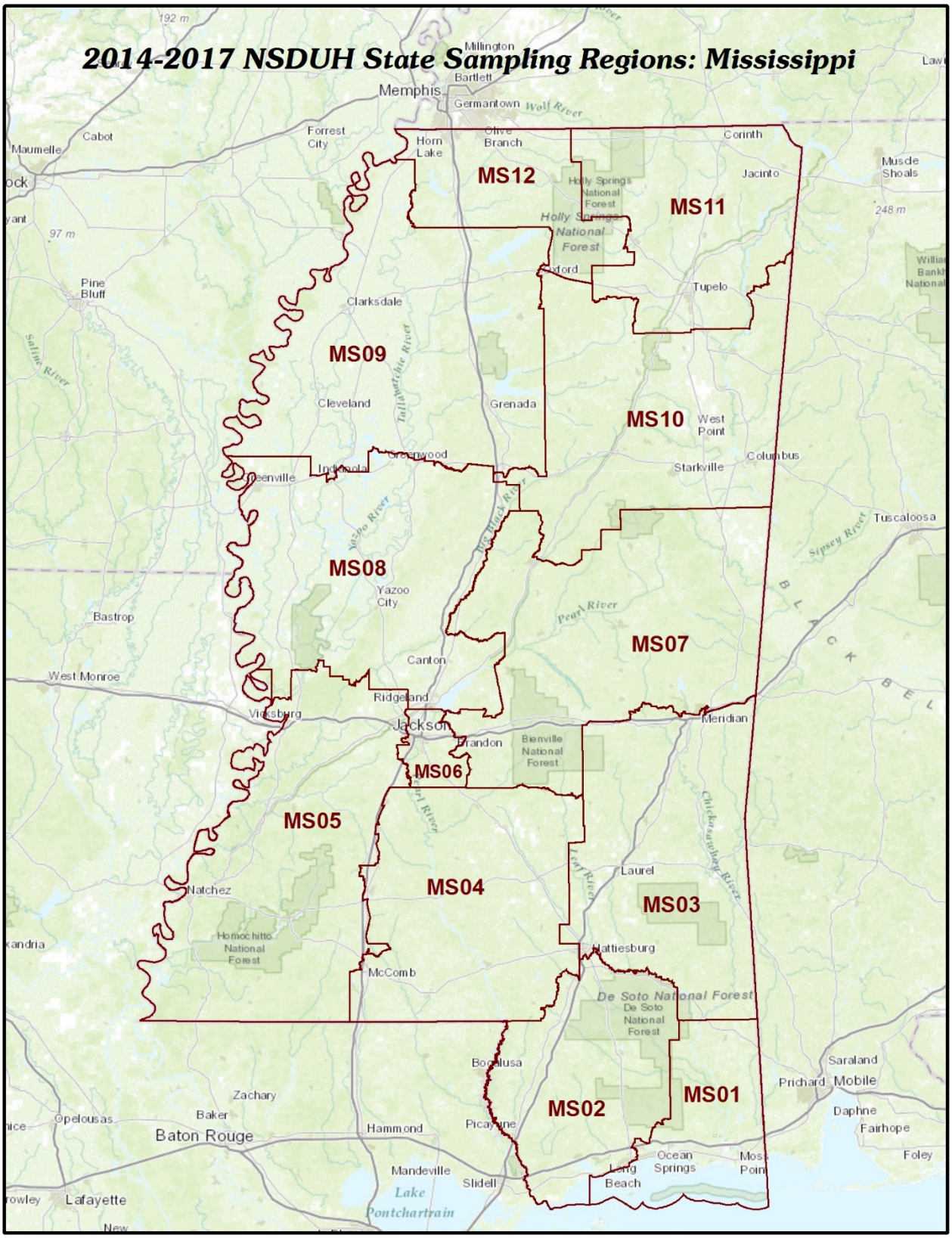

2014-2017 NSDUH State Sampling Regions: Maine
ME12
ME11
ME09
ME10
ME07
ME08
ME03
ME06
ME05
ME02
ME04
ME01
Québec
Lévis
Edmundston
NEW BRUNSWICK
Fredericton
Saint-Georges
Thetford Mines
Baxter State Park
Sherbrooke
Bangor
Waterville
Augusta
Berlin
White Mountains
Freeport
Falmouth
Portland
Scarborough
Laconia
Rochester
Kennebunk
North Berwick
Concord
Hooksett
Manchester
Londonderry
Amesbury
York
Rivière-du-Loup

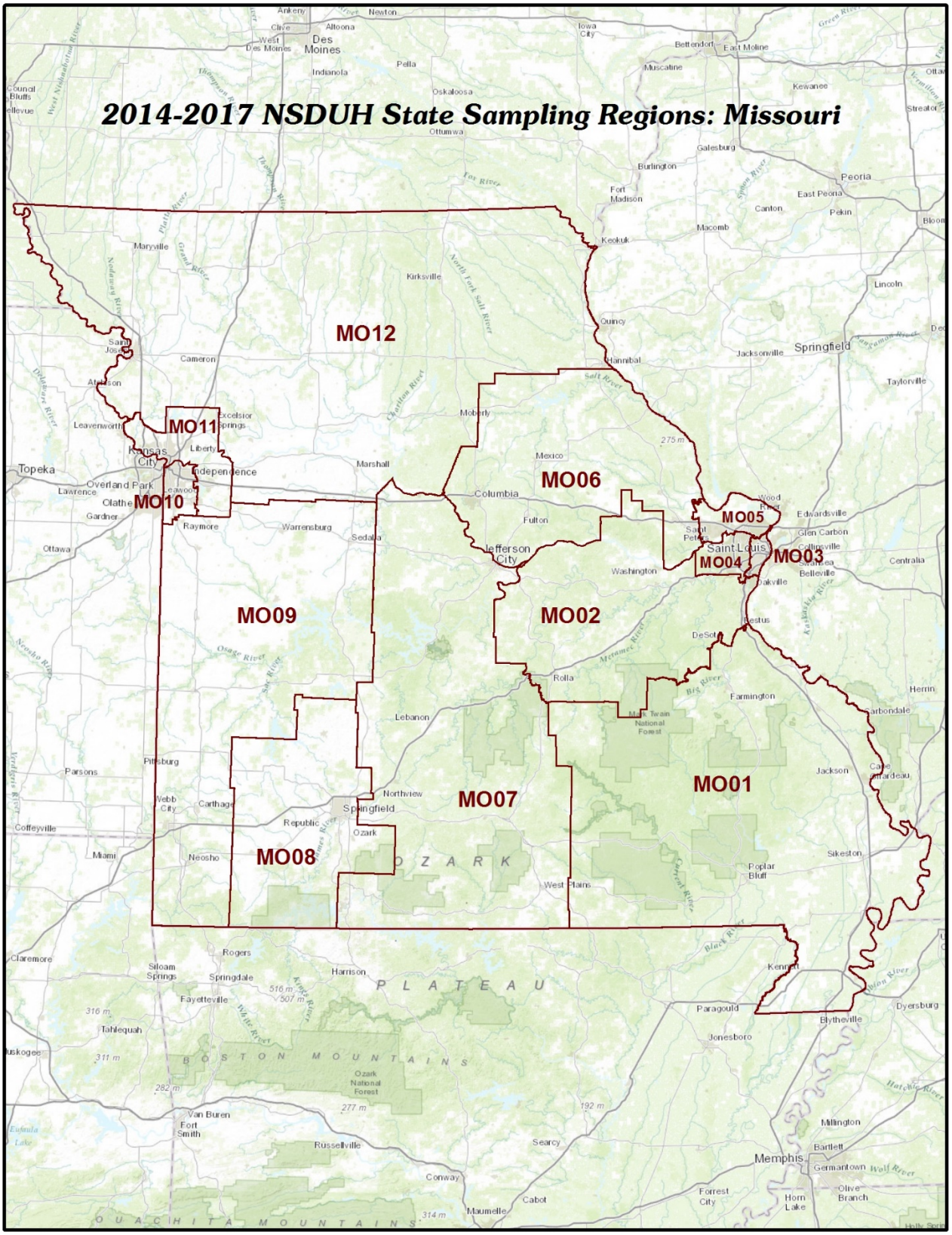

2014-2017 NSDUH State Sampling Regions: Maryland
MD01
MD02
MD03
MD04
MD05
MD06
MD07
MD08
MD09
MD10
MD11
MD12


2014-2017 NSDUH State Sampling Regions: Massachusetts
MA01
MA02
MA03
MA04
MA05
MA06
MA07
MA08
MA09
MA10
MA11
MA12
MASSACHUSETTS
CONNECTICUT
RHODE ISLAND
Boston
Providence
Hartford
Springfield
Worcester
Cape Cod Bay
Nantucket Sound
Long Island

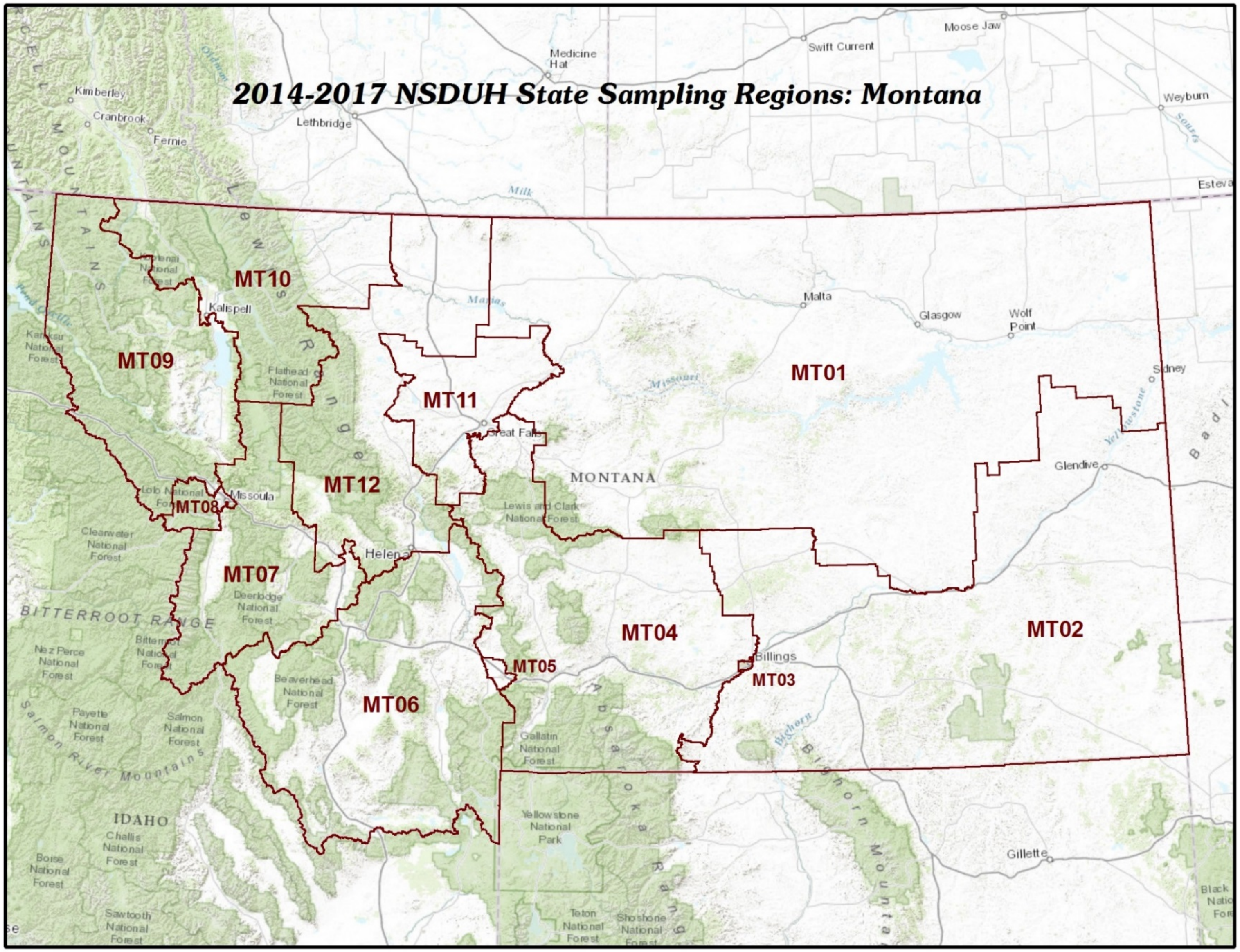

2014-2017 NSDUH State Sampling Regions: Michigan
Detroit Area
MI01
MI02
MI03
MI04
MI05
MI06
MI07
MI08
MI09
MI10
MI11
MI12
MI13
MI14
MI15
MI16
MI17
MI18
MI19
MI20
MI21
MI22
MI23
MI24
Lake Superior
Lake Michigan
Lake Huron
Fenton
Waterford
Farmington Hills
Sterling Heights
Royal Oak
Warren
Livonia
Detroit
Windsor
Taylor
Woodhaven
Monroe
Marquette
Escanaba
Sault Ste Marie
Traverse City
Cadillac
Grand Rapids
Lansing
Kalamazoo
Flint
Saginaw
Ann Arbor
Chicago
Toledo

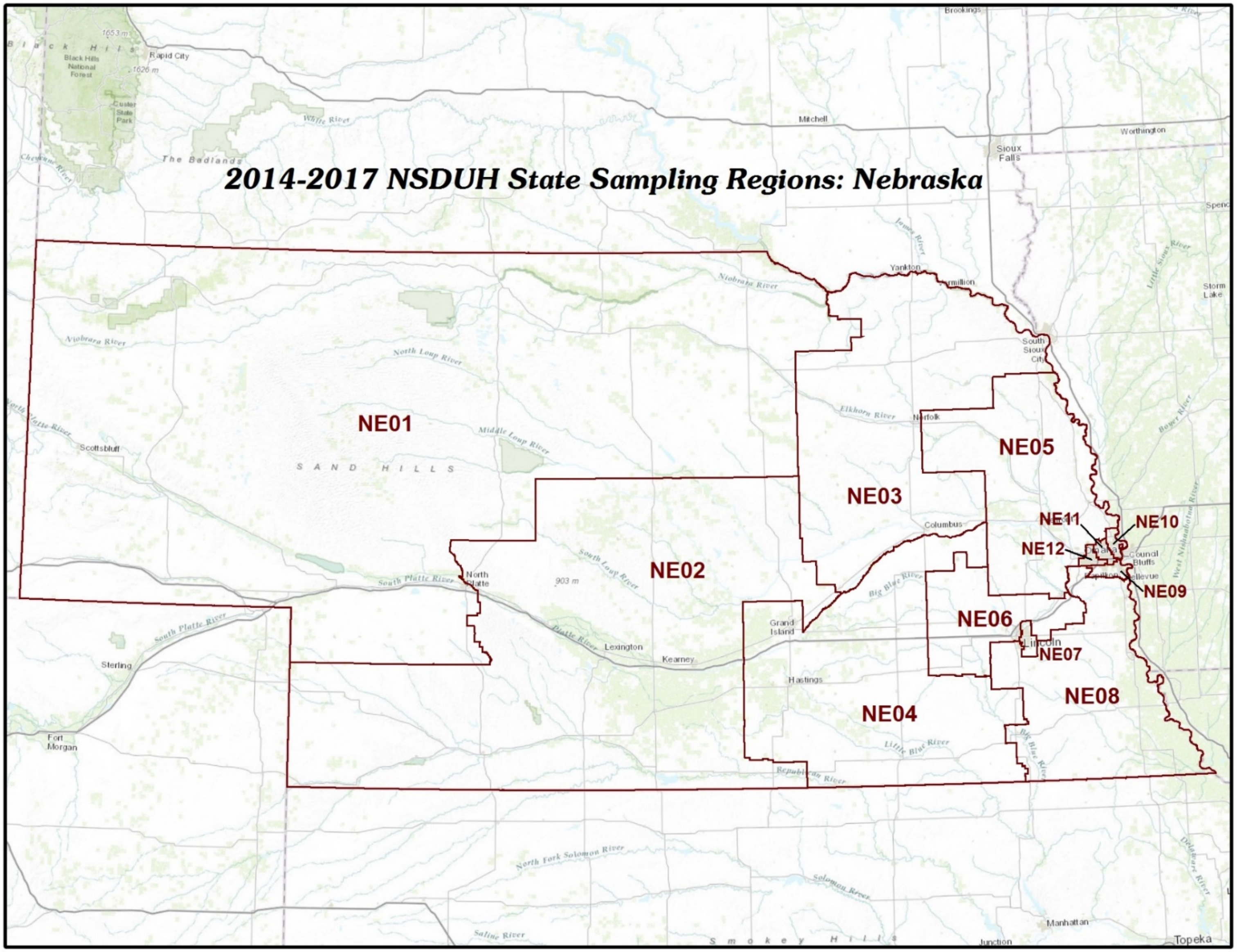

2014-2017 NSDUH State Sampling Regions: Minnesota
MN01
MN02
MN03
MN04
MN05
MN06
MN07
MN08
MN09
MN10
MN11
MN12
MINNESOTA
Winnipeg
Fargo
Moorhead
Duluth
Minneapolis
Saint Paul
Rochester
La Crosse
Sioux Falls

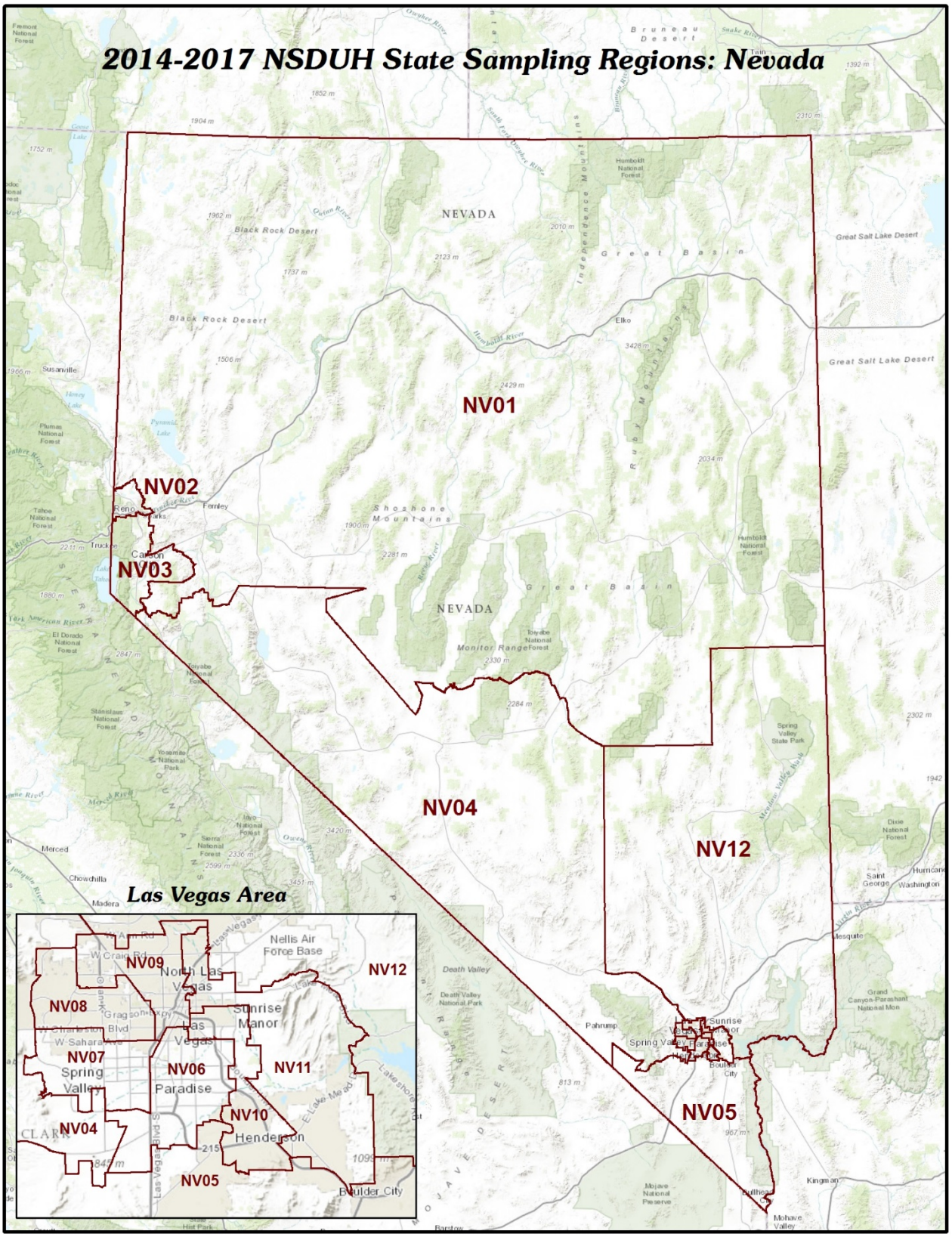

2014-2017 NSDUH State Sampling Regions: Mississippi
MS12
MS11
MS09
MS10
MS08
MS07
MS06
MS05
MS04
MS03
MS02
MS01
Memphis
Bartlett
Germantown
Millington
Wolf River
Cabot
Maumelle
Forrest City
Horn Lake
Olive Branch
Corinth
Jacinto
Holly Springs National Forest
Holly Springs National Forest
Oxford
Tupelo
Muscle Shoals
Pine Bluff
Clarksdale
Cleveland
Grenada
West Point
Columbus
Starkville
Greenwood
Indianola
Greenville
Yazoo City
Tuscaloosa
Sipsey River
Bastrop
Pearl River
Canton
West Monroe
Ridgeland
Vicksburg
Jackson
Meridian
Brandon
Bienville National Forest
Natchez
Laurel
Leaf River
Chickasawhay River
Homochitto National Forest
Hattiesburg
McComb
De Soto National Forest
De Soto National Forest
Bogalusa
Saraland
Prichard
Mobile
Zachary
Baker
Baton Rouge
Opelousas
Hammond
Picayune
Daphne
Fairhope
Foley
Ocean Springs
Long Beach
Mandeville
Slidell
Lake Pontchartrain
Lafayette
Tallahatchie River
Yazoo River
Big Black River
Satilla River
192 m
97 m
248 m


2014-2017 NSDUH State Sampling Regions: Missouri
MO12
MO11
MO10
MO06
MO05
MO04
MO03
MO09
MO02
MO07
MO01
MO08
Des Moines
Kansas City
Saint Louis
Springfield
Jefferson City
Columbia
Topeka
Memphis
OZARK
PLATEAU
BOSTON MOUNTAINS
OUACHITA MOUNTAINS

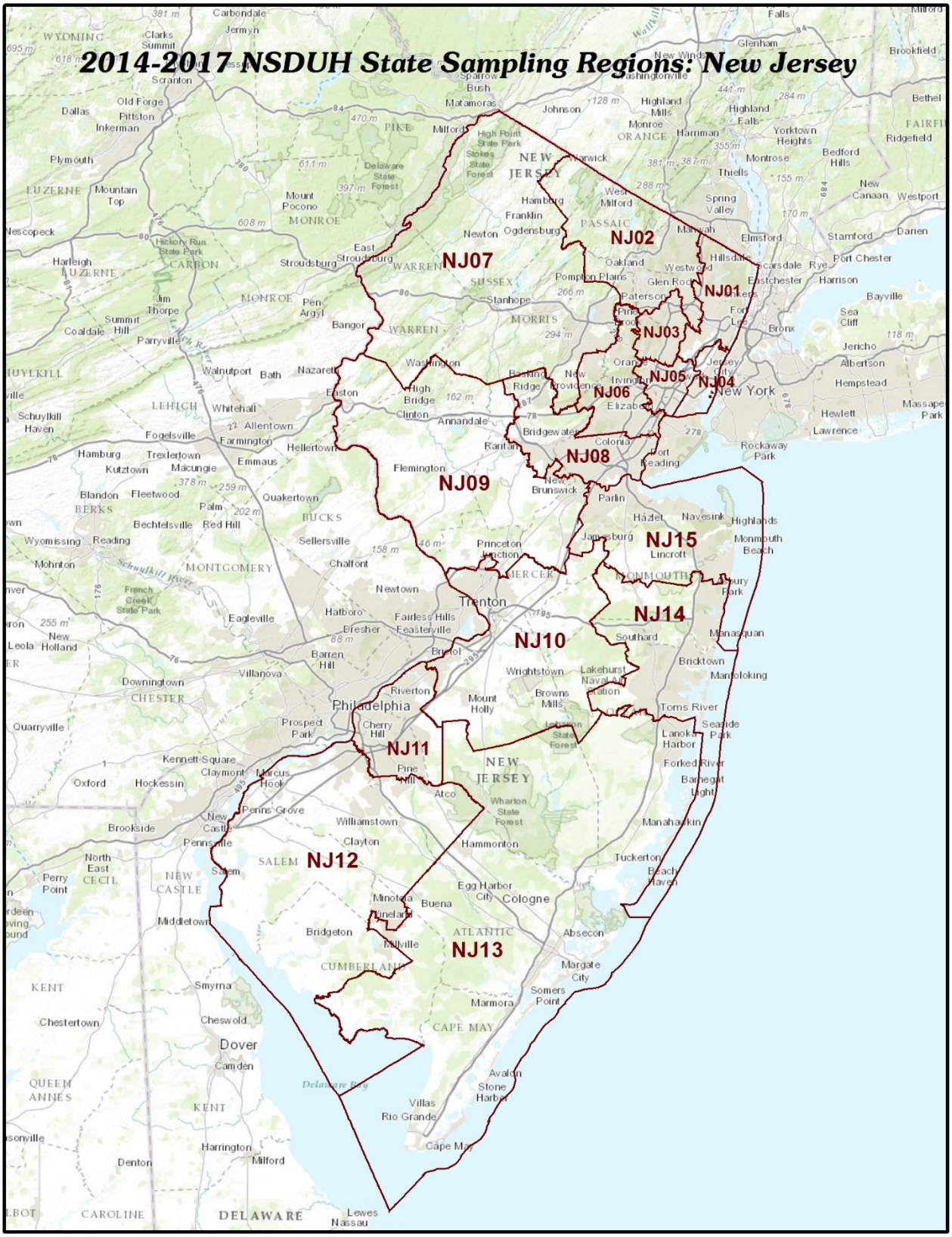
2014-2017 NSDUH State Sampling Regions: Montana

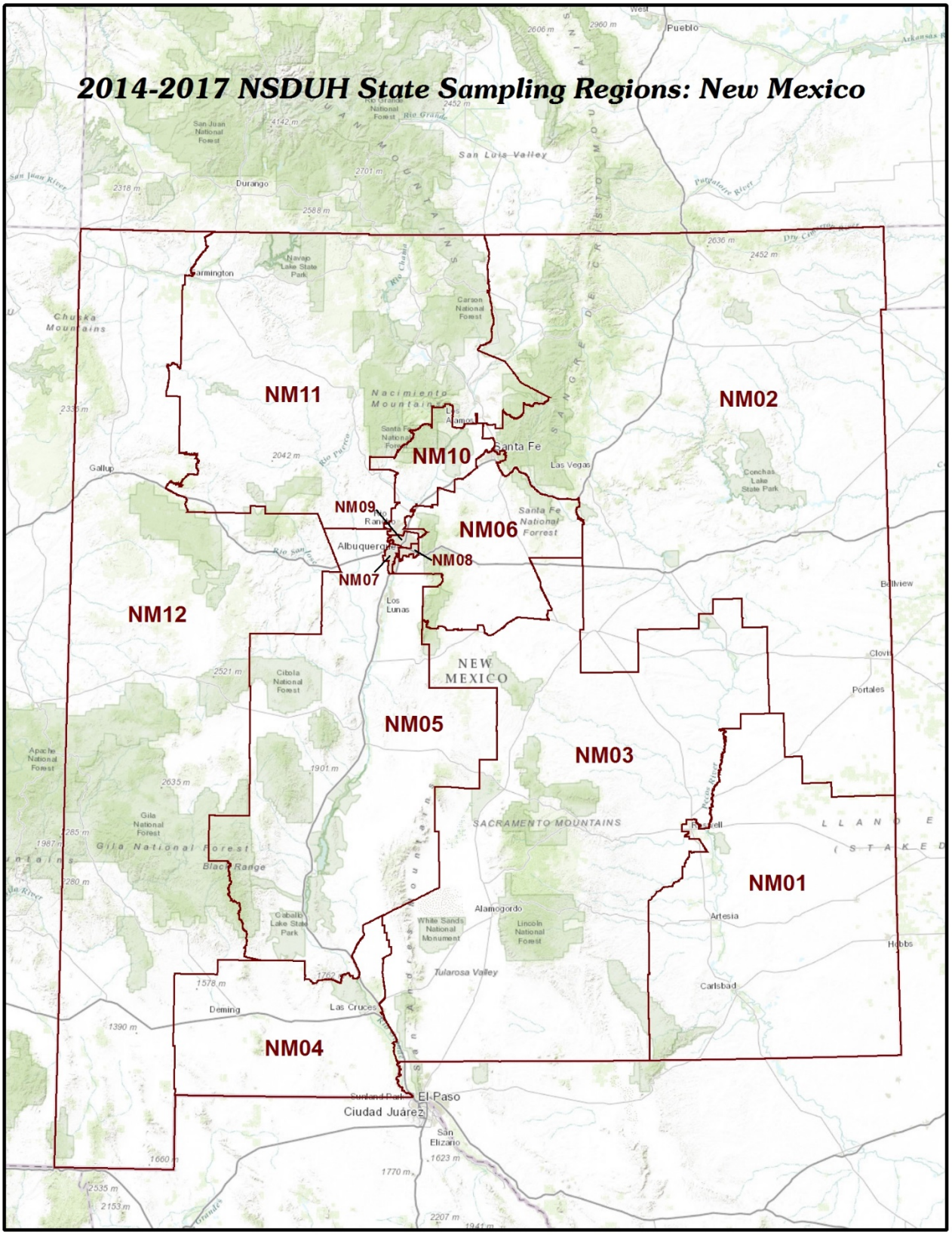


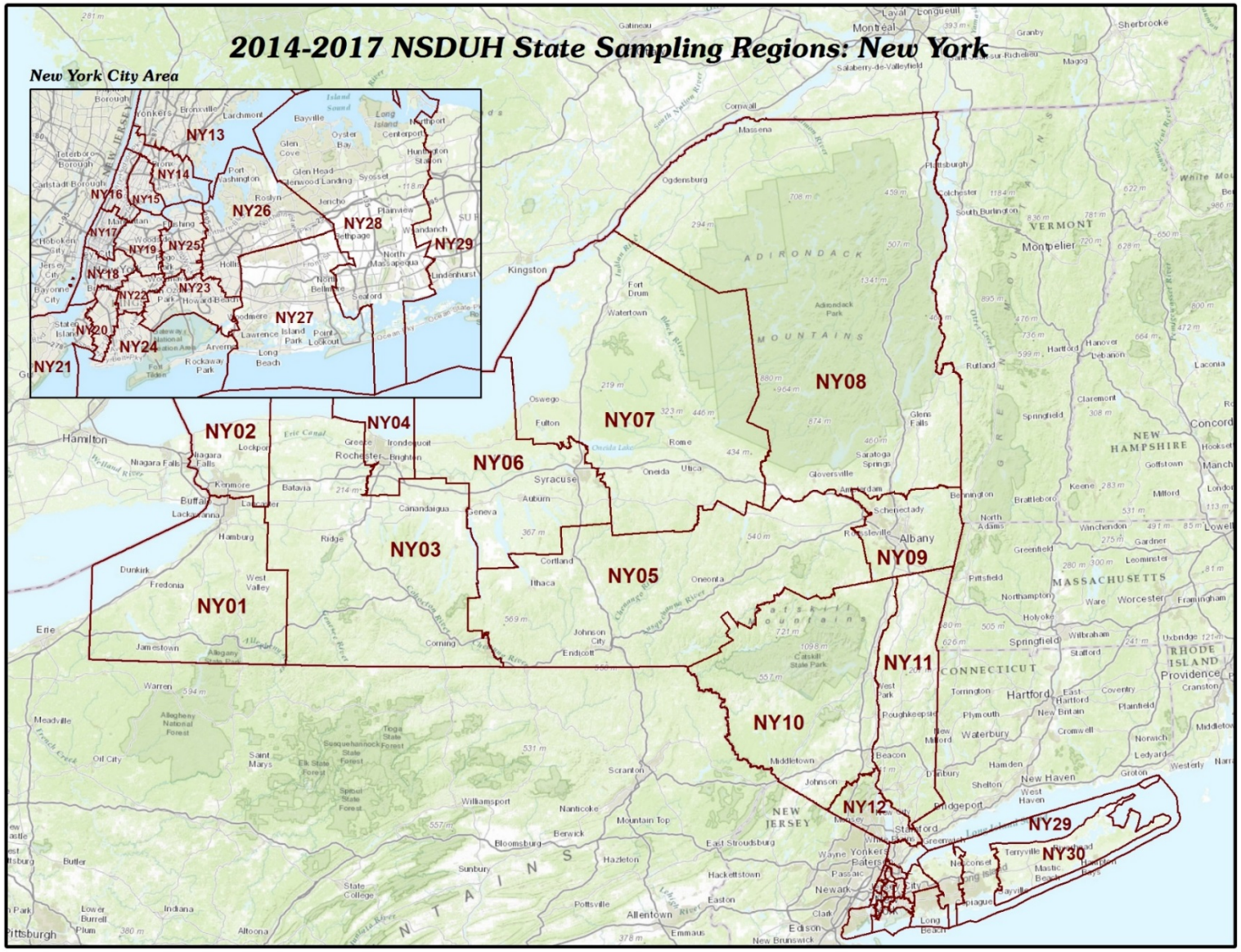

2014-2017 NSDUH State Sampling Regions: Nebraska
NE01
NE02
NE03
NE04
NE05
NE06
NE07
NE08
NE09
NE10
NE11
NE12
Scottsbluff
North Platte
Lexington
Kearney
Grand Island
Hastings
Columbus
Norfolk
Lincoln
Council Bluffs
Bellevue
South Sioux City
Yankton
Sioux Falls
Rapid City
Mitchell
Sterling
Fort Morgan
Manhattan
Topeka
S A N D H I L L S

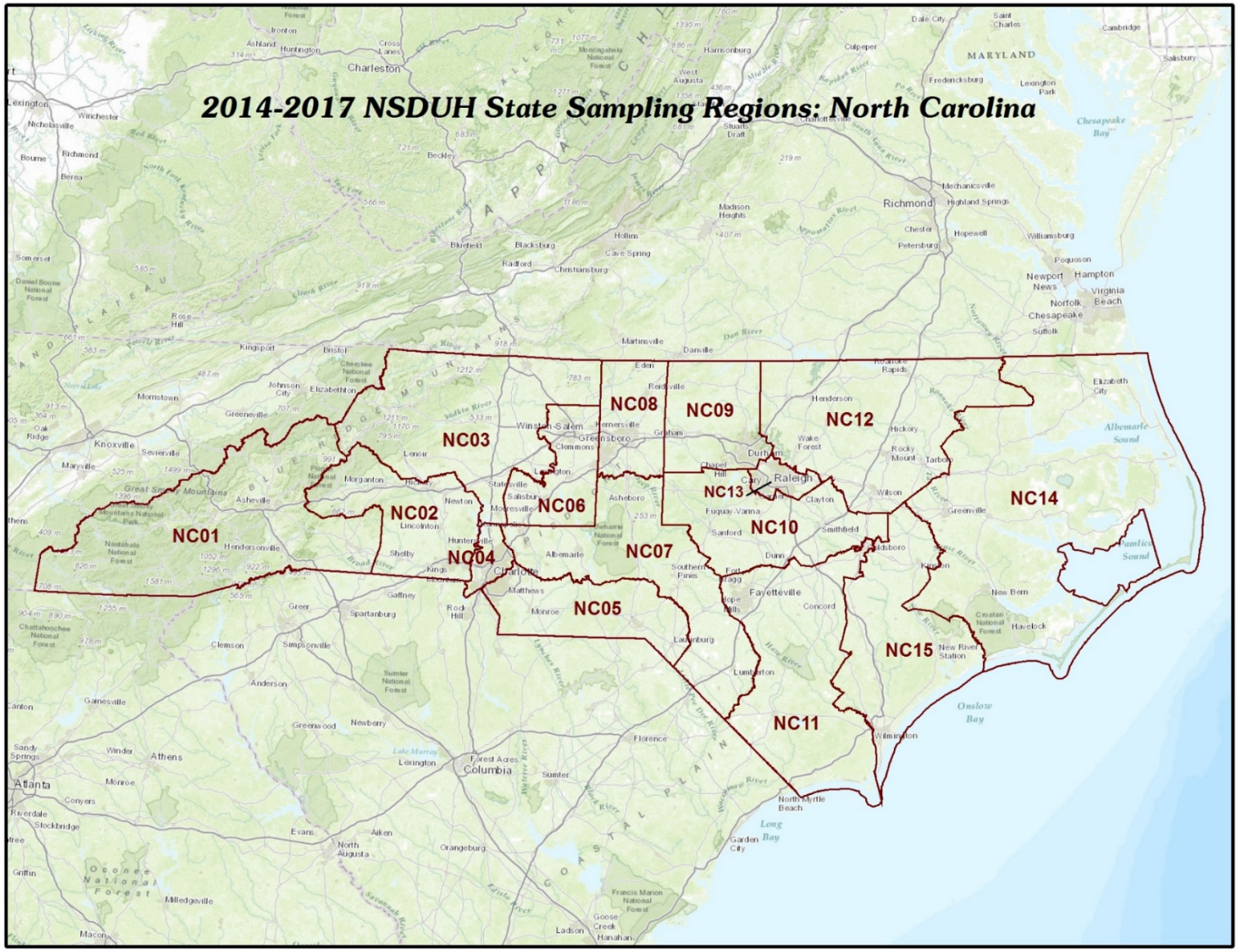

2014-2017 NSDUH State Sampling Regions: Nevada
NEVADA
NV01
NV02
NV03
NV04
NV05
NV12
Las Vegas Area
NV09
NV08
NV07
NV06
NV11
NV10
NV04
NV05
NV12
Nellis Air Force Base
North Las Vegas
Sunrise Manor
Las Vegas
Spring Valley
Paradise
Henderson
Boulder City

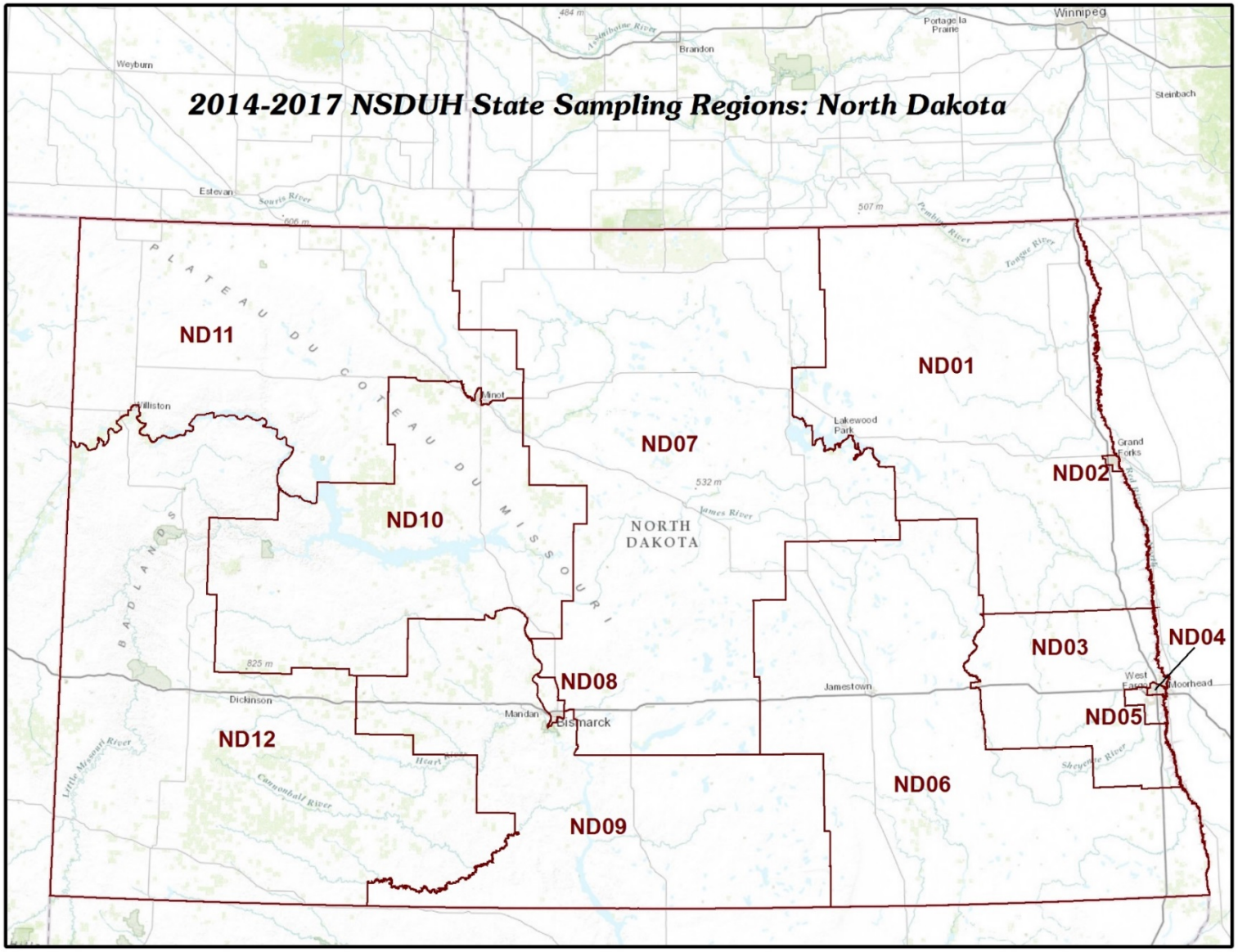

2014-2017 NSDUH State Sampling Regions: New Hampshire
NH01
NH02
NH03
NH04
NH05
NH06
NH07
NH08
NH09
NH10
NH11
NH12

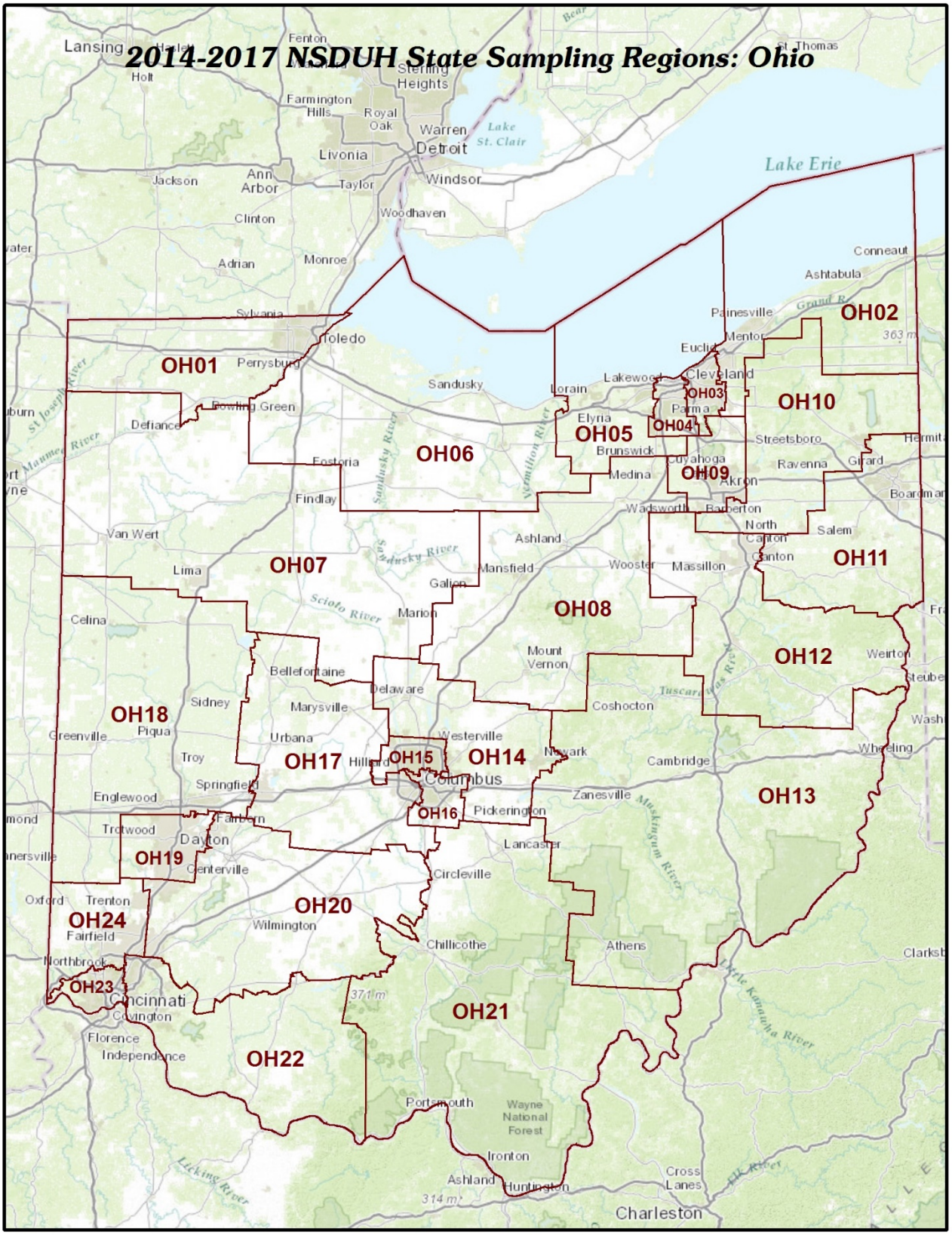

2014-2017 NSDUH State Sampling Regions: New Jersey
NJ01
NJ02
NJ03
NJ04
NJ05
NJ06
NJ07
NJ08
NJ09
NJ10
NJ11
NJ12
NJ13
NJ14
NJ15
NEW JERSEY
Philadelphia
Trenton
New York

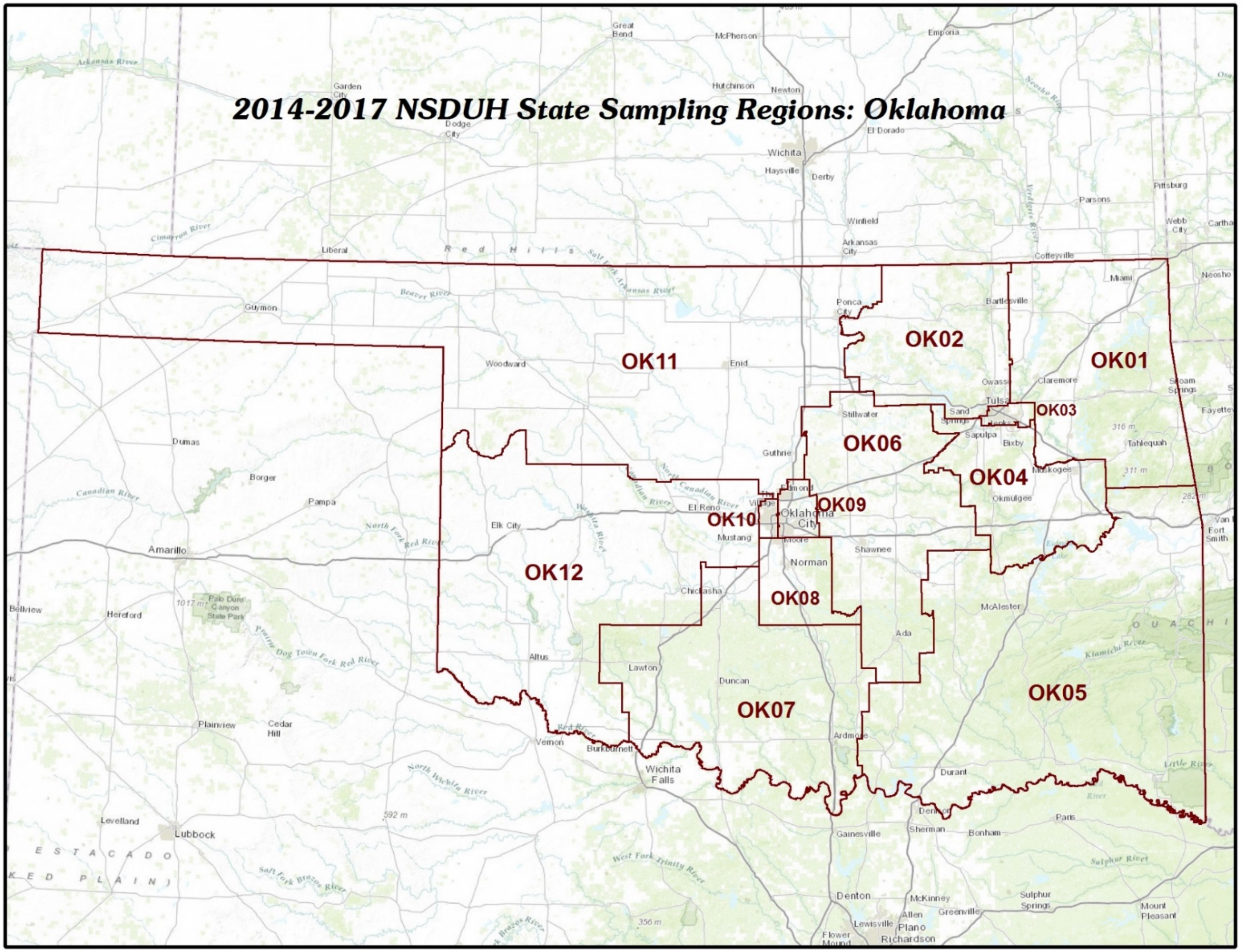

2014-2017 NSDUH State Sampling Regions: New Mexico
NM11
NM02
NM10
NM09
NM06
NM08
NM07
NM12
NM05
NM03
NM01
NM04
NEW MEXICO
Pueblo
Durango
San Luis Valley
Farmington
Gallup
Santa Fe
Las Vegas
Albuquerque
Los Lunas
Bellview
Clovis
Portales
Roswell
Artesia
Hobbs
Carlsbad
Alamogordo
Tularosa Valley
Las Cruces
Deming
El Paso
Ciudad Juárez
SACRAMENTO MOUNTAINS
Gila National Forest
Black Range
Nacimiento Mountains
Chuska Mountains

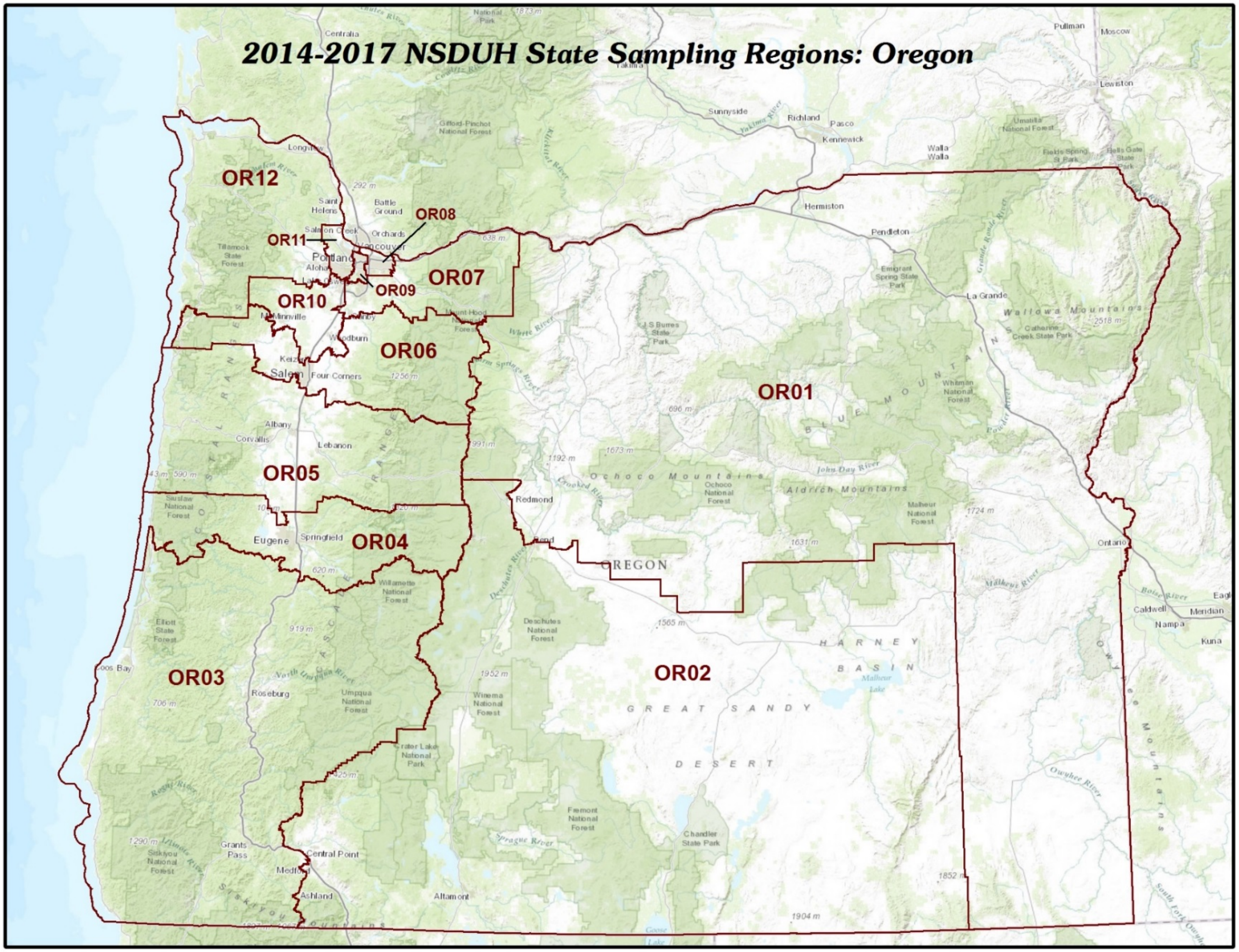

2014-2017 NSDUH State Sampling Regions: New York
New York City Area
NY01
NY02
NY03
NY04
NY05
NY06
NY07
NY08
NY09
NY10
NY11
NY12
NY13
NY14
NY15
NY16
NY17
NY18
NY19
NY20
NY21
NY22
NY23
NY24
NY25
NY26
NY27
NY28
NY29
NY30

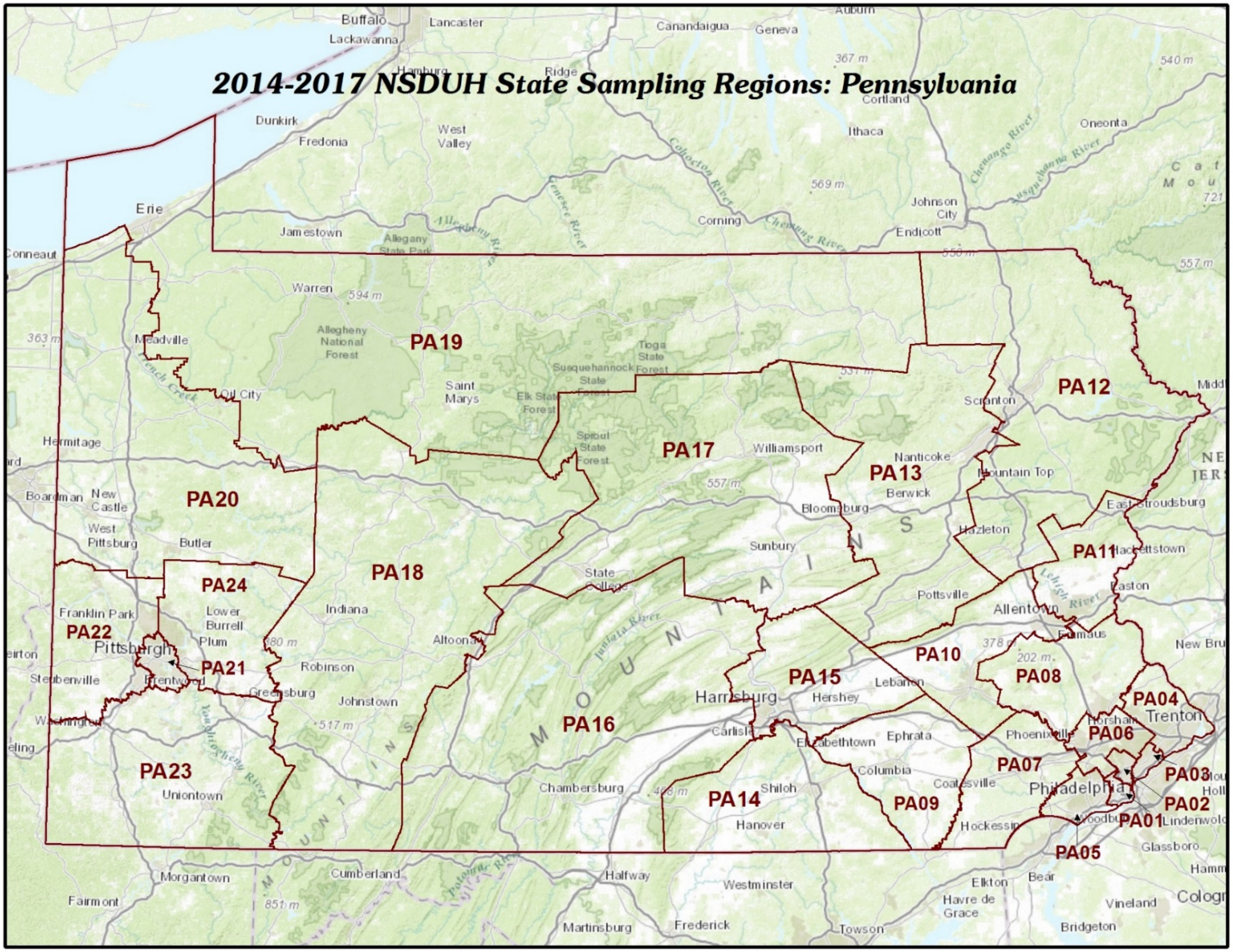

2014-2017 NSDUH State Sampling Regions: North Carolina
NC01
NC02
NC03
NC04
NC05
NC06
NC07
NC08
NC09
NC10
NC11
NC12
NC13
NC14
NC15

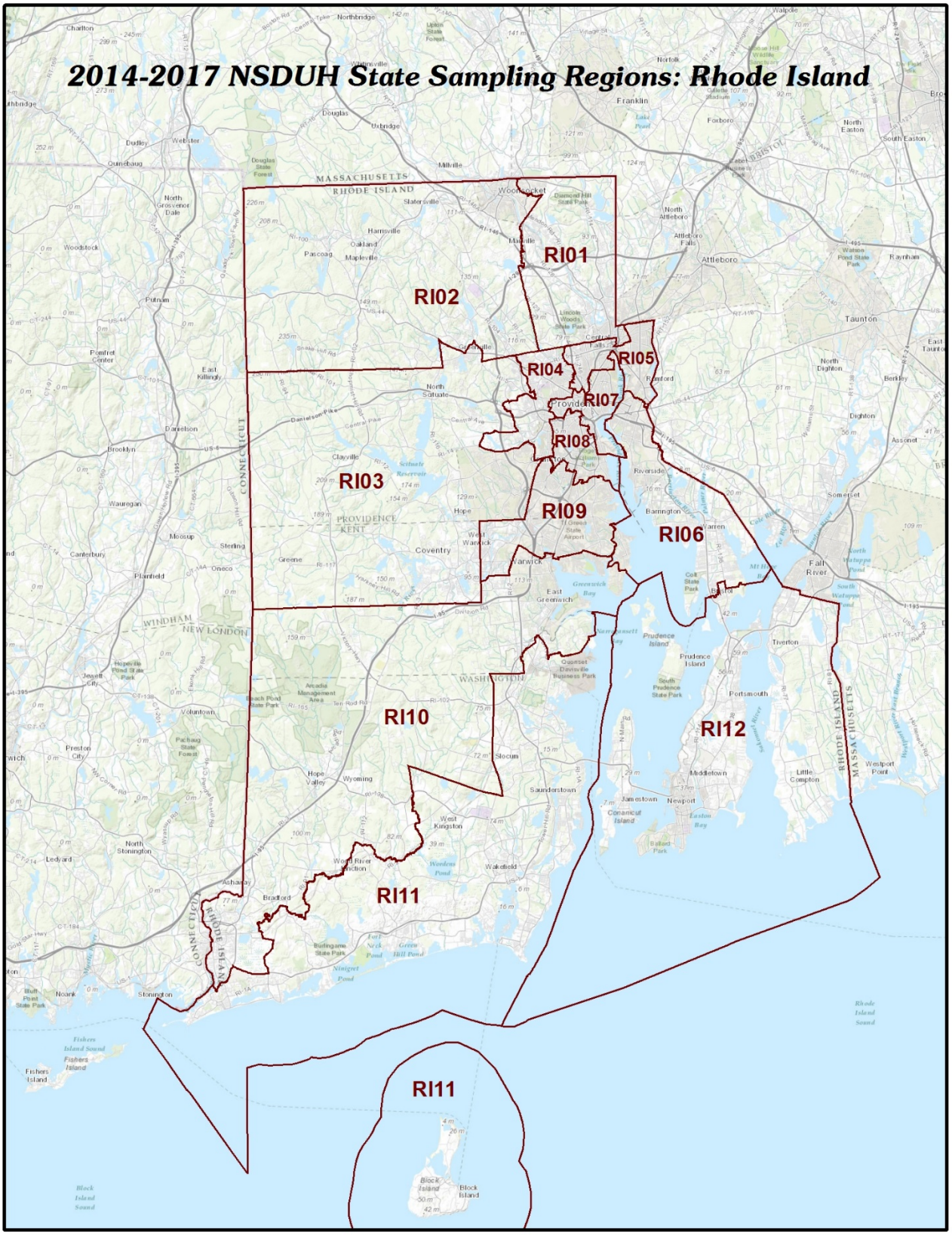

2014-2017 NSDUH State Sampling Regions: North Dakota
ND11
ND01
ND07
ND02
ND10
ND03
ND04
ND08
ND05
ND12
ND06
ND09
NORTH DAKOTA
PLATEAU DU COTEAU DU MISSOURI
BADLANDS
Winnipeg
Portage la Prairie
Brandon
Weyburn
Steinbach
Estevan
Souris River
Assiniboine River
Pembina River
Tongue River
Williston
Minot
Lakewood Park
Grand Forks
James River
Red River
Jamestown
Dickinson
Mandan
Bismarck
West Fargo
Moorhead
Little Missouri River
Heart
Cannonball River
Sheyenne River
484 m
507 m
606 m
532 m
825 m

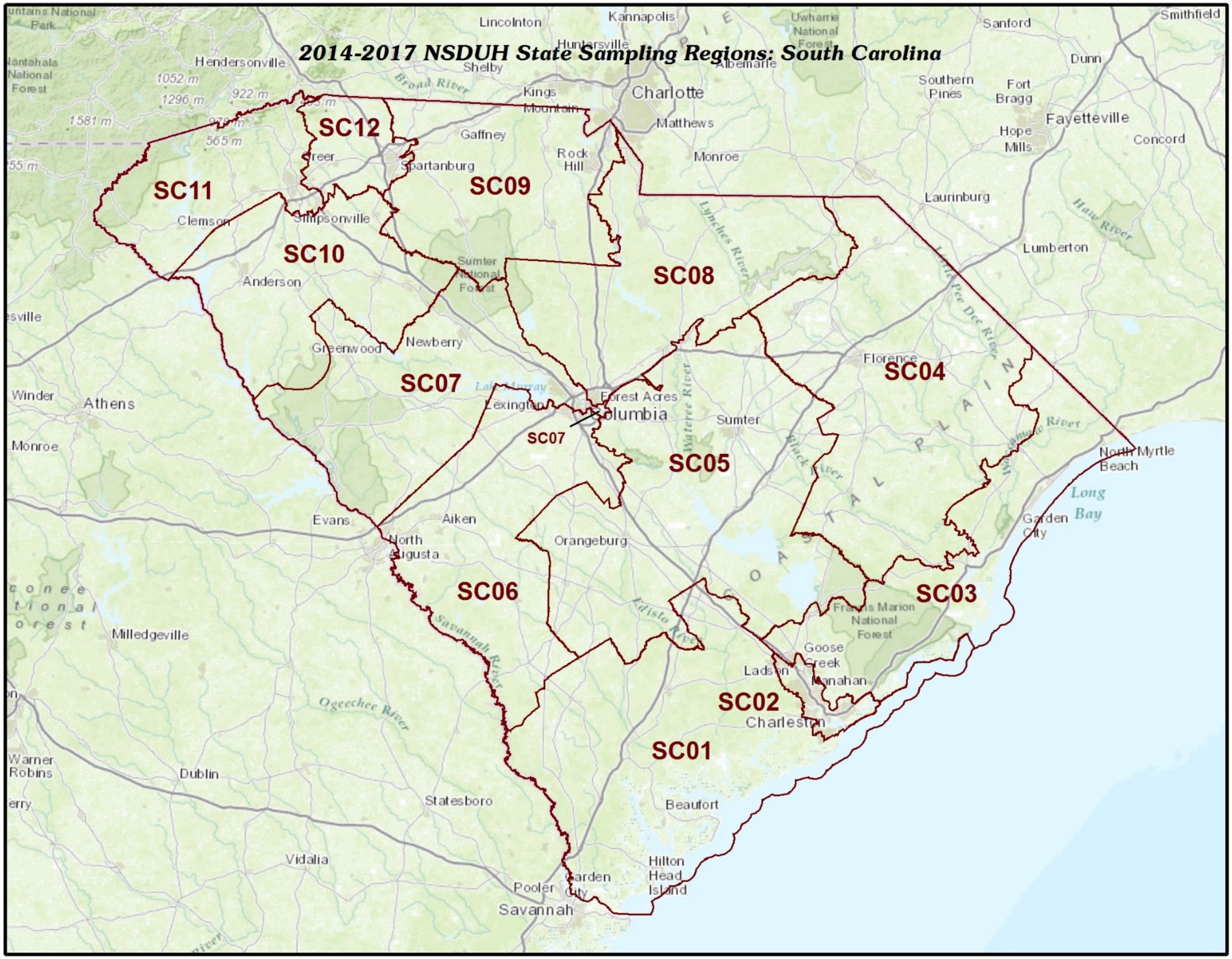

2014-2017 NSDUH State Sampling Regions: Ohio
OH01
OH02
OH03
OH04
OH05
OH06
OH07
OH08
OH09
OH10
OH11
OH12
OH13
OH14
OH15
OH16
OH17
OH18
OH19
OH20
OH21
OH22
OH23
OH24
Lake Erie
Lake St. Clair
Detroit
Toledo
Cleveland
Columbus
Cincinnati
Dayton
Akron
Canton
Lansing
Charleston

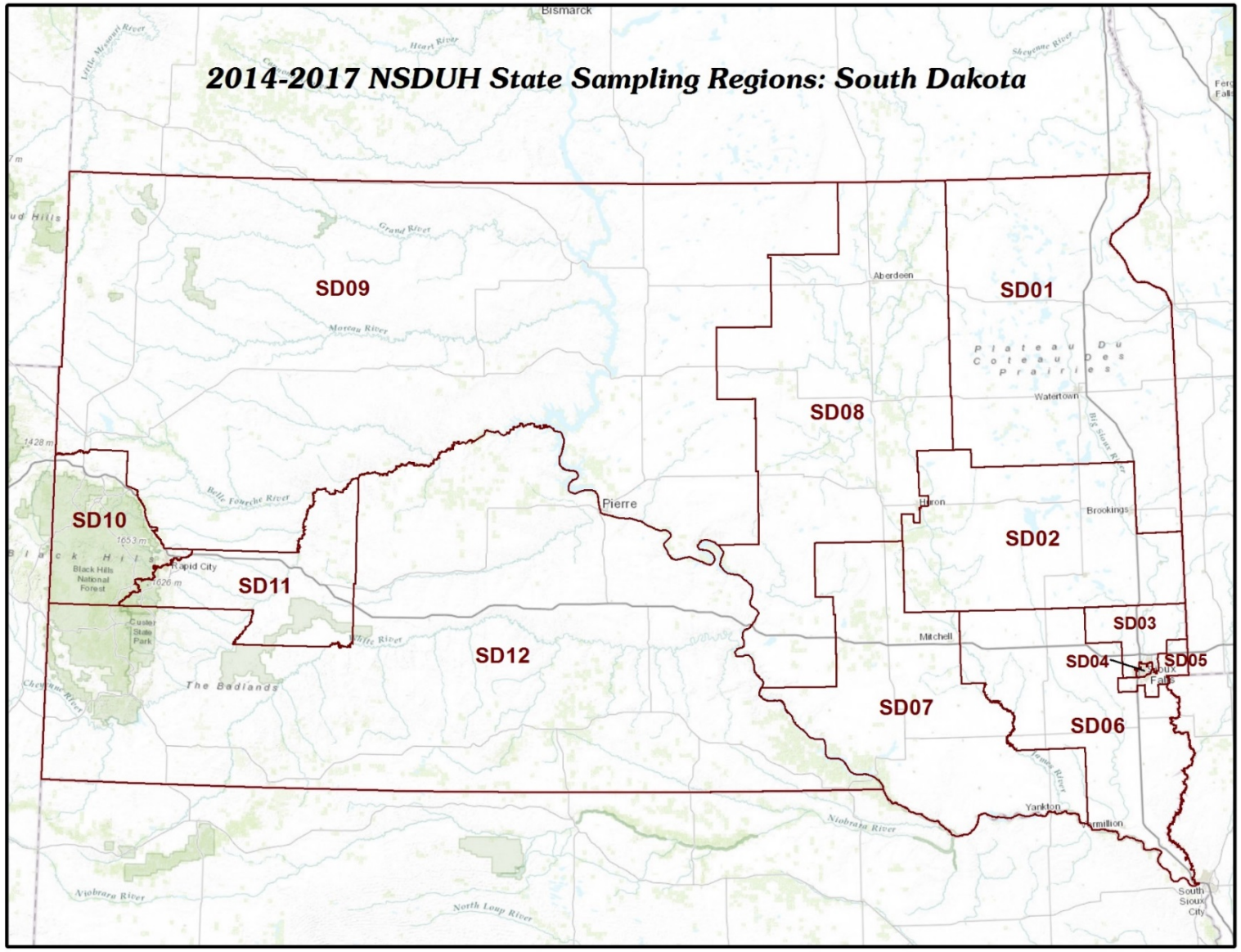

2014-2017 NSDUH State Sampling Regions: Oklahoma
OK01
OK02
OK03
OK04
OK05
OK06
OK07
OK08
OK09
OK10
OK11
OK12

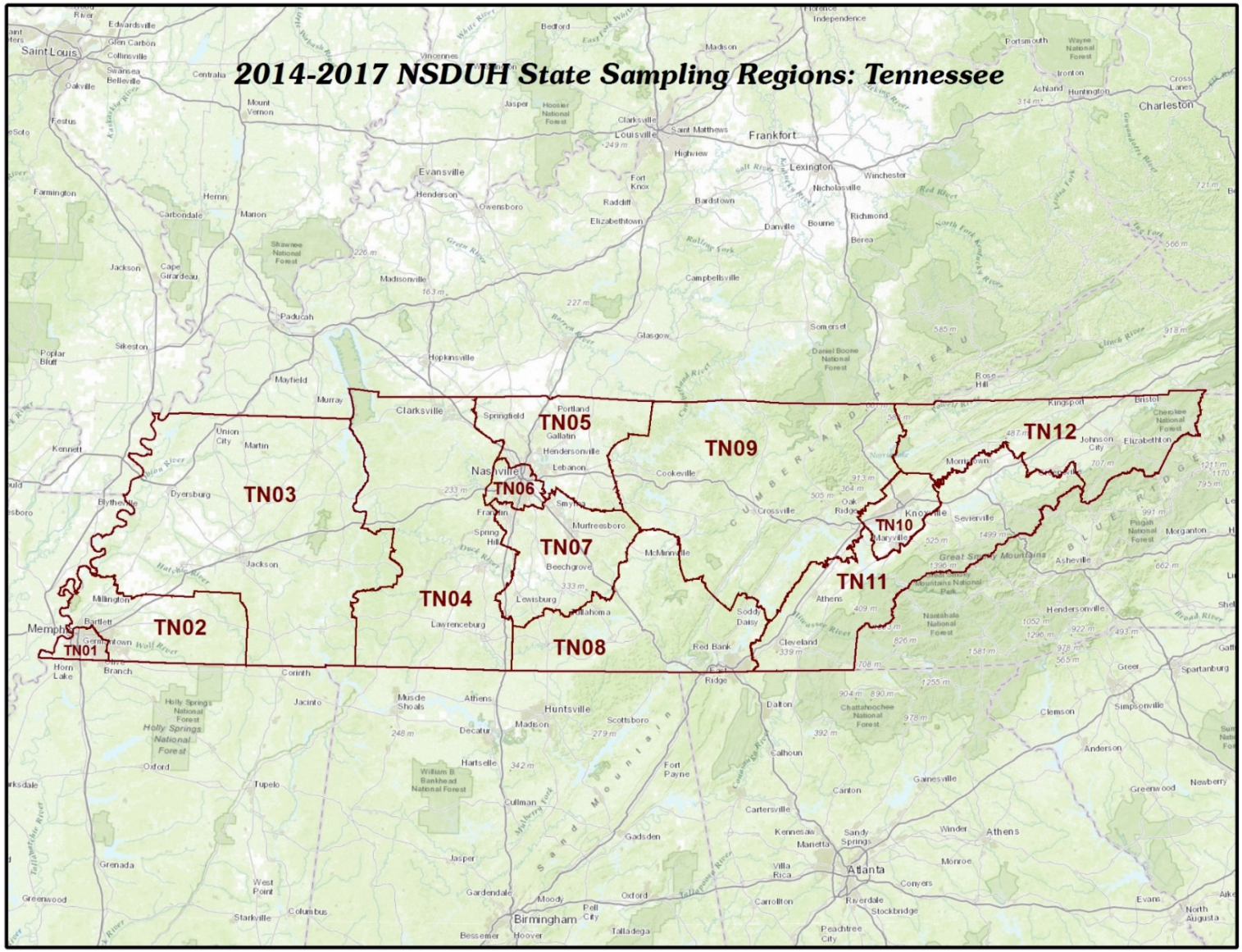

2014-2017 NSDUH State Sampling Regions: Oregon
OR12
OR11
OR08
OR09
OR07
OR10
OR06
OR01
OR05
OR04
OR03
OR02
OREGON
Portland
Salem
Eugene
Springfield
Redmond
Bend
Medford
Ashland
Roseburg
Coos Bay
Pendleton
Hermiston
La Grande
Ontario

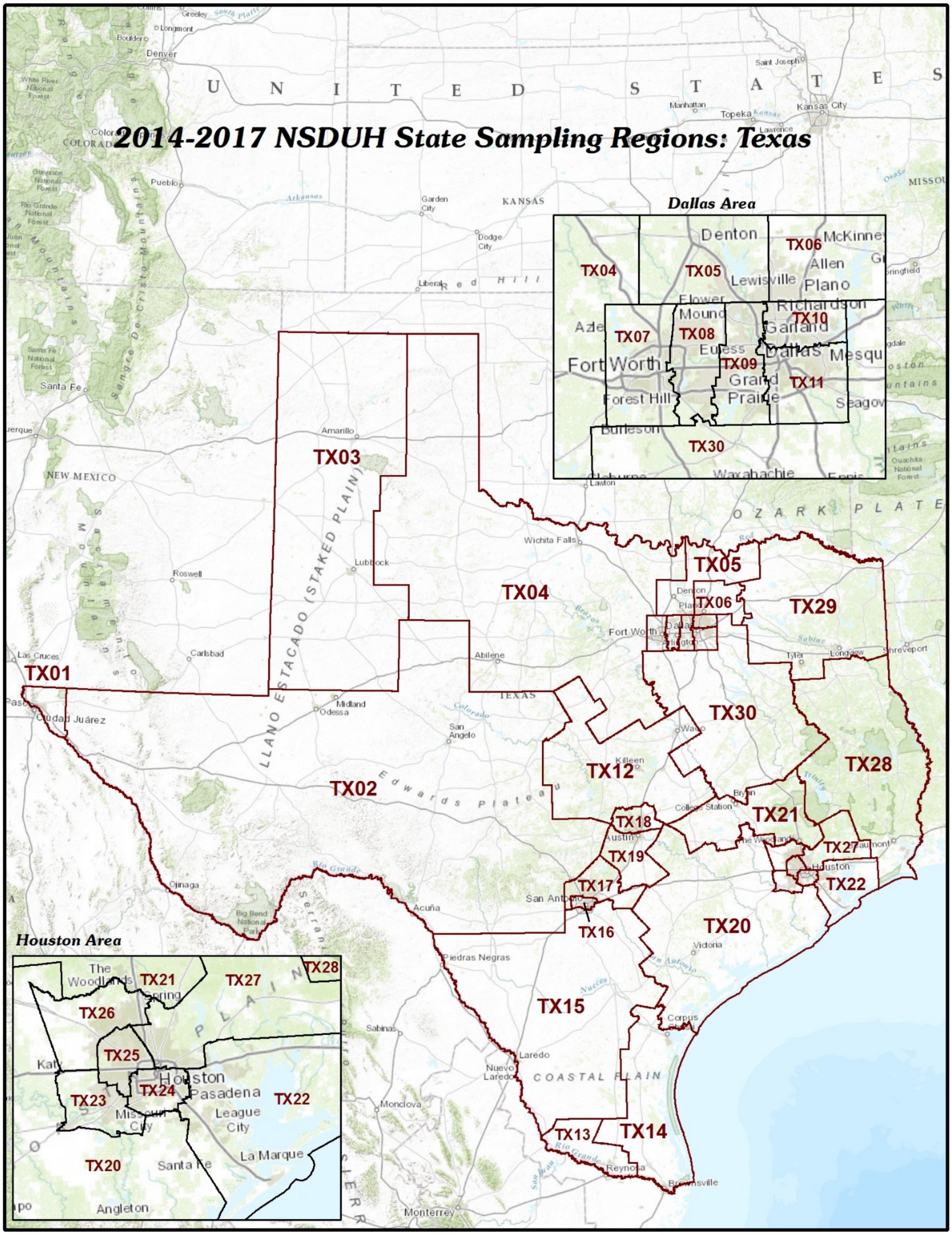

2014-2017 NSDUH State Sampling Regions: Pennsylvania
PA01
PA02
PA03
PA04
PA05
PA06
PA07
PA08
PA09
PA10
PA11
PA12
PA13
PA14
PA15
PA16
PA17
PA18
PA19
PA20
PA21
PA22
PA23
PA24
Erie
Meadville
Oil City
Warren
Jamestown
Buffalo
Dunkirk
Fredonia
Saint Marys
Williamsport
Scranton
Nanticoke
Berwick
Bloomsburg
Hazleton
Mountain Top
East Stroudsburg
Hackettstown
Easton
Allentown
Emmaus
Pottsville
Sunbury
State College
Altoona
Indiana
Butler
New Castle
Hermitage
Pittsburgh
Plum
Lower Burrell
Franklin Park
Robinson
Greensburg
Johnstown
Uniontown
Washington
Brentwood
Harrisburg
Carlisle
Hershey
Lebanon
Elizabethtown
Ephrata
Columbia
Coatesville
Phoenixville
Horsham
Trenton
Philadelphia
Chambersburg
Hanover
Shiloh
Hockessin
Morgantown
Cumberland
Fairmont
Halfway
Westminster
Frederick
Martinsburg
Towson
Elkton
Havre de Grace
Bear
Vineland
Bridgeton
Glassboro
Steubenville
Boardman
West Pittsburg
Ithaca
Corning
Johnson City
Endicott
Oneonta
Geneva
Canandaigua
Cortland
Lancaster
Lackawanna
West Valley
Conneaut

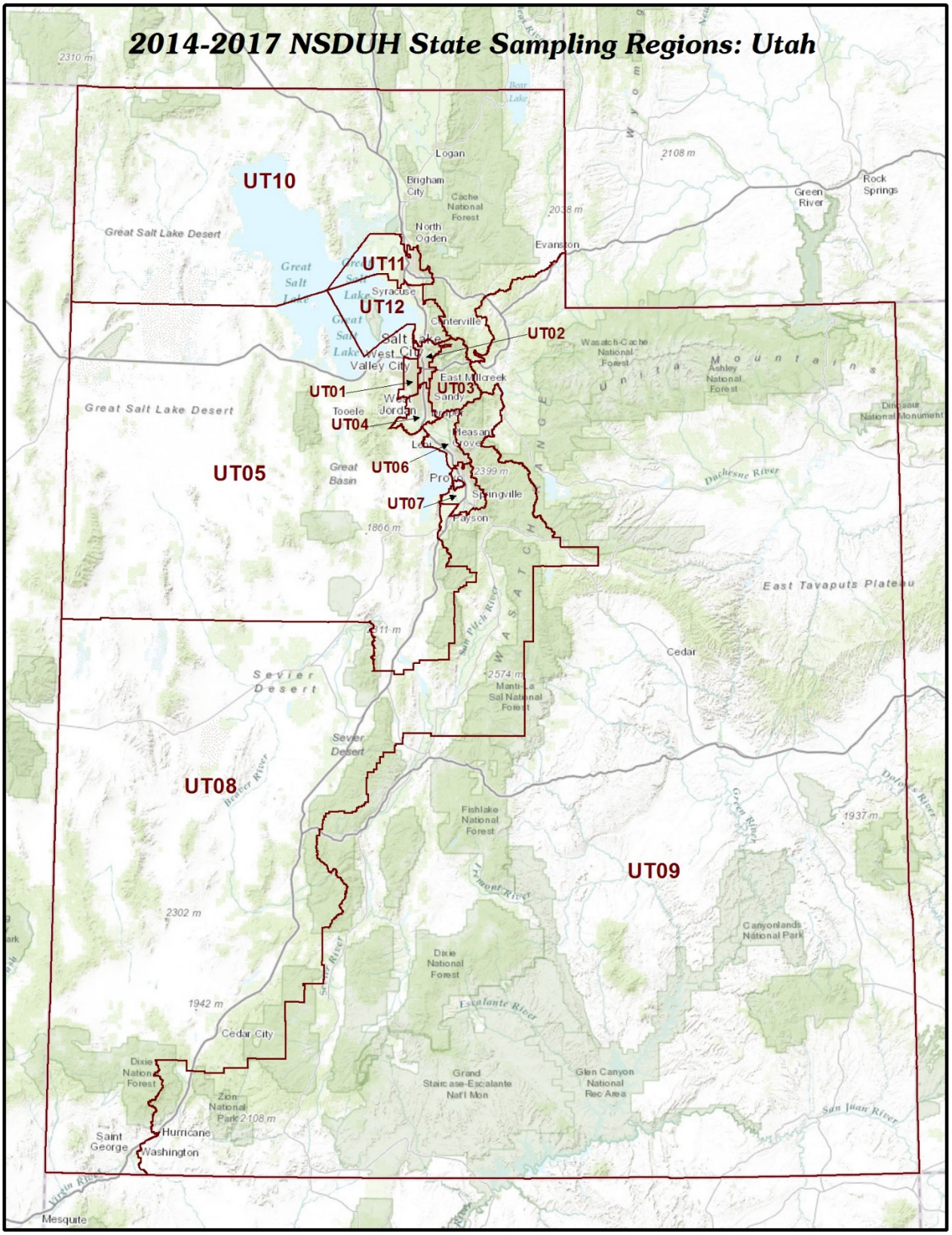

2014-2017 NSDUH State Sampling Regions: Rhode Island
RI01
RI02
RI03
RI04
RI05
RI06
RI07
RI08
RI09
RI10
RI11
RI12
RI11
MASSACHUSETTS
RHODE ISLAND
CONNECTICUT
PROVIDENCE
KENT
WASHINGTON
NEWPORT
BRISTOL
WINDHAM
NEW LONDON

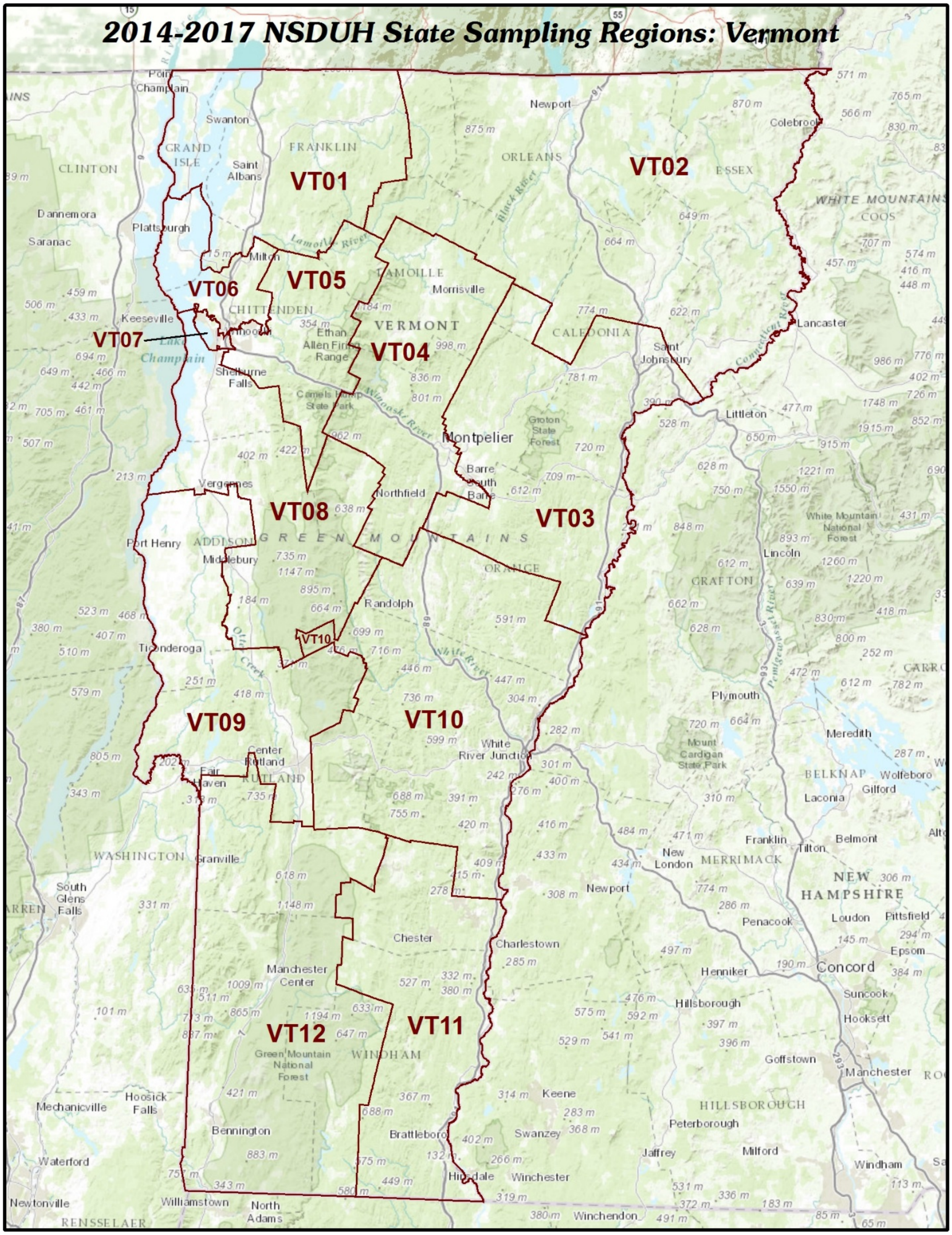

2014-2017 NSDUH State Sampling Regions: South Carolina
SC01
SC02
SC03
SC04
SC05
SC06
SC07
SC08
SC09
SC10
SC11
SC12
SC07
Charlotte
Columbia
Charleston
Spartanburg
Greenville
Anderson
Clemson
Greenwood
Newberry
Lexington
Forest Acres
Sumter
Florence
Orangeburg
Aiken
North Augusta
Beaufort
Hilton Head Island
Savannah
North Myrtle Beach
Garden City
Rock Hill
Gaffney
Greer
Simpsonville
Goose Creek
Hanahan
Ladson
Francis Marion National Forest
Sumter National Forest
Athens
Augusta
Evans
Laurinburg
Lumberton
Fayetteville
Long Bay

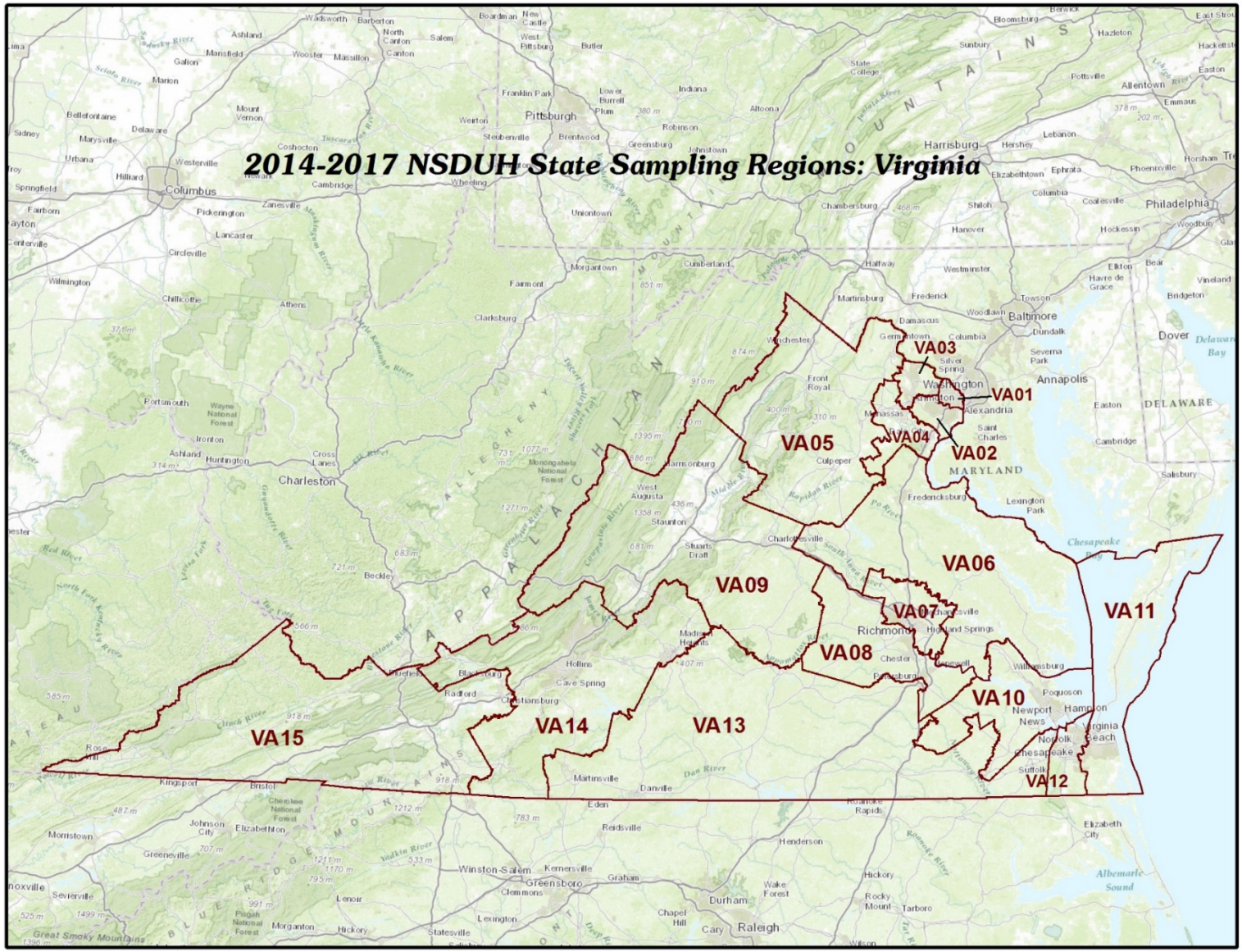

2014-2017 NSDUH State Sampling Regions: South Dakota
SD09
SD01
SD08
SD10
SD11
SD12
SD02
SD03
SD04
SD05
SD07
SD06
Bismarck
Aberdeen
Watertown
Pierre
Huron
Brookings
Mitchell
Rapid City
Yankton
Grand River
Moreau River
Belle Fourche River
White River
Niobrara River
North Loup River
Cheyenne River
Big Sioux River
Missouri River
Heart River
Sheyenne River
Plateau Du Coteau Des Prairies
Black Hills
Black Hills National Forest
Custer State Park
The Badlands
South Sioux City

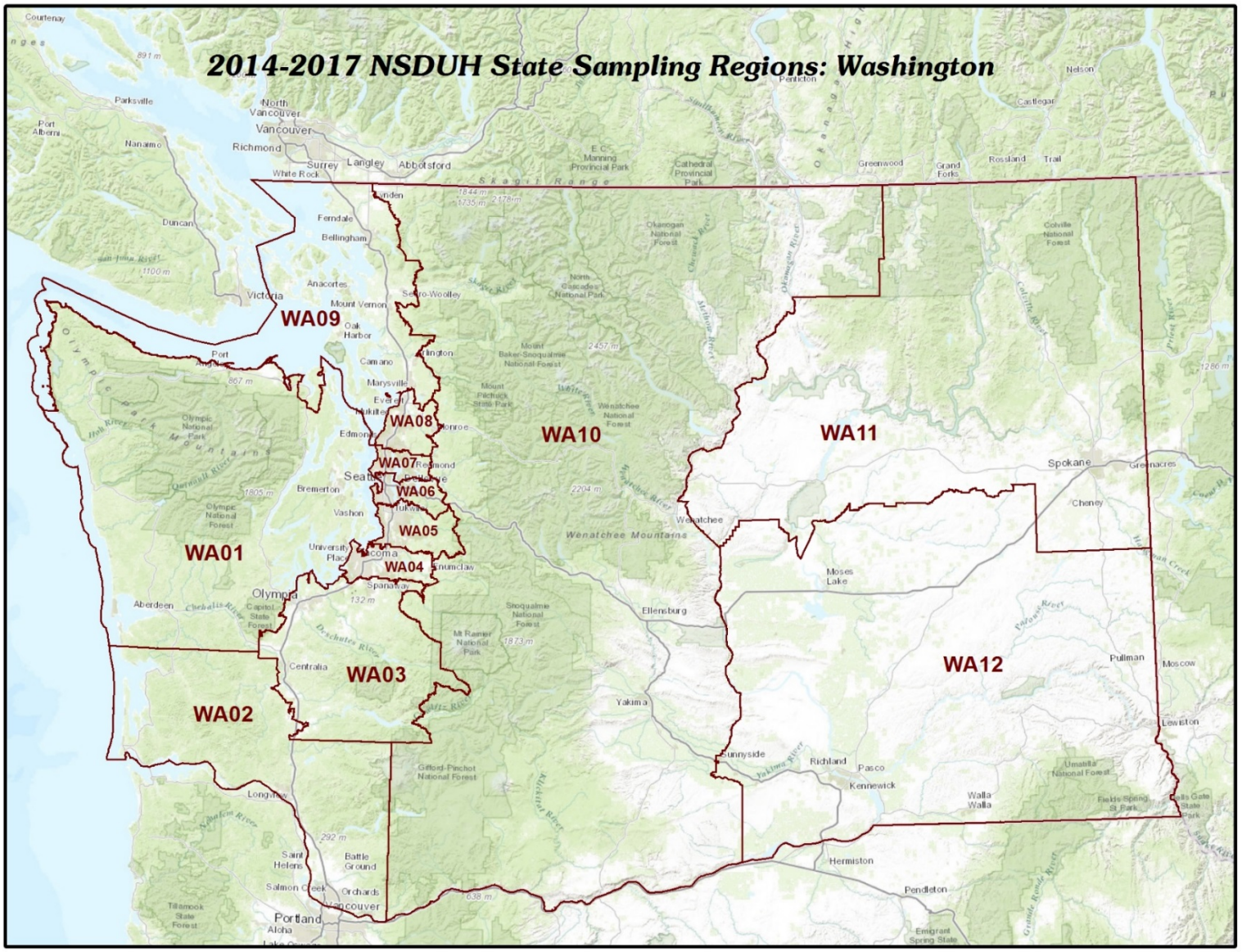

2014-2017 NSDUH State Sampling Regions: Tennessee
TN01
TN02
TN03
TN04
TN05
TN06
TN07
TN08
TN09
TN10
TN11
TN12

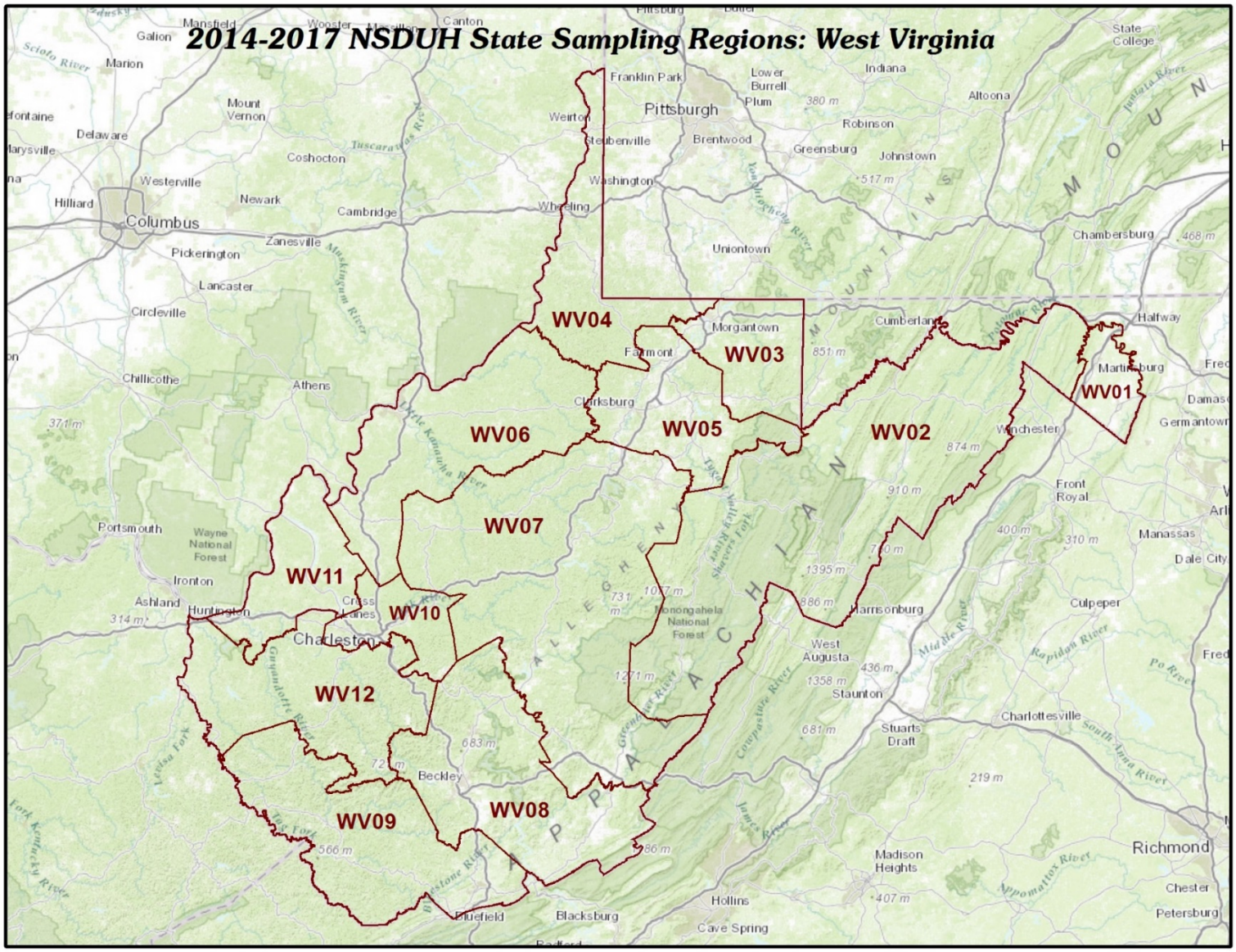

2014-2017 NSDUH State Sampling Regions: Texas
Dallas Area
Houston Area
TX01
TX02
TX03
TX04
TX05
TX06
TX07
TX08
TX09
TX10
TX11
TX12
TX13
TX14
TX15
TX16
TX17
TX18
TX19
TX20
TX21
TX22
TX23
TX24
TX25
TX26
TX27
TX28
TX29
TX30

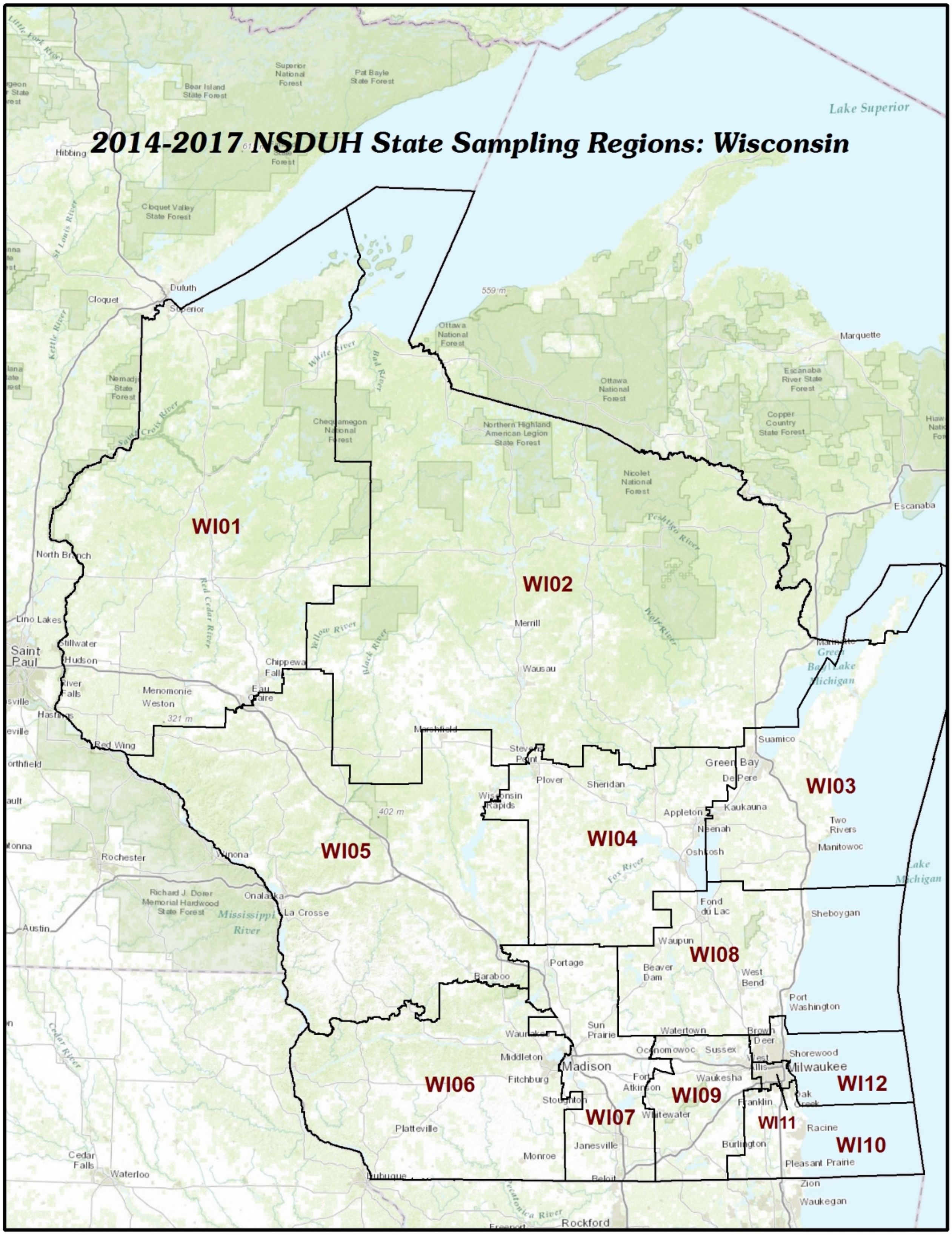

2014-2017 NSDUH State Sampling Regions: Utah
UT10
UT11
UT12
UT02
UT01
UT03
UT04
UT05
UT06
UT07
UT08
UT09
Great Salt Lake Desert
Great Salt Lake
Logan
Brigham City
Cache National Forest
North Ogden
Evanston
Green River
Rock Springs
Syracuse
Centerville
Salt Lake City
West Valley City
East Millcreek
Sandy
West Jordan
Tooele
Pleasant Grove
Provo
Springville
Payson
Great Basin
Sevier Desert
Wasatch-Cache National Forest
Ashley National Forest
Uinta Mountains
Dinosaur National Monument
Duchesne River
East Tavaputs Plateau
Cedar
Manti-La Sal National Forest
Fishlake National Forest
Green River
Canyonlands National Park
Dixie National Forest
Escalante River
Grand Staircase-Escalante Nat'l Mon
Glen Canyon National Rec Area
San Juan River
Cedar City
Zion National Park
Hurricane
Saint George
Washington
Mesquite
Virgin River
Beaver River
Sevier River
Fremont River
Dolores River
WASATCH RANGE
2310 m
2108 m
2038 m
2399 m
1866 m
2574 m
1937 m
2302 m
1942 m
2108 m

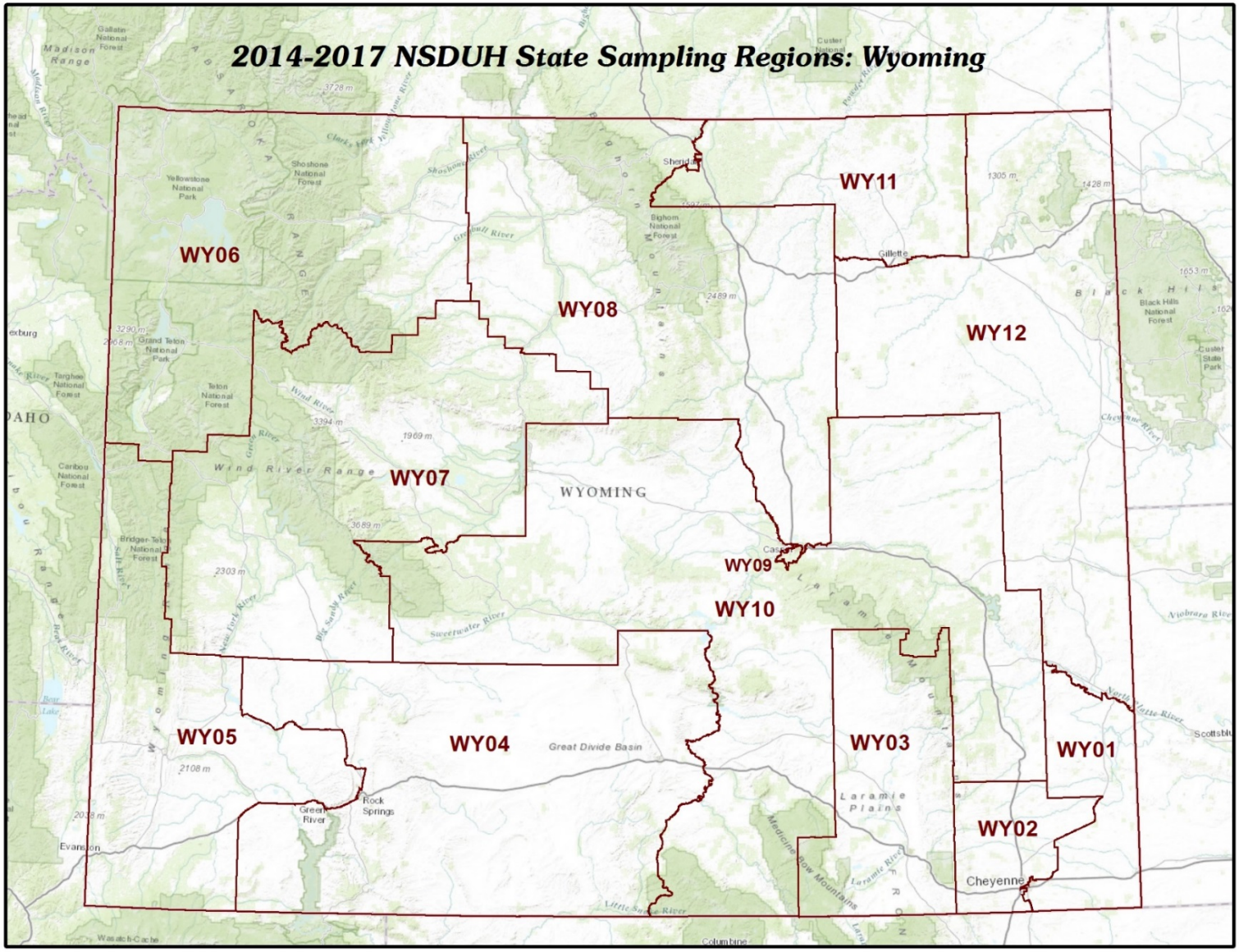

2014-2017 NSDUH State Sampling Regions: Vermont
VT01
VT02
VT03
VT04
VT05
VT06
VT07
VT08
VT09
VT10
VT11
VT12
VERMONT
NEW HAMPSHIRE
FRANKLIN
ORLEANS
ESSEX
GRAND ISLE
CHITTENDEN
LAMOILLE
CALEDONIA
WASHINGTON
ADDISON
ORANGE
RUTLAND
WINDHAM
CLINTON
COOS
GRAFTON
BELKNAP
MERRIMACK
HILLSBOROUGH
RENSSELAER
GREEN MOUNTAINS
WHITE MOUNTAINS
Montpelier
Burlington
Saint Albans
Swanton
Newport
Saint Johnsbury
Morrisville
Barre
Northfield
Randolph
Middlebury
Vergennes
Rutland
White River Junction
Chester
Manchester Center
Bennington
Brattleboro
Concord


2014-2017 NSDUH State Sampling Regions: Virginia
VA01
VA02
VA03
VA04
VA05
VA06
VA07
VA08
VA09
VA10
VA11
VA12
VA13
VA14
VA15
Pittsburgh
Columbus
Harrisburg
Philadelphia
Baltimore
Annapolis
Dover
Charleston
Richmond
Virginia Beach
Norfolk
Newport News
Raleigh
Durham
Greensboro
Winston-Salem
Roanoke
Knoxville
Great Smoky Mountains
MARYLAND
DELAWARE


2014-2017 NSDUH State Sampling Regions: Washington
WA01
WA02
WA03
WA04
WA05
WA06
WA07
WA08
WA09
WA10
WA11
WA12
Seattle
Spokane
Olympia
Yakima
Wenatchee
Moses Lake
Ellensburg
Pullman
Lewiston
Moscow
Richland
Pasco
Kennewick
Walla Walla
Hermiston
Pendleton
Sunnyside
Vancouver
Portland
Bellingham
Everett
Tacoma
Bremerton
Mount Vernon
Anacortes
Port Angeles
Aberdeen
Centralia
Longview
Cheney
Greenacres
Olympic Mountains
Wenatchee Mountains


2014-2017 NSDUH State Sampling Regions: West Virginia
WV01
WV02
WV03
WV04
WV05
WV06
WV07
WV08
WV09
WV10
WV11
WV12
Pittsburgh
Columbus
Charleston
Richmond
Morgantown
Fairmont
Clarksburg
Wheeling
Weirton
Steubenville
Washington
Uniontown
Cumberland
Martinsburg
Winchester
Harrisonburg
Staunton
Charlottesville
Beckley
Bluefield
Blacksburg
Roanoke
Lynchburg
Petersburg
Chambersburg
Altoona
Johnstown
Greensburg
Monongahela National Forest
Wayne National Forest


2014-2017 NSDUH State Sampling Regions: Wisconsin
WI01
WI02
WI03
WI04
WI05
WI06
WI07
WI08
WI09
WI10
WI11
WI12
Lake Superior
Lake Michigan
Duluth
Superior
Merrill
Wausau
Marshfield
Green Bay
De Pere
Appleton
Kaukauna
Neenah
Oshkosh
Two Rivers
Manitowoc
Sheboygan
Fond du Lac
Waupun
Beaver Dam
West Bend
Port Washington
Portage
Baraboo
Sun Prairie
Madison
Middleton
Fitchburg
Stoughton
Janesville
Beloit
Monroe
Platteville
Waukesha
Milwaukee
Shorewood
Sussex
Watertown
Whitewater
Franklin
Racine
Burlington
Pleasant Prairie
La Crosse
Onalaska
Winona
Rochester
Eau Claire
Chippewa Falls
Menomonie
Hudson
Saint Paul
Marquette
Escanaba
Rockford
Dubuque
Waukegan
Zion


2014-2017 NSDUH State Sampling Regions: Wyoming
WY01
WY02
WY03
WY04
WY05
WY06
WY07
WY08
WY09
WY10
WY11
WY12
WYOMING
Great Divide Basin
Wind River Range
Laramie Plains
Yellowstone National Park
Grand Teton National Park
Bighorn National Forest
Black Hills National Forest
Custer State Park
Rock Springs
Green River
Evanston
Cheyenne
Gillette
Sheridan
Casper
Wasatch-Cache
Columbine

Appendix B: 2016 NSDUH Pair Sampling and Selection of the Pair Sampling Parameter

The pair sampling algorithm in the National Survey on Drug Use and Health (NSDUH) is based on the Chromy and Penne (2002) adaptation of the Brewer (1963, 1975) method for selecting samples of size two as a means of selecting samples of 0, 1, or 2 persons within a selected dwelling unit (DU) containing at least one eligible person. Chromy and Penne (2002) also introduced a pair sampling parameter λ that governs the number of pairs selected.

Define the target selection probability for person *i* in DU *h* as P_{hi}. Then, to ensure that all pairs have a positive probability of selection, all person probabilities have to be strictly less than 1; and arbitrarily, the maximum P_{hi} is set to 0.99. In Brewer's (unadapted) method of sampling pairs, the sum of first-order inclusion probabilities is always equal to $n = 2$. However, because the NSDUH design calls for a selection of 0, 1, or 2 persons per DU, it is unlikely that the sum of person probabilities within a DU, $S_h = \Sigma_i P_{hi}$, equals 2. Adaptations were then applied to the sampling algorithm for the following types of DUs: DUs with S_h greater than or equal to 2 and DUs where S_h was less than 2.

B.1 Case I: DUs with $S_h \geq 2$

If $S_h \geq 2$, a multiplicative scaling factor, $F_h = 2/S_h$, was applied to all of the target selection probabilities so that they were scaled down to sum to exactly 2. Now, Brewer's method sets the pairwise selection probabilities at

$$P_{h(ij)} = \left[\frac{P_{h(i)}P_{h(j)}}{K}\right]\left[\frac{1}{1-P_{h(i)}} + \frac{1}{1-P_{h(j)}}\right]$$

by setting *K* at

$$K = 2 + \sum \frac{P_{h(i)}}{1-P_{h(i)}},$$

where

> *i* = *i*-th person in household *h* (whose selection probability depends on his or her age category: 1, 2, 3, 4, or 5) and
>
> *j* = *j*-th person in household *h* (whose selection probability depends on his or her age category: 1, 2, 3, 4, or 5).

Age category 1 corresponds to youths aged 12 to 17, 2 to adults aged 18 to 25, 3 to adults aged 26 to 34, 4 to adults aged 35 to 49, and 5 to adults aged 50 or older.

The sum of the pairwise selection probabilities taken over all unique pairs will be guaranteed to be exactly 1.

$$\sum_i \sum_{j>i} P_{h(ij)} = 1$$

It also guarantees that the sum of the pairwise selection probabilities for an individual is equal to the individual's selection probability

$$\sum_{j \neq i} P_{h(ij)} = P_{h(i)}$$

for all values of *i*.

Note that the above scheme always selects a pair of two eligible people.

B.2 Case II: DUs with $S_h < 2$

If $S_h < 2$, the problem was remedied by creating three dummy persons and distributing the remaining size measure ($2 - S_h$) to them equally (i.e., the inclusion of dummy persons in the selection could result in the selection of 0 or 1 actual persons). Operationally, this initially required the application of the following multiplicative scaling factor to the person probabilities:

$$F_h = \min\left\{\frac{2}{S_h}, \frac{0.99}{\max\left(P_{hi}\right)}\right\}.$$

However, a further modification was applied to this scaling factor that allowed some flexibility in the actual number of pairs selected. This modification was governed by the pair sampling parameter λ. Define

$$T(\lambda) = S_h + \lambda(2 - S_h); 0 \leq \lambda \leq 1.$$

Then the modified multiplicative scaling factor was expressed as

$$F_h^* = \min\left\{\frac{T\left(\lambda\right)}{S_h}, \frac{0.99}{\max\left(P_{hi}\right)}\right\}.$$

Denote S' as the sum of the selection probability after scale adjustment by F_h^*. If S' is exactly 2, then dummy people are not needed. If S' is less than 2, then three dummy people are added to the DU.

Now, for Brewer's method, set the pairwise selection probabilities as

$$P'_{h(ij)} = \left[\frac{P'_{h(i)}P'_{h(j)}}{K'}\right]\left[\frac{1}{1-P'_{h(i)}} + \frac{1}{1-P'_{h(j)}}\right]$$

by setting K' at

$$K' = 2 + \sum_i \frac{P'_{h(i)}}{1-P'_{h(i)}},$$

where

$P'_{h(i)}$ and $P'_{h(j)}$ = selection probabilities adjusted by the scaling factor F_h^*,

i = i-th person in the household (whose selection probability depends on his or her age category: 0, 1, 2, 3, 4, or 5), and

j = j-th person in the household (whose selection probability depends on his or her age category: 0, 1, 2, 3, 4, or 5).

Age category 0 corresponds to dummy people, and categories 1 to 5 are defined as in Case I.

Note that the following equation is now available: $\sum_{j \neq i} P'_{h(ij)} = P'_{h(i)}$. To maintain the original person selection probabilities despite the scale adjustment by F_h^*, Brewer's method is modified as follows. First, draw a random number, R, from a uniform (0,1) distribution. If $R \leq 1/F_h^*$, then select a pair using Brewer's method based on the adjusted pairwise selection probability. However, if $R > 1/F_h^*$, then no one is selected from the household. In this way, the probability for selecting a pair (i,j) in household h becomes $P^*_{h(ij)} = P'_{h(ij)} / F_h^*$, which, in turn, gives the original person selection probabilities, $P_{h(i)}$. Unlike Case I, where a pair of eligible persons was always selected, this adjusted selection scheme allows for 0, 1, or 2 persons to be selected from a DU.

B.3 Selection of λ

Simulation analyses resulted in the selection of $\lambda = 0.50$ for the 2002 through 2013 NSDUH sample designs. However, changes to the sample design in 2014 with respect to age group and state necessitated further simulation analyses to identify the value of λ best suited for the 2014 through 2017 design. Simulation analyses based on the 2012 screening data, modified to reflect the required 2014 through 2017 age group sample proportions (but not modified to reflect the new state proportions), were conducted, and $\lambda = 0.25$ was selected.

Table B.1 displays the expected pair selection counts (scaled to sum to 67,507) for different values of λ in the simulation exercise, and Table B.2 displays the corresponding response rates.

Table B.1 Projected Pair Selection Counts, by Age Group Pairs (Three Age Groups: 12 to 17, 18 to 25, and 26 or Older)

	Simulated 2014-2017 Age-Based Sampling Design					2012 Sampling Design	
	$\lambda =$					Simulated	Observed
Age Group Pair	0.00	0.25	0.50	0.75	1.00	$\lambda = 0.50$	(Unscaled)[1]
12+, 12+	18,054	22,752	28,630	34,047	37,809	26,664	27,035
12-17, 12-17	2,951	3,041	3,169	3,340	3,489	4,417	4,507
12-17, 18-25	2,170	2,326	2,517	2,671	2,775	3,624	3,627
12-17, 26+	5,211	6,208	7,317	7,726	7,956	5,359	5,489
18-25, 18-25	2,728	3,185	3,606	4,142	4,576	5,529	5,476
18-25, 26+	2,962	3,833	4,908	5,629	5,867	3,672	3,735
26+, 26+	2,032	4,160	7,113	10,538	13,146	4,063	4,201

[1] Observed counts in 2012 sum to 68,309, whereas the simulated counts sum to 67,507.

Table B.2 Projected Pair Response Rates, by Age Group Pairs (Three Age Groups: 12 to 17, 18 to 25, and 26 or Older)

	Simulated 2014-2017 Age-Based Sampling Design					2012 Sampling Design	
	$\lambda =$					Simulated	Observed Response
Age Group Pair	0.00	0.25	0.50	0.75	1.00	$\lambda = 0.50$	Rate[1]
12+, 12+	72.7	71.4	70.3	69.3	68.7	72.0	72.0
12-17, 12-17	81.4	81.4	81.4	81.4	81.4	81.4	81.4
12-17, 18-25	76.1	76.1	76.1	76.1	76.1	76.1	76.1
12-17, 26+	74.8	74.8	74.8	74.9	74.8	74.8	74.7
18-25, 18-25	71.2	71.2	71.2	71.2	71.2	71.2	71.2
18-25, 26+	67.1	67.1	67.1	67.1	67.1	67.1	67.1
26+, 26+	61.7	60.7	60.4	60.1	59.8	60.2	60.1

[1] Observed response rates are based on the questionnaire age.

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

The 2014 through 2017 National Survey on Drug Use and Health (NSDUH) design uses 750 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). After 2014, one half of the segments (one per quarter in each SSR) will be replaced each year with a fresh subset of a combined 4-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 3,000 (750 SSRs × 4 quarters) first-stage strata with two replicates (area segments in each stratum). This approach provides 3,000 degrees of freedom (*df*) for variance estimation for national estimates and 48 to 144 degrees of freedom for state estimates, depending on the state. 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. A similar approach for the 2014 to 2017 design would yield 750 national variance estimation strata and 12 to 36 variance estimation strata per state. Under the NSDUH design, 1 degree of freedom for variance estimation is associated with each variance estimation stratum.

For the 2005 through 2013 design, an alternate stratum-collapsing strategy was defined that had the combined effect of maintaining adequate degrees of freedom for national estimates while obtaining higher degrees of freedom for state-level estimates. This stratum-collapsing strategy will be used for the 2014 through 2017 NSDUHs and provides 750 degrees of freedom for national estimates and 48 to 144 degrees of freedom for state-level estimates. 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

$$T_Y = \sum_{h=1}^{3,000} \sum_{i=1}^{N_h} \sum_{j=1}^{N_{hi}} Y_{hij},$$

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 3,000 strata and a variable number of persons, n_{hi}, per area segment. The total sample of persons is targeted at 67,507 for the 2014 through 2017 samples or an average of 11.25 responding persons per segment.

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

$$\hat{T}_Y = \sum_{h=1}^{3,000} \sum_{i=1}^{2} \sum_{j=1}^{n_{hi}} w_{hij} y_{hij},$$

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 3,000 strata can be written as

$$v(\hat{T}_Y) = \sum_{h=1}^{3,000} \sum_{i=1}^{2} 2(z_{hi} - \bar{z}_h)^2,$$

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 3,000 strata into K (< 3,000) strata, each containing H_k of the original strata, and such that $\sum_{k=1}^{K} H_k = 3,000$. 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 $z'_{ki} - \bar{z}'_k = \sum_{h \in k}^{H_k} (z_{hi} - \bar{z}_h)$.

[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.

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

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

The first term can be shown to have the same value as $v\left(\hat{T}_Y\right)$ by rearranging the summation as

$$\sum_{k=1}^{K}\sum_{i=1}^{2} 2\sum_{h\in k}^{H_k}(z_{hi} - \bar{z}_h)^2 = \sum_{k=1}^{K}\sum_{h\in k}^{H_k}\sum_{i=1}^{2} 2(z_{hi} - \bar{z}_h)^2 = \sum_{h=1}^{3{,}000}\sum_{i=1}^{2} 2(z_{hi} - \bar{z}_h)^2 .$$

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. Although the expectation of the second term is zero, it has a positive variance. As a result, there is a loss in statistical efficiency from using the collapsed method. That is to say, the relative variance of the collapsed variance estimator is higher than that of the uncollapsed variance estimator. This is reflected in its smaller degrees of freedom.

This page intentionally left blank

Appendix D: 2016 NSDUH Census Block Assignment Procedures

D.1 Introduction

A coordinated sample design was developed for the 2014 through 2017 National Surveys on Drug Use and Health (NSDUHs). The multiyear design consists of a deeply stratified, multistage area probability design. The first stages of selection involve selecting census tracts within state sampling regions (SSRs), census block groups within selected census tracts, and area segments within selected census block groups. Segments consist of one or more adjacent census blocks and are defined using 2010 census geography. After segments are selected, specially trained field listers visit the areas and enumerate all eligible dwelling units (DUs) within each segment's boundaries. This process, called "counting and listing," creates the sample frame for the fifth stage of selection (DUs). To increase the precision of year-to-year trend estimates and to reduce the costs associated with counting and listing, the multiyear design facilitates 50 percent overlap in area segments within each successive 2-year period from 2014 through 2017. Thus, only half of the segments will be listed for the 2015 through 2017 surveys.

Because the counting and listing of sample segments for NSDUH occurs at the segment level rather than at the census block level, only the group of blocks in which a NSDUH respondent resides is known. A requirement of the 2014 through 2017 surveys is to associate each NSDUH respondent's DU to a specific census block. Using desktop computer mapping software, census block information is recorded by manually comparing field listings with electronic segment maps that were created using census Topologically Integrated Geographic Encoding and Referencing (TIGER) data. As a result, block-level data can be associated with NSDUH respondents, improving the accuracy of geographic variables used for data analyses.

This appendix describes the procedures used to assign each NSDUH respondent's DU to a census block (Section D.2). In addition, Section D.3 describes the quality control (QC) procedures that are used to ensure the accuracy of the census block assignments. The census block assignments and their corresponding QC procedures are completed on a quarterly basis.

D.2 Step-by-Step Procedures

1. **Create respondent DU file.** Approximately 2 weeks following the completion of each survey quarter, a respondent DU file is created from the preliminary person-level weight file. Because the completed case rule has not been implemented when the respondent DU file is created, some of these DUs may not be included on the final data file. First, to reduce the number of DU assignments required, segments consisting of only one block are flagged, and all respondents in the segment are assigned to the census block. Next, duplicates are eliminated such that multiple respondents at the same DU only require one assignment. This last step includes eliminating duplicate group quarter (GQ) units within the same GQ structure such that only one census block assignment is required. For example, a college dormitory with X rooms within the same building will have X-1 duplicate GQ units.

2. **Append address to each respondent DU.** Addresses are appended to the respondent DU file created in Step 1, and the file is uploaded to the system.

3. **Open census block assignment application.** The census block assignment application is included as a feature of the web-Setting and Zooming (web-SAZ) utility that was developed by RTI for NSDUH map production. Setting and zooming is the process of assigning each census block in a segment to a map page and zooming congested areas to allow sufficient space for marking DU locations on the resulting block listing maps. For census block assignments, the user follows the same steps that are used to open web-SAZ for map setting and zooming.

4. **Navigate to segment maps.** After specifying the state and quarter within the appropriate survey year's tracking system, the user selects a segment and selects "Start" to open the segment maps.

5. **Select a DU for assignment.** In web-SAZ, the user selects the "Locate DU" button in the toolbar, identified by the house icon (see the top left side of Exhibit 3, which appears after this appendix's text discussion). The user then selects a DU from the DU dropdown box at the bottom of the page. DUs needing assignment are in red font, and those with assignments already made are in black font. Once selected, the DU's address appears in the "DU Address" box next to the DU dropdown box.

6. **Locate DU on original field listing map.** The user opens the hard-copy segment maps that were listed in the field and locates the DU on the map. As shown in Exhibit 4, the DU locations are marked on the hard-copy map, but the census block boundaries are not visible. The user verifies that the address at the line number in the segment listing matches the address in the "DU Address" field. If the addresses do not match (e.g., if the address was edited during screening and interviewing and not updated in the hard-copy listing), further investigation is required to ensure that the DU on the hard-copy map and in web-SAZ are the same.

   If the DU was added to the sample during screening and interviewing (i.e., it was located on the property or premises of a sample DU or was added using special "bust" procedures, as described in Section F.6 of Appendix F), the DU location may not be specified on the map. In this case, further investigation is required to identify the sample DU to which the added DU is linked. The added DU is then placed at the same location as the sample DU. A bust situation involves a large number of missed DUs (e.g., a new or missed road or subdivision). Thus, DUs added through the special bust procedures are almost always drawn on the map.

7. **Specify map page of the DU.** The user selects the DU's map page from the "Zoom to Page for DU work" dropdown box (see Exhibit 3). The map page containing the DU's location is displayed.

8. **Click on map at DU location.** The user clicks the point on the electronic map where the DU has been spotted by the lister. An "X" will appear on the map with the associated DU

number above and the census block number below the "X." Although only the federal information processing standards (FIPS) code for the block is displayed for readability, the system captures the state, county, tract, and block group codes to uniquely identify the census block.

As shown in Exhibit 5, the census block boundaries are displayed with a thick gray line, which allows the user to see the census block boundaries even when a road or stream is placed over them. The FIPS code for the block is shown with a dark blue label. Three blocks make up the example in the exhibit, although all DUs are located within the same census block. The red square around the area displays the map page boundaries.

The census block number is automatically assigned based on the DU location inside the census block map layer. Thus, the system will only allow users to associate DUs with blocks that are part of the segment. If the user clicks on an area outside the segment, an error is displayed, and the user is asked to try again.

9. **Repeat Steps 5 through 8 for each respondent DU in the segment.** Once the census block assignments have been made for all respondent DUs in the segment, a pop-up will appear saying, "All DUs have been assigned to blocks for [current segment]." The same message appears at the bottom of the page as shown in Exhibit 6. The user then clicks the "complete" button in the toolbar, which is identified by the checkmark icon. The user is not able to mark a segment as complete until the census blocks for all respondent DUs in the segment have been assigned.

10. **Repeat Steps 4 through 9 for all segments.**

11. **Output census block data to a file.** The file containing the census block assignments for all respondent DUs is processed, quality checked, and output to a file that can be easily appended to the master data file. The variables that are appended to the data file include tract (TRACT10), block group (BLKGRP10), and census block (BLOCK10).[25]

D.3 Quality Control Procedures

Throughout the process, QC procedures are implemented to ensure the accurate assignment of respondents to census blocks. Some quality checks are built into the application, while others are completed during the postprocessing of the data.

D.3.1 Built-In Quality Control Checks

1. The user does not enter any census block numbers. Census blocks are automatically assigned based on the DU location inside the census block map layer. This ensures that there are no data entry errors and that only census blocks contained in the segment are assigned.

2. The user is not able to specify a DU location that is outside of the segment boundaries.

[25] State and county are already known for each segment and are delivered to the master data file separately.

3. The user cannot mark a segment as complete until the census blocks for all respondent DUs in the segment have been assigned.

D.3.2 Postprocessing Quality Control Checks

1. Confirm that all respondent DUs are assigned to a census block.

2. For each DU, confirm that the assigned census block is part of the segment associated with the DU.

3. Confirm that all final respondents have a census block assignment by merging to the master data file.

Exhibit 3 Census Block Assignment Application with No Dwelling Units Assigned

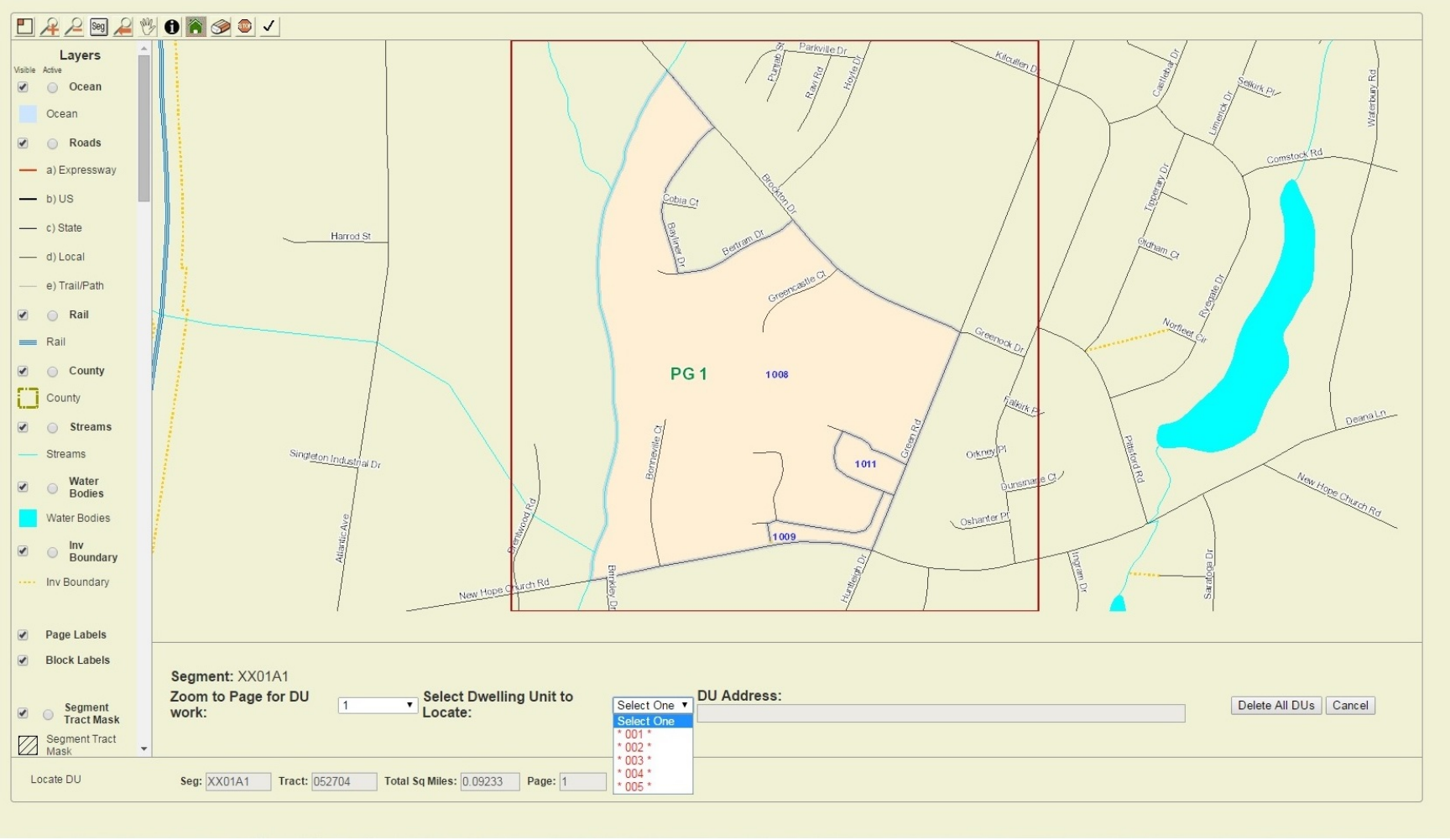


Exhibit 4 Field Listing Map

Exhibit 5 Census Block Assignment Application with Some Dwelling Units Assigned

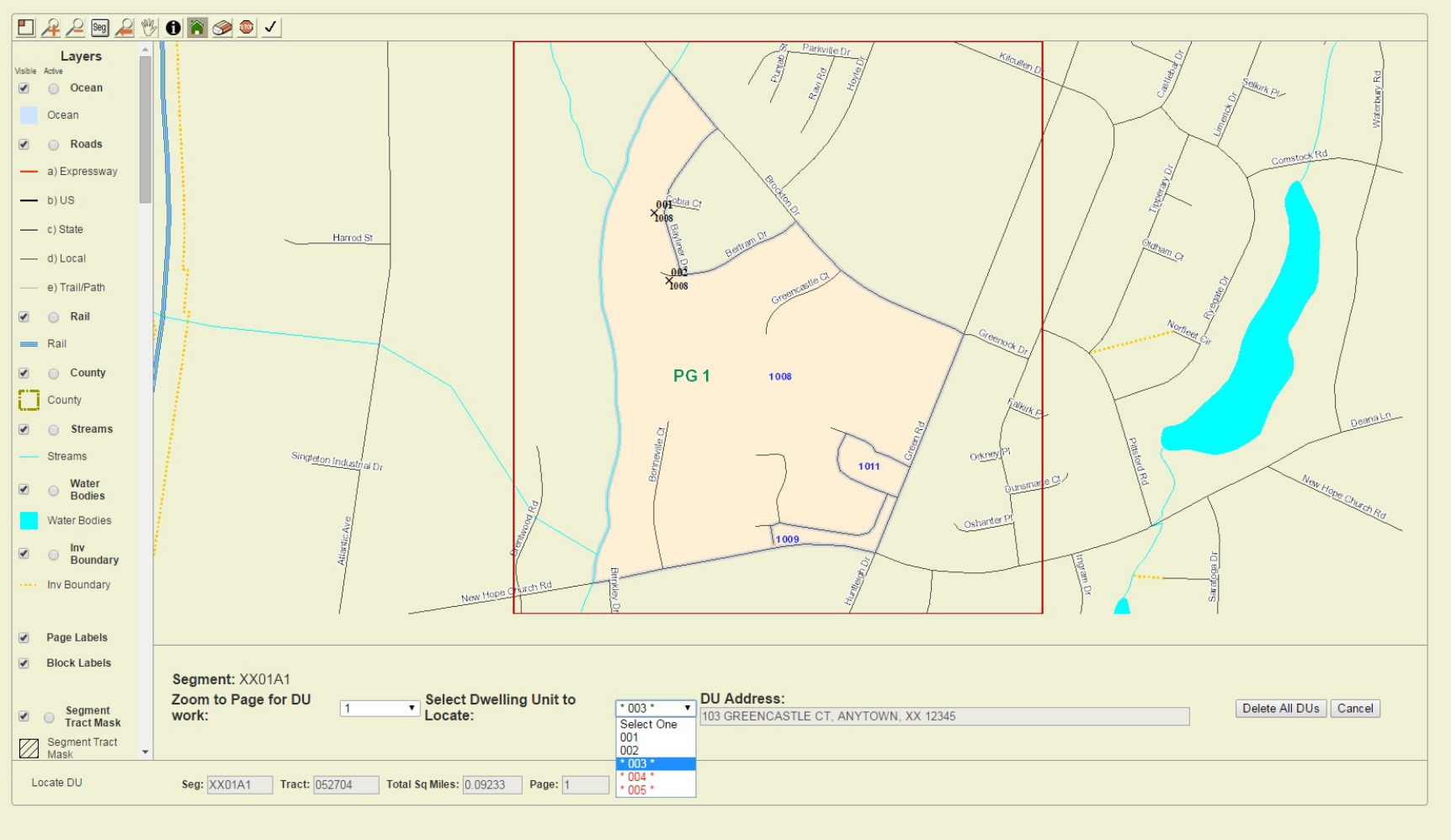
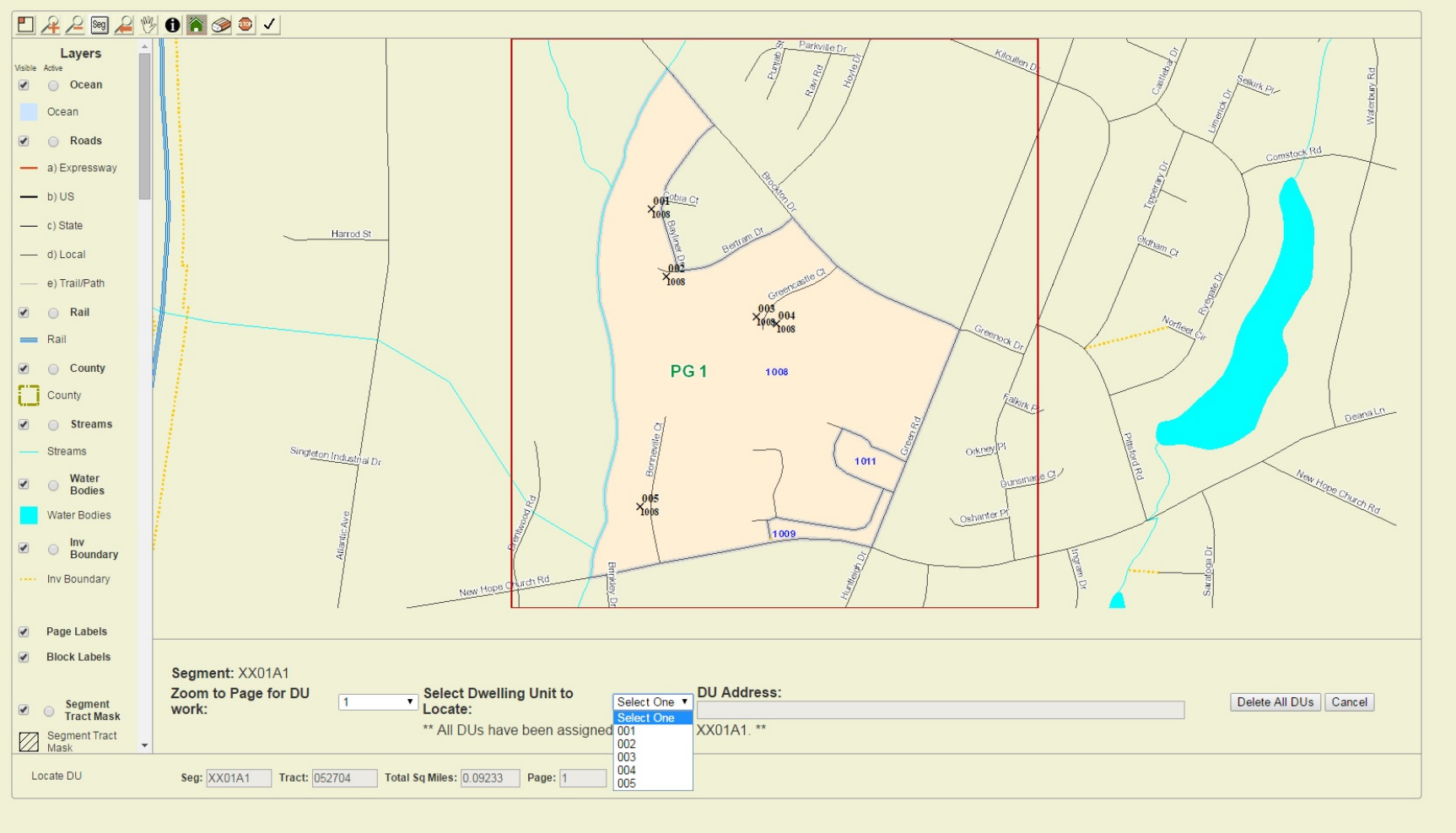


Exhibit 6 Census Block Assignment Application with All Dwelling Units Assigned



Appendix E: 2016 NSDUH Procedures for Subsegmenting

E.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.

E.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 census blocks, square miles, and dwelling units (DUs).[26] For the 2016 NSDUH, the two criteria for identifying candidate segments were as follows:

- number of census blocks > 1, square miles ≥ 75, and DU count ≥ 200; or
- number of census blocks > 1 and DU count ≥ 400.

Candidate segments then were evaluated to determine whether they could be subsegmented without input from the field. If feasible, the subsegmenting was 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.

E.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 E.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.

[26] DU counts were obtained from 2010 census data supplemented with revised population counts from Nielsen Claritas.

E.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, ...).

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 F: 2016 NSDUH Procedures for Adding Missed Dwelling Units

F.1 Introduction

The 2016 National Survey on Drug Use and Health (NSDUH) required field interviewers (FIs) to visit sample segments and conduct screenings and interviews in 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 2015. Because the listing was done a short time before the 2016 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 the majority of DUs to be given a chance of being selected, the 2016 NSDUH had in place a procedure for checking for and adding missed DUs on the premises of sampled DUs. During the screening interview, FIs asked the screening respondent about other units on the property of the selected DU. If the number of added DUs linked to any particular 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. However, if either of these limits was exceeded, the FI contacted the sampling support office for subsampling to be considered.

Unlike the 2005 through 2013 NSDUHs but similar to the 2014 and 2015 NSDUHs, the half-open interval (HOI) rule was not implemented in the 2016 NSDUH. This procedure requires FIs to look between each selected DU and the next listed DU to identify any unlisted units. Prior research suggested that only 0.2 percent of the total DUs on the NSDUH frame were added through the HOI rule (Iannacchione et al., 2012). Eliminating the HOI rule in 2014 decreased the burden on interviewers and simplified FI training and the screening process. This decrease in burden outweighed the small amount of coverage afforded by the HOI rule.

To minimize bias associated with large numbers of missed DUs (e.g., from new construction or a missed subdivision), FIs were instructed to contact their supervisors if they noticed large discrepancies between the segment listing and what they encountered in the field. If the FI encountered 150 or more total unlisted units or 50 or more missed DUs following any one DU, the situation qualified as a "bust" and special sampling procedures were employed (see Section F.6).

This appendix outlines the 2016 NSDUH procedures for adding missed DUs. For this appendix, procedures for adding missed DUs are classified into three categories: adding HUs, adding GQ units, and busts. Table F.1 compares the 2014 through 2016 procedures with those used for the 2013 NSDUH.

Table F.1 Comparison of 2013 and 2014-2016 NSDUH Procedures for Adding Missed Dwelling Units

Missed Dwelling Unit Scenario	2013 Procedure	2014-2016 Procedure
Regular housing units (e.g., houses, townhouses, duplexes, trailers)	During the screening interview, the respondent was asked the following question: "Are there any other living quarters within this structure or on this property, such as a separate apartment with a separate entrance?" If the response was "yes," the FI recorded the address of the possible missed unit and added the unit to the sample if it was not on the original handwritten list of DUs. In addition, the FI checked the interval between the SDU and the next listed DU for these types of DUs. Missed DUs found in the HOI were added to the sample.	The FI continued to ask the screening respondent about other DUs on the property of the SDU. Missed DUs identified during the screening interview were added to the sample. The HOI rule was not implemented to pick up these types of DUs in the interval between an SDU and the next listed DU.
Units in apartment and condo buildings	Missed DUs in an apartment building were picked up via the HOI rule; that is, each missed unit was linked to the preceding listing unit in the path of travel.	Missed DUs in apartment buildings were not added unless the situation qualified as a "bust" or in the unusual situation that there was a "unit within a unit" (e.g., maid or nanny quarters within a large apartment or flat).
Missed apartment buildings	Missed apartment buildings were picked up via the HOI rule.	Missed apartment buildings were not added unless the situation qualified as a bust.
Missed DUs in GQ structures (e.g., dormitories, shelters, boarding houses)	Missed GQ units were sampled at the same rate as the original listing.	Missed GQ units continued to be sampled at the same rate as the original listing.
Missed GQ structures	GQ structures were added via the HOI rule.	GQ structures were not added unless the situation qualified as a bust.
Major discrepancies (new subdivision, missed floor or wing in an apartment building, etc.)	The FI was instructed to call his or her supervisor if he or she noticed large omissions or changes to the area. These situations were handled using bust procedures; that is, a subsample of the missed DUs were added if there were 50 or more missed DUs following any one DU or 150 or more total missed DUs in the segment.	FIs were instructed to call their supervisors if they noticed major discrepancies between the segment listing and ground truth. Bust procedures were implemented if the situation qualified as a bust.

DU = dwelling unit; FI = field interviewer; GQ = group quarters; HOI = half-open interval; SDU = sample dwelling unit.

Note: 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.

F.2 Subsampling of Dwelling Units

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 and continued through the 2016 survey:

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

F.3 Procedure for Adding Housing Units

This section refers to HUs that were 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 was used to resolve added DU problems.

1. Once the limit of 5 (or 10) rule was exceeded, the FI was instructed to stop screening and interviewing activities on added HUs and contact the sampling support office. The FI was then instructed to do a quick check of the segment for major discrepancies in the segment listing. At this time, the FI completed a paper list of added HUs for the entire segment.
2. Once the final list of added HUs was received by the sampling support office, the following was done:
   (a) Sampling examined the added HUs and determined whether they were linked to a sample dwelling unit (SDU).
   (b) If the number of added HUs linked to any *one* SDU exceeded 50, these units were treated as a bust (see Section F.6).
   (c) If the number of added HUs linked to any *one* nonsampled DU exceeded 50, these units also were treated using the procedure for busts (see Section F.6).
   (d) Sampling staff calculated 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 screening interview (i.e., on the premises of selected DUs).
   (e) If the total number of added DUs exceeded 10, a subsampling rate was determined using the criteria above.
3. The computing division added the DUs to the system and subsampled if necessary:
   (a) Data entry of the added DUs was done. Entries were made for all units that collectively qualified as a bust and units obtained through the missed DU

procedures—not for all missed DUs found in the segment. The link number then was entered and a line number was assigned.[27] For DUs obtained through the bust procedure, the sampling link number (SLN) also was recorded. Finally, it was necessary to check that none of the DUs had already been entered in the handheld computer so that DUs did not appear in the system twice.

(b) DUs were selected from the added DUs at the rate defined above. The subsampling rate was recorded in a data field.

(c) Probabilities of selection were brought over as appropriate for the segment.

(d) A random number was added for the screening selection algorithm.

4. Selected DUs were added to the FI's assignment during the next transmission.

5. A sample weight was assigned to each added DU. If the total number of added DUs was fewer than or equal to 10, each added DU was assigned the weight of the original selected DUs in the segment. If subsampling was required, the selected DU weight was adjusted by the inverse of the subsampling rate for each added DU.

F.4 Procedure for Adding Group Quarters Structures

If an entire GQ structure was not listed (or was erroneously listed as an HU), the GQ structure was not added to the sample. The exception to this rule was if the number of GQ units in the missed GQ structure exceeded 50. In this case, the bust procedure was applied (see Section F.6).

F.5 Procedure for Adding Group Quarters Units

In the case of discrepant GQ listings, two approaches were taken depending on whether the actual number of GQ units was less than or greater than the number of GQ units listed in advance.

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

In the case that there were extra GQ units listed, the units at the end of the list were assigned an ineligible code, such as "Listing Error." All other units remained eligible.

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

If there were more GQ units in the structure than were previously listed, a complete list was made, and the units were 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

[27] 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
5
6
7
8
9
10
11

The additional units then were numbered consecutively, and an SLN corresponding to each of the originally listed units was assigned. Next, the added GQ units with SLNs corresponding to the original selected units were 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**

F.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 was identified to relist the segment as quickly as possible. Otherwise, the following approach was employed.

First, if any DUs disappeared since the time of the listing, all selected "disappears" were assigned an "ineligible" final screening code. Then any new DUs were listed consecutively, assigned an SLN, and added to the sample if the SLN corresponded 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 were 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 was greater than 10, then resampling occurred from all nonfinalized DUs as described in Section F.3.

F.7 Quality Control

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

- Sampling staff ensured that the correct information was keyed by data entry.
- Checks within the computing division were performed.
- Sampling staff checked the number of selected DUs and the person probabilities of selection assigned to each DU selected in the subsampling routine.