SAMHSA
Substance Abuse and Mental Health Services Administration

2020 National Survey on Drug Use and Health (NSDUH) Methodological Resource Book

Section 2: Sample Design Report

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

May 2021

2020 National Survey on Drug Use and Health (NSDUH) Methodological Resource Book, Section 2: Sample Design Report

Acknowledgments

This report was prepared for the Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services (HHS), under Contract No. HHSS283201700002C with RTI International. Rong Cai served as the government project officer and as the contracting officer representative.

Electronic Access

This publication may be downloaded at https://www.samhsa.gov/data/.

Recommended Citation

Center for Behavioral Health Statistics and Quality. (2021). *2020 National Survey on Drug Use and Health (NSDUH) methodological resource book, Section 2: Sample design report*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Originating Office

Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, 5600 Fishers Lane, Room 15-E09D, Rockville, MD 20857. For questions about this report, please e-mail CBHSQrequest@samhsa.hhs.gov.

U.S. Department of Health and Human Services
Substance Abuse and Mental Health Services Administration
Center for Behavioral Health Statistics and Quality
Populations Survey Branch

May 2021

Table of Contents

Chapter **Page**

1. Introduction ... 1
   - 1.1 Purpose ... 1
   - 1.2 Target Population ... 1
   - 1.3 Overview ... 2
   - 1.4 Impact of COVID-19 on Sample Design and Data Collection ... 3

2. Coordinated Sample Design for the 2014-2022 NSDUHs ... 5
   - 2.1 Multiyear Design ... 5
   - 2.2 Stratification and First-, Second-, and Third-Stage Sample Selections ... 6
   - 2.3 Sample Dwelling Units and Persons ... 9
   - 2.4 Impact of COVID-19 on the Coordinated Sample Design ... 11

3. Selection of First-, Second-, and Third-Stage Sampling Units for the 2014-2022 Coordinated Sample ... 13
   - 3.1 Formation of and Objectives for Using the Composite Size Measures ... 13
   - 3.2 Stratification ... 14
   - 3.3 First-, Second-, and Third-Stage Sample Selection ... 15
   - 3.4 Survey Year and Quarter Assignment ... 17
   - 3.5 Impact of COVID-19 on Selection of First-, Second-, and Third-Stage Sampling Units ... 18

4. Selection of Dwelling Units and Persons for the 2020 NSDUH ... 21
   - 4.1 Notation ... 22
   - 4.2 Determining Person Sample Sizes, by State and Age Group ... 24
     - 4.2.1 Original Sample ... 24
     - 4.2.2 Supplemental Sample ... 25
     - 4.2.3 Expanded Sample ... 28
   - 4.3 Fourth-Stage Sample Determination for Each Segment ... 31
     - 4.3.1 Dwelling Unit Frame Construction—Counting and Listing ... 31
     - 4.3.2 Determining Dwelling Unit Sample Size ... 35
   - 4.4 Determining Fifth-Stage Sample (Person) Selection Probabilities for Each Segment ... 36
   - 4.5 Sample Size Constraints: Guaranteeing Sufficient Sample for Additional Studies and Reducing Field Interviewer Burden ... 36
   - 4.6 Dwelling Unit Selection and Release Partitioning ... 37
   - 4.7 Procedures for Adding Dwelling Units ... 38
   - 4.8 Quarter-by-Quarter Deviations ... 38
   - 4.9 Impact of COVID-19 on Dwelling Unit and Person Samples ... 42

5. Sample Weighting and Variance Estimation ... 45
   - 5.1 Sample Weighting Procedures ... 45
     - 5.1.1 Main Study Design Weights ... 45
     - 5.1.2 Main Study Weight Adjustments ... 45
     - 5.1.3 Main Study Analysis Weights ... 46
     - 5.1.4 Quality Control Measures in Design-Based Weighting Procedures ... 49

Table of Contents (continued)

Chapter **Page**

5.2 Creation of Variance Strata and Variance Primary Sampling Units for Person-Level Estimates 49
5.3 Creation of Variance Strata and Variance Primary Sampling Units for Pair Analyses 51
5.4 Other Sampling-Related Variables 52
5.5 Impact of COVID-19 on Sample Weighting and Variance Estimation 53

6. Clinical Validation Study Sample Design 55
6.1 Overview 55
6.2 CVS Sample Size Determination 56
6.3 CVS Sample Allocation 58
6.4 CVS Sample Selection 60
6.5 CVS Sample Results 60
6.6 CVS Weighting and Variance Estimation 61
6.7 Impact of COVID-19 on the CVS Sample Design 61

References 63

List of Contributors 67

Appendix

A Comparison of the 2014 through 2022 NSDUH Sample Design with the 2005 through 2013 NSDUH Sample Design 69

B 2014 through 2022 NSDUH State Sampling Regions 75

C 2020 NSDUH Pair Sampling and Selection of the Pair Sampling Parameter 127

D 2020 NSDUH Procedures for Subsegmenting 131

E 2020 NSDUH Procedures for Adding Missed Dwelling Units 133

F Expected Value of the Collapsed Stratum Estimator as Applied to the NSDUH "with Replacement" Variance Estimator 139

G 2020 NSDUH Census Block Assignment Procedures 143

List of Tables

Table		Page
2.1	Annual National Sample of Area Segments and Respondents	6
3.1	Number of Census Tracts, Block Groups, and Segments on Sampling Frame, by State	16
3.2	Segment Identification Number Suffixes and Quarter Assignment	19
4.1	2020 NSDUH Sample Sizes and Projected Respondents, by State and Age Group	26
4.2	2020 NSDUH Sample Sizes, by State	27
4.3	Relative Standard Errors and Sample Sizes for Key Outcome Measures, by Demographic Domain	30
4.4	Number of Map Pages, by State and Segment	32
4.5	Segment and Dwelling Unit Summary	34
6.1	Expected Sample Sizes and Outcome Rates for the Clinical Validation Study (Quarter 1)	57
6.2	Quarter 1 CVS Sample Allocation and Sampling Rates, by CAM Stratum	59
6.3	Achieved Sample Sizes and Outcome Rates for the Clinical Validation Study	60
A.1	Sources of Geographic Data Used in NSDUH Sample Designs: 1999 through 2022	69
A.2	Sample Sizes and Targeted Respondents, by State and Age Group: 2005 through 2013	69
A.3	Sample Sizes and Projected Respondents, by State and Age Group: 2014 through 2022	71
A.4	Population and Sample Percentages, by State: 2013 and 2014 through 2022	71
C.1	Expected Pair Selection Counts and Response Rates for $\lambda = 0.25$	130
E.1	Comparison of 2013 and 2014 with 2020 NSDUH Procedures for Adding Missed Dwelling Units	134

This page intentionally left blank

List of Exhibits

Exhibit		Page
1	2020 NSDUH Expected Respondent Samples	3
2	Hierarchy of Census and NSDUH Geographic Entities	7
3	Sample Weight Components	48
4	Structure of the Clinical Validation Study Sampling Design	56
5	Census Block Assignment Application with No Dwelling Units Assigned	147
6	Field Listing Map	148
7	Census Block Assignment Application with Some Dwelling Units Assigned	149
8	Census Block Assignment Application with All Dwelling Units Assigned	150

This page intentionally left blank

1. Introduction

1.1 Purpose

The goal of this report is to document the sample design for the 2020 National Survey on Drug Use and Health (NSDUH),[1] including the design of a Clinical Validation Study (CVS) that was conducted during the first 3 months of the 2020 NSDUH.[2] The 2020 NSDUH sample experiences will be further documented in the forthcoming 2020 NSDUH sample experience report.

The report is organized into six chapters and includes a list of cited references, a list of contributors, and several appendices. The remainder of Chapter 1 describes the target population, provides a general overview of the 2020 NSDUH sample design, and summarizes the impact of the COVID-19 pandemic on the 2020 NSDUH. Chapter 2 provides a summary of the 2014 through 2022 coordinated design, while Chapter 3 discusses the first three stages of selection in greater detail. Chapter 4 describes features of the sample design that were employed in 2020 (i.e., fourth and fifth stages of selection), and Chapter 5 describes sample weighting and variance estimation procedures for the 2020 NSDUH. Finally, Chapter 6 summarizes the CVS sample design.

1.2 Target Population

The respondent universe for the 2020 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 2020 NSDUH universe included residents of housing units (e.g., single-family houses, townhomes, apartments, mobile homes, rented rooms), noninstitutional group quarters[3] (e.g., shelters, rooming houses, dormitories, and group homes), and civilians residing on military bases. Coverage before the 1991 survey was limited to residents of the coterminous 48 states (i.e., excluding residents of Alaska and Hawaii), and it excluded residents of group quarters and all persons (including civilians) living on military bases. Persons excluded from the 2020 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] This report presents information from the 2020 NSDUH. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[2] The CVS was originally planned for the first 6 months (quarters 1 and 2) of the 2020 NSDUH.

[3] NSDUH defines group quarters as dwelling units in which 10 or more unrelated persons reside. The following dwelling units are considered group quarters regardless of the number or relationship of occupants: college dormitories, sorority houses, and fraternity houses; quarters for live-in staff members of institutions; and missions and shelters. Other noninstitutional group quarters include boarding or rooming houses, communes, convents and monasteries, halfway houses, migratory workers' camps, and nonmilitary barracks. Institutional group quarters, which are ineligible for NSDUH, include correctional facilities, mental institutions, nursing homes, and hospitals.

1.3 Overview

A coordinated sample design was developed for the 2014 through 2017 NSDUHs. A large reserve sample of area clusters or segments was selected at the time the 2014 through 2017 NSDUH sample was selected. This reserve sample was (or will be) used to field the 2018 through 2022 NSDUHs. Thus, the 2018 through 2022 NSDUH designs simply continue the coordinated design. The coordinated sample design is state-based, with an independent, multistage area probability sample within each state and the District of Columbia.

Similar to NSDUHs dating back to 1999, the 2014 through 2022 surveys are conducted using computer-assisted interviewing (CAI) methods and provide sufficient sample sizes to support state and national estimates. Furthermore, NSDUH was redesigned in 2014 to a more cost-efficient sample allocation, which increases the sample size in the largest states while maintaining adequate sample sizes in smaller states to support reliable state estimates based on small area estimation methodology. Reliable direct state estimates are also possible (in any state) by pooling multiple years of data.

A new pair sampling strategy was implemented in 2002 that increased the number of pairs selected in dwelling units with older persons on the roster (Chromy & Penne, 2002). With the increase in the number of pairs came a moderate decrease in the response rate for older persons. 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 through 2020 NSDUHs.

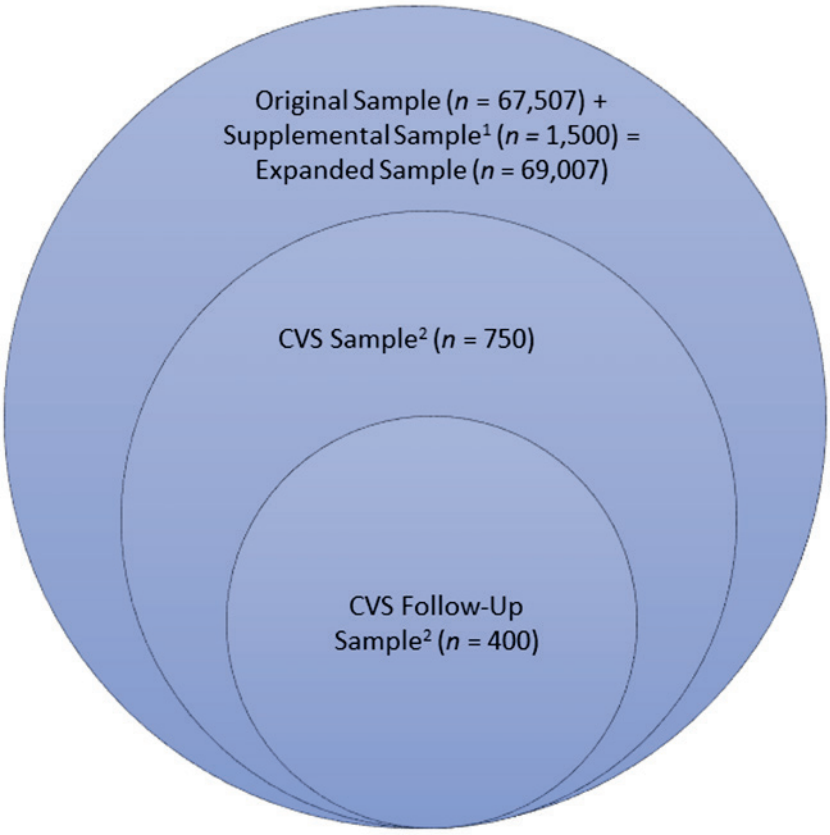
For the 2020 NSDUH, the original target national sample size of 67,507 was distributed across five age groups as follows: 25 percent for youths aged 12 to 17, 25 percent for young adults aged 18 to 25, 15 percent for adults aged 26 to 34, 20 percent for adults aged 35 to 49, and 15 percent for adults aged 50 or older. This original target sample size would allow the Substance Abuse and Mental Health Services Administration (SAMHSA) to report precise estimates for demographic subgroups at the national level without needing to oversample specially targeted demographics. To maintain the original sample sizes by state and age group if the decision was later made to remove the CVS cases (e.g., if CVS methods were found to significantly affect survey responses), the quarter 1 and 2 sample targets were supplemented with a total of 1,500 interviews proportionally allocated to states and age groups (see Section 4.2).[4] The expanded sample size of 69,007 interviews, which includes the original sample and the supplemental sample, is also referred to as the "main study sample."

The CVS was originally planned for the first 6 months of the 2020 NSDUH to assess a revised NSDUH module on substance use disorders (SUDs) designed to be consistent with the criteria in the *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition (DSM-5) (American Psychiatric Association [APA], 2013). Because of COVID-19, the CVS was limited to quarter 1. A probability subsample of 750 adult and youth respondents (the CVS sample) was expected to be selected by the CAI instrument from the expanded sample (i.e., the original and

[4] Although the CVS was conducted in only quarter 1 of the 2020 NSDUH, the quarter 1 and 2 interview targets were increased by 750 interviews because the dwelling unit samples were selected before the decision was made to suspend data collection. As noted in Sections 1.4 and 4.8, the quarter 2 sample was further supplemented to partially compensate for the negative impact of the pandemic on data collection and response rates.

supplemental samples) in quarter 1. Based on their responses to questions about past year use of cigarettes, alcohol, and marijuana, respondents selected for the CVS were routed to the new DSM-5 SUD modules instead of the modules using the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (APA, 1994). At the end of the NSDUH interview, the interviewers invited CVS sample respondents to participate in a follow-up clinical interview by telephone. Of the 750 respondents selected for the CVS sample, 400 were expected to complete the clinical follow-up interview in quarter 1.[5] Exhibit 1 shows the planned relationship between the expanded sample, the CVS sample, and the CVS follow-up sample.

Exhibit 1 2020 NSDUH Expected Respondent Samples



CVS = Clinical Validation Study.
[1] Included in the quarter 1 and 2 samples.
[2] Included in only quarter 1 of the 2020 NSDUH .

1.4 Impact of COVID-19 on Sample Design and Data Collection

On March 16, 2020, given the public health emergency related to COVID-19 and considering the safety of the field staff and the public, SAMHSA made the decision to immediately suspend NSDUH in-person data collection. NSDUH project management at RTI

[5] See Section 6.2 for a discussion of the CVS sample size determination.

International[6] began working closely with SAMHSA and RTI's Infectious Disease Response Team (IDRT), Executive Leadership Team (ELT), and Institutional Review Board (IRB) to determine when and where it was safe to resume in-person data collection. To assess the feasibility of resuming in-person data collection, the IDRT advised (and SAMHSA, the ELT, and the IRB approved) a small-scale data collection that was conducted from July 16 to 22, 2020, in select counties in two states where it was deemed safe based on state- and county-level health metrics. During the 1-week data collection, RTI monitored health metrics daily to determine whether it was safe to continue data collection. During this period, the majority of the counties became ineligible, and data collection activities had to be stopped in those counties.

For the remainder of 2020, the COVID-19 pandemic made it nearly impossible to perform the conventional in-person data collection, which would potentially reduce the respondent sample size to an unacceptable level. As a measure of minimizing the pandemic's impact on NSDUH data collection, SAMHSA approved the addition of web data collection. As a result, quarter 4 became a period of multimode data collection. During this time, all sample dwelling units (SDUs) originally selected from the quarter 2, 3, and 4 area segments for in-person data collection were released for web or in-person data collection. SDUs in the quarter 2, 3, and 4 area segments deemed safe for in-person data collection were released to the field. SDUs in the remaining quarter 2, 3, and 4 area segments (except in segments that were worked by field interviewers during the 1-week, small-scale data collection) were mailed an invitation to participate via web. In addition, as described in Section 4.8, additional SDUs were selected in some quarter 2 and 3 segments and released to web data collection in quarter 4 to partially compensate for the negative impact of the pandemic on data collection and response rates. Throughout the multimode data collection period, health metrics were monitored, and areas that became ineligible for in-person data collection were transferred to web data collection.

As a result of the COVID-19 pandemic, the achieved sample for the 2020 NSDUH was 36,284 persons with 17,082 interviews completed in person and 19,202 interviews completed via web. Further, 15,628 interviews were completed in quarter 1, 42 were completed during the July 2020 small-scale data collection, and 20,614 were completed during the multimode data collection period. The achieved sample for the 2020 NSDUH is sufficient for producing national estimates for persons aged 12 or older and adults aged 18 or older but may not be sufficient for producing estimates for smaller demographic groups (Morton et al., 2020).

As noted previously, the CVS was originally planned for quarters 1 and 2 of the 2020 NSDUH. Because data collection was suspended on March 16, 2020 and did not resume until the fourth calendar quarter, the CVS was limited to the first 3 months of data collection. In quarter 1, a total of 699 respondents were selected for the CVS, and 424 of these respondents completed the follow-up interview (see Chapter 6).

Additional detail on the impact of the COVID-19 pandemic on each stage of the 2020 NSDUH sample design is provided in the chapters that follow.

[6] RTI International is a trade name of Research Triangle Institute. RTI and the RTI logo are U.S. registered trademarks of Research Triangle Institute.

2. Coordinated Sample Design for the 2014-2022 NSDUHs

2.1 Multiyear Design

A coordinated sample design was developed for the 2014 through 2017 National Surveys on Drug Use and Health (NSDUHs) and is being extended to the 2018 through 2022 NSDUHs. Appendix A provides a comparison of the 2014 through 2022 design with the 2005 through 2013 design. The 2014 through 2022 coordinated design facilitates 50 percent overlap in third-stage units (area segments) within each successive 2-year period from 2014 through 2022. This designed sample overlap may slightly increase the precision of estimates of year-to-year trends when there is a small positive correlation in successive survey years due to an overlapped segment being somewhat homogeneous. The 50 percent overlap of segments significantly reduces segment listing costs because only one half of the segments needs to be listed for the 2015 through 2022 surveys.

The 2014 through 2022 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 2.1, each annual survey 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 allows 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 is being treated separately from the remainder of Hawaii for sample allocation and sample size management purposes. The annual sample in Hawaii consists 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 2.1 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.

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

Exhibit 2 displays the relationship between census and NSDUH geographic entities. First, state sampling regions (SSRs) were formed by combining census tracts within each state. Each state was geographically partitioned into roughly equal-sized regions according to a composite size measure (i.e., population weighted by state and age group sampling rates). In other words, regions were formed such that each area within a state yielded, in expectation, roughly the same number of interviews during each data collection period. This partitioning divided the United States into 750 SSRs. Maps for these regions can be found in Appendix B.

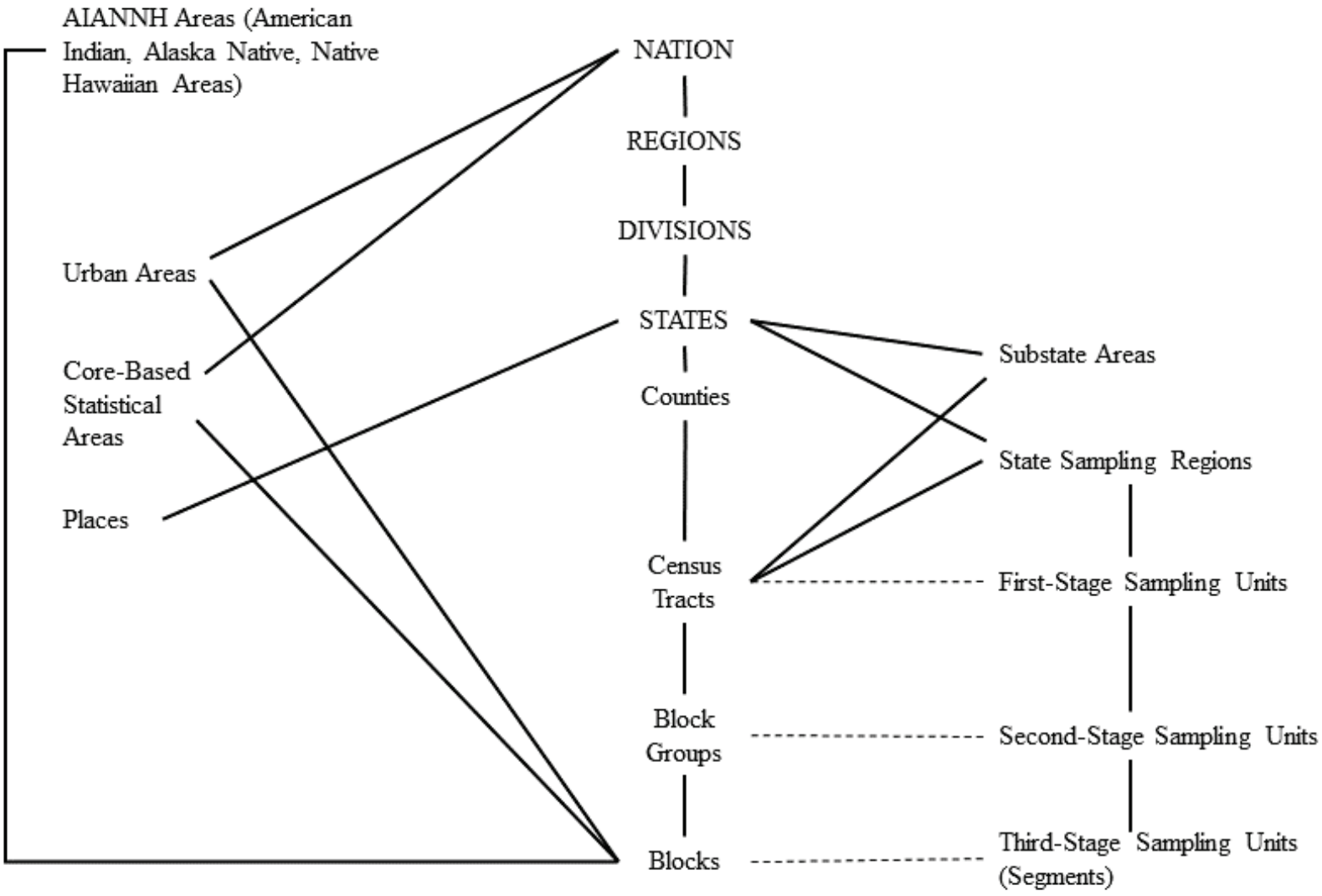
Similar to the 2005 through 2013 NSDUHs, the first stage of selection for the 2014 through 2022 NSDUHs was census tracts.[7] This stage was included to contain sample segments within a single census tract to the extent possible.[8] Segments that cross census tract boundaries make merging to external 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

[7] 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, 2001).

[8] Some census tracts had to be aggregated in order to meet the minimum dwelling unit (DU) requirement.

Exhibit 2 Hierarchy of Census and NSDUH Geographic Entities



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, Ohio, Pennsylvania, Texas, and Virginia, this minimum size requirement was 250 DUs[9] in urban areas and 200 DUs in rural areas.[10] 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,[11] additional implicit stratification was achieved by sorting the first-stage sampling units by a CBSA/SES[12] (core-based statistical area/socioeconomic status) indicator[13] and by the percentage of the population that is non-Hispanic and white.[14] 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[15] 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 in the 2014 through 2022 NSDUHs to facilitate possible transitioning to an address-based sampling (ABS) design.

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

[9] DU counts were obtained from the 2010 census data supplemented with revised population counts from Claritas, which is a market research firm headquartered in Cincinnati, Ohio (see https://www.claritas.com/).

[10] 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 sample states, and 200 DUs are sufficient to support three samples in the larger sample states.

[11] 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.

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

[13] 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."

[14] Although the large sample size eliminates the need for the oversampling of specially targeted demographic subgroups as was required prior to the 1999 National Household Survey on Drug Abuse, 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 selecting neighboring and possibly similar segments at a lower probability than if the selection were done completely at random.

[15] 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[16] of DUs by aggregating adjacent census blocks.[17] Consistent with the terminology used in previous NSDUHs, these geographic clusters of blocks are referred to as "segments." 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 the Substance Abuse and Mental Health Services Administration 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 3.4.

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 2020, the first panel segments (panel G) were used for the 2019 and 2020 surveys, constituting the overlap sample. The second panel segments (panel H) were used for the 2020 survey and are being used again for the 2021 survey.

2.3 Sample Dwelling Units and Persons

After sample segments for the 2020 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. A DU in NSDUH refers to either a housing unit or a group quarters listing unit, such as a dormitory room or a shelter bed. The lists of DUs compiled by the field household listers 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 subsegmentation inflation factor,[18] if any; the probability of selecting a person in the age group (equal to the

[16] 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).

[17] 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, 2001).

[18] 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.

maximum, or 0.99, for the driving age group); and an adjustment for the "maximum of two" rule.[19] In addition to these factors, historical data from the 2018, 2019, and 2020 NSDUHs were used to compute predicted screening and interviewing response rate adjustments.[20,21] 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 120 per segment or half of the actual listing unit count.[22] 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 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 2005 to 2013 may be selected by chance in 2014 to 2022. No mechanism is currently in place for identifying duplicate persons across years within a sample design (e.g., 2014 to 2022) 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 included in the 2014 through 2022 NSDUHs was that any pair of 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). This pair sampling algorithm 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

[19] 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.

[20] Screening and interview response rates were computed using the American Association for Public Opinion Research (AAPOR) formulas for Response Rate 1 (RR1) and Response Rate 2 (RR2), respectively. For more information, see pp. 61 and 62 of the 9th edition of their *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (AAPOR, 2016).

[21] Data from quarter 1 of the 2020 NSDUH were used to compute screening and interview response rate adjustments for the quarter 3 and 4 samples (see Section 4.8).

[22] Prior to the 2020 NSDUH, the listing unit sample size could not exceed 100 per segment or half of the actual listing unit count.

of pairs selected. Appendix C describes the simulation analyses that were conducted to select the pair sampling parameter for the 2014 through 2022 NSDUHs.

As in previous years, during the screening interview, the interviewer asked the screening respondent about other units within the sampled DU or on the property (e.g., a garage apartment). When found on the premises of a sampled DU, any new or missed dwelling was selected into the 2020 NSDUH. However, unlike the 2005 through 2013 NSDUHs, the half-open interval (HOI) procedure[23] was discontinued 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 et al., 2012). Excluding the HOI procedure decreases the burden on field interviewers 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 between the segment listing and what they encountered in the field. Then special "bust" procedures were implemented, as described in Section 4.7.

2.4 Impact of COVID-19 on the Coordinated Sample Design

Because area samples were selected simultaneously for the 2014 through 2022 NSDUHs, COVID-19 had no impact on stratification or the first three stages of selection in the coordinated sample design. The impact on the selection of DUs and persons is discussed in Section 4.9.

[23] 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.

This page intentionally left blank

3. Selection of First-, Second-, and Third-Stage Sampling Units for the 2014-2022 Coordinated Sample

As was mentioned previously, the first three stages of sample were selected simultaneously for each of the 2014 through 2022 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 2014 through 2017 main studies and a number of large field tests. A number of "reserve" segments also were selected to carry the sample through the next decennial census, if desired. This reserve sample is being used to support the 2018 through 2022 NSDUHs.

3.1 Formation of and Objectives for Using the Composite Size Measures

The composite size measure procedure is used to obtain self-weighting[24] 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 resulting 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.

[24] 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

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

where N_{hij} equals the number of blocks within segment j of SSR i and state h.

3.2 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. Using desktop computer mapping software,[25] a technician selected whole or partial counties (census tracts) and assigned them to an SSR. This process was repeated until each aggregate area had a scaled composite size measure of roughly 100.[26] When forming the regions, the technician also considered 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 2022 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

[25] RTI developed this software for the purpose of forming NSDUH SSRs.

[26] Prior to forming the SSRs, composite size measures were scaled so that the aggregate composite size measure was roughly 100 per region. The scaling factor was equal to 100 divided by the expected sample size per SSR in the state (i.e., the state target sample size divided by the number of SSRs). For example, in California, the scaling factor was equal to 100/(4,560/36) = 0.789. This scaling made it easier for the technician when forming the regions but served no other purpose. Without scaling, the composite size measures would sum to the expected sample size per region, which varies by state.

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.

3.3 First-, Second-, and Third-Stage Sample Selection

Once the SSRs were formed, the first-stage sampling units were created by aggregating adjacent census tracts within regions until the resulting first-stage sampling units met the minimum DU 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 core-based statistical area/socioeconomic status indicator and by the percentage of the population that is non-Hispanic and white. Table 3.1 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 within selected census tracts until the resulting second-stage sampling units met the minimum DU requirement. Second-stage units or "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 proportional 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 aggregated 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 centers or centroids were serpentine-sorted by latitude and longitude (i.e., the blocks were sorted in an alternating ascending and descending pattern so that any two consecutive blocks on the sorted frame within a block group were geographically adjacent).[27] 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.

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

Table 3.1 Number of Census Tracts, Block Groups, and Segments on Sampling Frame, by State

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 2014-2022 Sample	Unique Segments in 2014-2022 Sample
Total United States			72,006	96,898	99,137	36,000	30,000	28,608
Northeast								
Connecticut	CT	09	824	1,837	1,646	576	480	477
Maine	ME	23	351	1,037	1,960	576	480	430
Massachusetts	MA	25	1,458	2,100	1,690	576	480	480
New Hampshire	NH	33	292	907	1,883	576	480	415
New Jersey	NJ	34	1,988	2,417	1,321	720	600	600
New York	NY	36	4,798	4,956	2,260	1,440	1,200	1,200
Pennsylvania	PA	42	3,184	3,849	2,202	1,152	960	960
Rhode Island	RI	44	240	796	1,486	576	480	381
Vermont	VT	50	183	515	1,363	576	480	373
Midwest								
Illinois	IL	17	3,107	3,706	2,311	1,152	960	959
Indiana	IN	18	1,503	1,968	2,114	576	480	479
Iowa	IA	19	823	1,781	1,830	576	480	475
Kansas	KS	20	758	1,728	1,840	576	480	473
Michigan	MI	26	2,736	3,476	2,078	1,152	960	960
Minnesota	MN	27	1,332	1,891	1,996	576	480	478
Missouri	MO	29	1,386	2,014	2,035	576	480	480
Nebraska	NE	31	529	1,471	1,602	576	480	447
North Dakota	ND	38	205	569	1,452	576	480	382
Ohio	OH	39	2,931	3,729	2,253	1,152	960	959
South Dakota	SD	46	221	644	1,555	576	480	394
Wisconsin	WI	55	1,390	1,975	1,925	576	480	479
South								
Alabama	AL	01	1,174	1,801	2,224	576	480	480
Arkansas	AR	05	683	1,708	2,146	576	480	475
Delaware	DE	10	214	546	1,480	576	480	388
District of Columbia	DC	11	178	448	1,057	576	480	345
Florida	FL	12	4,148	4,009	3,984	1,440	1,200	1,199
Georgia	GA	13	1,949	2,142	1,820	720	600	600
Kentucky	KY	21	1,104	1,875	1,876	576	480	480
Louisiana	LA	22	1,121	1,869	1,912	576	480	479
Maryland	MD	24	1,384	1,713	1,662	576	480	479
Mississippi	MS	28	654	1,727	2,199	576	480	472
North Carolina	NC	37	2,158	2,171	1,739	720	600	599
Oklahoma	OK	40	1,036	1,781	1,961	576	480	477
South Carolina	SC	45	1,082	1,711	2,270	576	480	477
Tennessee	TN	47	1,475	1,708	2,310	576	480	480
Texas	TX	48	5,191	4,493	3,426	1,440	1,200	1,198
Virginia	VA	51	1,864	2,174	1,412	720	600	597
West Virginia	WV	54	484	1,485	1,801	576	480	457

See notes at end of table.

(continued)

Table 3.1 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 2014-2022 Sample	Unique Segments in 2014-2022 Sample
West								
Alaska	AK	02	165	516	1,212	576	480	351
Arizona	AZ	04	1,508	1,721	1,951	576	480	479
California	CA	06	7,935	5,155	3,570	1,728	1,440	1,440
Colorado	CO	08	1,234	1,760	2,166	576	480	477
Hawaii	HI	15	309	789	1,380	576	480	397
Idaho	ID	16	296	939	2,166	576	480	434
Montana	MT	30	268	814	1,765	576	480	387
Nevada	NV	32	678	1,448	1,718	576	480	466
New Mexico	NM	35	497	1,301	2,000	576	480	459
Oregon	OR	41	825	1,816	2,063	576	480	475
Utah	UT	49	582	1,448	1,944	576	480	468
Washington	WA	53	1,440	2,055	1,816	576	480	480
Wyoming	WY	56	131	409	1,305	576	480	332

FIPS = federal information processing standards.

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 3.1 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 2014 through 2017 NSDUHs, an additional 28 segments were selected to serve as replacements when segment DUs are depleted, to support any supplemental studies embedded within NSDUH, and to extend the sample to the next decennial census, if desired. As noted previously, these 28 segments constitute the "reserve" sample and are being used to support the 2018 through 2022 NSDUHs.

3.4 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 five reserve subsamples are being used to field the 2018 through 2022 surveys.

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 they were selected.

Using the survey year and quarter assignments, a segment identification number (SEGID) then was assigned. Table 3.2 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 2020 main survey corresponds to segment suffixes G1 through G4 and H1 through H4.

3.5 Impact of COVID-19 on Selection of First-, Second-, and Third-Stage Sampling Units

COVID-19 had no impact on stratification or the first three stages of selection for the 2014 through 2022 coordinated sample. The area samples were selected simultaneously in advance of the 2014 to 2022 NSDUHs.

Table 3.2 Segment Identification Number Suffixes and Quarter Assignment

Segment Suffix	2014 NSDUH	2015 NSDUH	2016 NSDUH	2017 NSDUH	2018 NSDUH	2019 NSDUH	2020 NSDUH	2021 NSDUH	2022 NSDUH	Variance PSU
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)	x (Q1)					1
E2				x (Q2)	x (Q2)					1
E3				x (Q3)	x (Q3)					1
E4				x (Q4)	x (Q4)					1
F1					x (Q1)	x (Q1)				2
F2					x (Q2)	x (Q2)				2
F3					x (Q3)	x (Q3)				2
F4					x (Q4)	x (Q4)				2
G1						x (Q1)	x (Q1)			1
G2						x (Q2)	x (Q2)			1
G3						x (Q3)	x (Q3)			1
G4						x (Q4)	x (Q4)			1

See notes at end of table.

(continued)

Table 3.2 Segment Identification Number Suffixes and Quarter Assignment (continued)

Segment Suffix	2014 NSDUH	2015 NSDUH	2016 NSDUH	2017 NSDUH	2018 NSDUH	2019 NSDUH	2020 NSDUH	2021 NSDUH	2022 NSDUH	Variance PSU
H1							x (Q1)	x (Q1)		2
H2							x (Q2)	x (Q2)		2
H3							x (Q3)	x (Q3)		2
H4							x (Q4)	x (Q4)		2
I1								x (Q1)	x (Q1)	1
I2								x (Q2)	x (Q2)	1
I3								x (Q3)	x (Q3)	1
I4								x (Q4)	x (Q4)	1
J1									x (Q1)	2
J2									x (Q2)	2
J3									x (Q3)	2
J4									x (Q4)	2

PSU = primary sampling unit.
Note: The segment suffix is defined as the last two digits of the segment identification number (SEGID).

4. Selection of Dwelling Units and Persons for the 2020 NSDUH

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 2020 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 2022 NSDUHs, SAMHSA specified the original state and age group sample sizes (n_m). In 2020, a supplemental sample of 1,500 interviews distributed proportionally to states and age groups, n_s, was added in quarters 1 and 2 (approximately 750 interviews were added to each quarterly sample's target). For each state h (51) and age group a (5), the computation of the 255 expanded sample sizes, $m_{ha} = n_m + n_s$, is described in further detail in Section 4.2.
2. Using the m_{ha} from Step 1, the next step was to determine the optimal (i.e., minimum) 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.[28] This step is described in further detail in Section 4.3.
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.[29] The details of this final step are provided in Section 4.4. 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 4.5.

[28] Because of the COVID-19 pandemic, data collection was suspended at the end of the first calendar quarter and did not resume until the fourth calendar quarter. However, 2020 NSDUH DU samples were still selected on a quarterly basis. At the time, it was not clear when data collection would resume.

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

4.1 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.

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 4.2. 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 population in state h and age group a. The 2020 population is estimated using the 2010 census data adjusted to the 2013 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.[30] Then the 2020 population (y^*_{ha}) is computed using the original formula and this time allowing x to be 10. Finally, the 2020 population is adjusted by the ratio of estimated eligible listed DUs to the Claritas DU counts (U_{hj}). This adjustment factor considers the number of added DUs expected to be found on the premises of sampled DUs (1.004) and the probability of a DU being eligible (ε_h), both determined via historical data. The coefficient adjustment of 1.004 is estimated using 2018 NSDUH data and is the proportion of all screened DUs (includes added DUs) over the original total of selected DUs (excluding added DUs). So, $y_{ha} = \{[1.004 * \varepsilon_h * L_{hj} * (1/I_{hj})/U_{hj}]\} * y^*_{ha}$, 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 = 1-5].

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

[30] 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.

the selected segments, these probabilities are adjusted by a factor of 4 so that weights will add to the yearly totals.

I_{hj} = Subsegmentation inflation factor. For segments too large to count and to list efficiently in both time and cost, field listing personnel may request that a portion of the segment be randomly sampled. First, they perform a quick count (best guess: $L_1^*{}_{hj}$) of the entire segment. The sampling staff then subdivides the segment into roughly equal-sized subdivisions or subsegments (using a best guess estimate of the number of DUs in each subsegment based on counts obtained from the field lister: $G_{1\,hj}^*$) and selects one for regular counting and listing. 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:

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

= ($G_{1\,hj}^* / L_{1\,hj}^*$)*($G_{2\,hj}^* / L_{2\,hj}^*$), 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 2018 NSDUH quarter 4 and 2019 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 2018 NSDUH and quarters 1 through 3 of the 2019 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 2018 NSDUH quarter 4 and 2019 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 2018 NSDUH, the adjustment was computed by taking the average maximum-of-two rule adjustment within each state.

4.2 Determining Person Sample Sizes, by State and Age Group

4.2.1 Original Sample

The first step in the design of the fifth stage of selection was to determine the 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 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 2022 design. Compared with the sample allocation in prior years, the 2014 through 2022 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 at the state and substate levels. Furthermore, the 2014 through 2022 design increases the 26 or older sample size to more accurately estimate drug use and related mental health measures among this age group.

The original requirement of 67,507 completed interviews for the 2020 NSDUH was derived from the following objectives:

- 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.

4.2.2 Supplemental Sample

In the first two quarters of the 2020 NSDUH, the original sample was supplemented with approximately 1,500 interviews (see Exhibit 1).[31] This supplemental sample was allocated proportionally to states and age groups. Although removal of the Clinical Validation Study (CVS) sample from the expanded 2020 NSDUH sample was not planned, the supplemental sample was included to maintain the original sample sizes by state and age group if the CVS sample were removed (e.g., if CVS methods were found to significantly affect survey responses). Based on 2018 NSDUH data, Table 4.1 displays the supplemental sample distribution by age group, and Table 4.2 shows the expected supplemental sample sizes by state.

[31] The supplemental sample was selected from the same segments as the original sample.

Table 4.1 2020 NSDUH 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	233,248	100,746	17,015	17,066	10,344	13,836	10,746	69,007
Northeast										
Connecticut	12	96	3,284	1,401	240	240	144	192	144	960
Maine	12	96	3,284	1,401	240	240	144	192	144	960
Massachusetts	12	96	3,284	1,401	240	240	144	192	144	960
New Hampshire	12	96	3,284	1,401	240	240	144	192	144	960
New Jersey	15	120	5,131	2,188	375	375	225	300	225	1,500
New York	30	240	11,289	4,815	825	825	495	660	495	3,300
Pennsylvania	24	192	8,210	3,501	600	600	360	480	360	2,400
Rhode Island	12	96	3,284	1,401	240	240	144	192	144	960
Vermont	12	96	3,284	1,401	240	240	144	192	144	960
Midwest										
Illinois	24	192	8,210	3,501	600	600	360	480	360	2,400
Indiana	12	96	3,284	1,401	240	240	144	192	144	960
Iowa	12	96	3,284	1,401	240	240	144	192	144	960
Kansas	12	96	3,284	1,401	240	240	144	192	144	960
Michigan	24	192	8,210	3,501	600	600	360	480	360	2,400
Minnesota	12	96	3,284	1,401	240	240	144	192	144	960
Missouri	12	96	3,284	1,401	240	240	144	192	144	960
Nebraska	12	96	3,284	1,401	240	240	144	192	144	960
North Dakota	12	96	3,284	1,401	240	240	144	192	144	960
Ohio	24	192	8,210	3,501	600	600	360	480	360	2,400
South Dakota	12	96	3,284	1,401	240	240	144	192	144	960
Wisconsin	12	96	3,284	1,401	240	240	144	192	144	960
South										
Alabama	12	96	3,284	1,401	240	240	144	192	144	960
Arkansas	12	96	3,284	1,401	240	240	144	192	144	960
Delaware	12	96	3,284	1,401	240	240	144	192	144	960
District of Columbia	12	96	3,284	1,401	240	240	144	192	144	960
Florida	30	240	11,289	4,815	825	825	495	660	495	3,300
Georgia	15	120	5,131	2,188	375	375	225	300	225	1,500
Kentucky	12	96	3,284	1,401	240	240	144	192	144	960
Louisiana	12	96	3,284	1,401	240	240	144	192	144	960
Maryland	12	96	3,284	1,401	240	240	144	192	144	960
Mississippi	12	96	3,284	1,401	240	240	144	192	144	960
North Carolina	15	120	5,131	2,188	375	375	225	300	225	1,500
Oklahoma	12	96	3,284	1,401	240	240	144	192	144	960
South Carolina	12	96	3,284	1,401	240	240	144	192	144	960
Tennessee	12	96	3,284	1,401	240	240	144	192	144	960
Texas	30	240	11,289	4,815	825	825	495	660	495	3,300
Virginia	15	120	5,131	2,188	375	375	225	300	225	1,500
West Virginia	12	96	3,284	1,401	240	240	144	192	144	960

(continued)

Table 4.1 2020 NSDUH 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
West										
Alaska	12	96	3,284	1,401	240	240	144	192	144	960
Arizona	12	96	3,284	1,401	240	240	144	192	144	960
California	36	288	15,599	6,653	1,140	1,140	684	912	684	4,560
Colorado	12	96	3,284	1,401	240	240	144	192	144	960
Hawaii	12	96	3,308	1,411	242	242	145	193	145	967
Idaho	12	96	3,284	1,401	240	240	144	192	144	960
Montana	12	96	3,284	1,401	240	240	144	192	144	960
Nevada	12	96	3,284	1,401	240	240	144	192	144	960
New Mexico	12	96	3,284	1,401	240	240	144	192	144	960
Oregon	12	96	3,284	1,401	240	240	144	192	144	960
Utah	12	96	3,284	1,401	240	240	144	192	144	960
Washington	12	96	3,284	1,401	240	240	144	192	144	960
Wyoming	12	96	3,284	1,401	240	240	144	192	144	960
Supplemental Sample (National)	N/A	N/A	2,321	2,257	138	189	218	335	620	1,500

N/A = not applicable.

Table 4.2 2020 NSDUH Sample Sizes, by State

State	Original Sample (n_m)	Supplemental Sample (n_s)	Expanded Sample (m_h)
Total United States	67,507	1,500	69,007
Alabama	960	22	982
Alaska	960	3	963
Arizona	960	32	992
Arkansas	960	14	974
California	4,560	182	4,742
Colorado	960	26	986
Connecticut	960	17	977
Delaware	960	4	964
District of Columbia	960	3	963
Florida	3,300	99	3,399
Georgia	1,500	47	1,547
Hawaii	967	6	973
Idaho	960	8	968
Illinois	2,400	59	2,459
Indiana	960	31	991
Iowa	960	14	974
Kansas	960	13	973
Kentucky	960	20	980
Louisiana	960	21	981
Maine	960	6	966

(continued)

Table 4.2 2020 NSDUH Sample Sizes, by State (continued)

State	Original Sample (n_m)	Supplemental Sample (n_s)	Expanded Sample (m_h)
Maryland	960	28	988
Massachusetts	960	33	993
Minnesota	960	26	986
Mississippi	960	14	974
Missouri	960	28	988
Montana	960	5	965
Nebraska	960	9	969
Nevada	960	14	974
New Hampshire	960	6	966
New Jersey	1,500	42	1,542
New Mexico	960	10	970
New York	3,300	93	3,393
North Carolina	1,500	47	1,547
North Dakota	960	3	963
Ohio	2,400	54	2,454
Oklahoma	960	18	978
Oregon	960	19	979
Pennsylvania	2,400	60	2,460
Rhode Island	960	5	965
South Carolina	960	23	983
South Dakota	960	4	964
Tennessee	960	31	991
Texas	3,300	126	3,426
Utah	960	13	973
Vermont	960	3	963
Virginia	1,500	39	1,539
Washington	960	34	994
West Virginia	960	9	969
Wisconsin	960	27	987
Wyoming	960	3	963

4.2.3 Expanded Sample

The expanded sample included both the original sample of 67,507 and the supplemental sample of 1,500. Each calendar quarter, the state and age group sample sizes (m_{ha}) for the expanded sample were computed as follows:

$$m_{ha} = n_m + n_s ,$$

where n_m was the original state and age group sample target and n_s was the supplemental sample state and age group target.[32] Then the expanded sample was selected in the usual manner based on m_{ha} (see Sections 4.3 and 4.4). Selecting a combined sample rather than two separate samples simplified the process because the probabilities of selection (and design weights) are the same

[32] In quarters 3 and 4, n_s was equal to zero.

for all DUs in a segment. If two separate samples were selected, the design weights would need to be adjusted to combine the two samples.

Using the expanded sample state and age group sample requirements and 2018 NSDUH data, estimates and relative standard errors for 11 outcome measures and several domains of interest were modeled and are presented in Table 4.3 (Center for Behavioral Health Statistics and Quality [CBHSQ], 2019b). 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 (BNGDRKMON)*. 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 (PNRNMMON = 1).*
6. *Past year alcohol use disorder (ABODALC).* Dependence or abuse of alcohol during the past 12 months.
7. *Past year illicit drug use disorder (UDPYILL).* Dependence or abuse of illicit drugs during the past 12 months.
8. *Past year substance use disorder (UDPYILAL).* Dependence or abuse of illicit drugs or alcohol during the past 12 months.
9. *Past year specialty substance use treatment (TXYRSPILAL).* 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_U).* 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 4.3 Relative Standard Errors and Sample Sizes for Key Outcome Measures, by Demographic Domain

Data File Variable Name	Measure	Domain	2018 Prevalence	Projected RSE (2020)	Expected Sample Size (2020)
ALCMON	Past Month Alcohol Use	12+	0.5108	0.0072	69,007
ALCMON	Past Month Alcohol Use	12-20	0.1883	0.0223	23,383
ALCMON	Past Month Alcohol Use	50+	0.5017	0.0135	10,746
ALCMON	Past Month Alcohol Use	API, 12+	0.3905	0.0394	4,103
ALCMON	Past Month Alcohol Use	AIAN, 12+	0.3590	0.1007	777
ALCMON	Past Month Alcohol Use	Pregnant, 12-44	0.0990	0.1470	770
BNGDRKMON	Past Month Binge Alcohol Use	18-25	0.3489	0.0147	17,066
BNGDRKMON	Past Month Binge Alcohol Use	12+	0.2450	0.0105	69,007
MRJMON	Past Month Marijuana Use	12+	0.1011	0.0195	69,007
MRJMON	Past Month Marijuana Use	12-17	0.0666	0.0379	17,015
MRJMON	Past Month Marijuana Use	18-25	0.2212	0.0212	17,066
MRJMON	Past Month Marijuana Use	50+	0.0489	0.0630	10,746
MRJMON	Past Month Marijuana Use	API, 12+	0.0570	0.1294	4,103
MRJMON	Past Month Marijuana Use	AIAN, 12+	0.1458	0.1656	777
MRJMON	Past Month Marijuana Use	Pregnant, 12-44	0.0473	0.1754	770
CIGMON	Past Month Cigarette Use	12-17	0.0270	0.0584	17,015
CIGMON	Past Month Cigarette Use	12+	0.1715	0.0144	69,007
PNRNMMON	Past Month Pain Reliever Misuse	18-25	0.0139	0.0770	17,066
PNRNMMON	Past Month Pain Reliever Misuse	12+	0.0104	0.0555	69,007
ABODALC	Past Year Alcohol Use Disorder	12+	0.0541	0.0238	69,007
UDPYILL	Past Year Illicit Drug Use Disorder	12+	0.0296	0.0329	69,007
UDPYILAL	Past Year Substance Use Disorder	50+	0.0394	0.0612	10,746
TXYRSPILAL	Past Year Specialty Substance Use Treatment	12+	0.0086	0.0627	69,007
SMIYR_U	Past Year SMI	18+	0.0457	0.0264	51,992
AMDEYR	Past Year MDE	18+	0.0720	0.0214	51,992

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 2020 state and age sample allocations in a variance component model. All model components were updated using 2018 NSDUH data.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2018.

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, the original sample size of 67,507 respondents was necessary to meet all sample size requirements while ensuring sufficient precision for key outcome measures, and the supplement sample size of 1,500 respondents was included to support the CVS. The large expanded 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 2020 NSDUH was designed to oversample the younger age groups as shown in Table 4.1. 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.

4.3 Fourth-Stage Sample Determination for Each Segment

Once the desired respondent sample size for each state and age group (m^*_{ha}) was established, the next step was to determine the minimal number of DUs to select for each segment to meet the targeted sample sizes. At the planning stage, it was estimated that roughly 233,248 DUs would need to be selected to obtain a sample of 69,007 responding persons distributed by state and age group as shown in Table 4.1 (CBHSQ, 2019b).[33] The actual fourth-stage 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, and the description is specific to quarter 1. Any modifications or corrections were made in subsequent quarters and are explained in detail in Section 4.8. For example, adjustments were made to the samples from quarters 3 and 4 to compensate for the suspension of data collection due to the COVID-19 pandemic.

4.3.1 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 Maryland and Nevada have fewer map pages on average, while segments in states like North Dakota and Pennsylvania 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 4.4. 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 subsegmented in the

[33] The sample was simulated using 2018 NSDUH data and the expected state and age group sample distribution in Table 4.1. After accounting for eligibility, nonresponse, and the fifth-stage sample selection procedures, the results suggested that roughly 233,248 sample DUs would yield 100,746 selected persons and a sample of 69,007 responding persons.

2020 NSDUH sample are summarized in Table 4.5. For more information on the subsegmenting procedures, see Appendix D.

Table 4.4 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	36,146	6.0
Alabama	96	582	6.1
Alaska	96	449	4.7
Arizona	96	559	5.8
Arkansas	96	655	6.8
California	288	1,554	5.4
Colorado	96	487	5.1
Connecticut	96	373	3.9
Delaware	96	476	5.0
District of Columbia	96	299	3.1
Florida	240	1,469	6.1
Georgia	120	752	6.3
Hawaii	96	338	3.5
Idaho	96	635	6.6
Illinois	192	1,415	7.4
Indiana	96	584	6.1
Iowa	96	703	7.3
Kansas	96	682	7.1
Kentucky	96	493	5.1
Louisiana	96	488	5.1
Maine	96	519	5.4
Maryland	96	391	4.1
Massachusetts	96	475	4.9
Michigan	192	1,264	6.6
Minnesota	96	551	5.7
Mississippi	96	680	7.1
Missouri	96	550	5.7
Montana	96	714	7.4
Nebraska	96	754	7.9
Nevada	96	389	4.1
New Hampshire	96	518	5.4
New Jersey	120	776	6.5
New Mexico	96	546	5.7
New York	240	1,244	5.2
North Carolina	120	723	6.0
North Dakota	96	974	10.1
Ohio	192	1,250	6.5

(continued)

Table 4.4 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
Oklahoma	96	669	7.0
Oregon	96	507	5.3
Pennsylvania	192	1,697	8.8
Rhode Island	96	493	5.1
South Carolina	96	533	5.6
South Dakota	96	730	7.6
Tennessee	96	525	5.5
Texas	240	1,742	7.3
Utah	96	495	5.2
Vermont	96	542	5.6
Virginia	120	733	6.1
Washington	96	455	4.7
West Virginia	96	553	5.8
Wisconsin	96	494	5.1
Wyoming	96	667	6.9

Table 4.5 Segment and Dwelling Unit Summary

State	Total Segments	Total Subsegmented Segments	Second-Level Subsegmented Segments	Listed Dwelling Units
Total Population	6,000	1,529	27	1,509,356
Alabama	96	10	0	22,976
Alaska	96	11	0	23,331
Arizona	96	21	0	23,511
Arkansas	96	6	0	21,235
California	288	133	2	78,554
Colorado	96	10	0	21,799
Connecticut	96	9	0	23,042
Delaware	96	14	0	22,884
District of Columbia	96	18	0	29,246
Florida	240	131	5	71,534
Georgia	120	62	0	33,031
Hawaii	96	18	1	24,510
Idaho	96	8	0	20,278
Illinois	192	83	2	52,132
Indiana	96	7	0	23,523
Iowa	96	9	0	21,605
Kansas	96	5	0	20,603
Kentucky	96	4	0	24,219
Louisiana	96	7	1	22,450
Maine	96	9	0	21,994
Maryland	96	13	0	25,268
Massachusetts	96	7	0	23,401
Michigan	192	79	0	55,863
Minnesota	96	13	0	20,642
Mississippi	96	8	0	21,401
Missouri	96	6	0	22,224
Montana	96	13	0	20,176
Nebraska	96	7	0	22,376
Nevada	96	22	1	24,653
New Hampshire	96	10	0	21,303
New Jersey	120	60	1	33,803
New Mexico	96	6	0	21,663
New York	240	132	1	67,759
North Carolina	120	60	0	33,066
North Dakota	96	11	0	20,421
Ohio	192	90	0	54,485
Oklahoma	96	7	0	22,992
Oregon	96	13	1	22,179
Pennsylvania	192	94	4	54,183
Rhode Island	96	8	1	22,022

(continued)

Table 4.5 Segment and Dwelling Unit Summary (continued)

State	Total Segments	Total Subsegmented Segments	Second-Level Subsegmented Segments	Listed Dwelling Units
South Carolina	96	17	0	22,585
South Dakota	96	14	0	19,554
Tennessee	96	7	0	22,515
Texas	240	129	5	67,822
Utah	96	9	0	20,801
Vermont	96	15	0	23,949
Virginia	120	68	2	34,489
Washington	96	13	0	22,017
West Virginia	96	3	0	22,228
Wisconsin	96	8	0	21,831
Wyoming	96	12	0	19,228

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,[34] 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 2020 counting and listing general manual (RTI International, 2019).

4.3.2 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, formulated as $\left[f_{ha} / \left(\phi_h * \lambda_{ha} * \delta_{ha} \right) \right] = F_h$ and known as the driving age group (see Section 2.3), is set to the largest allowable selection probability (S_{hja}) of 0.99. D_{hj} then is computed as

[34] 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.

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

4.4 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_{hj} and S_{hja} are set to 0.

4.5 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 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 time frame. These concerns prompted adjustments to the D_{hj} sample size:

1. Number of selected DUs for screening: < 120 or < ½ L_{hj}. Adjustments were made by adjusting the D_{hj} counts to equal the minimum of 120 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 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.

4.6 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 4.3.1, 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 4.2), 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 5.1, 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 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 as well as the impact of the COVID-19 pandemic on sample releases will be provided in the forthcoming 2020 NSDUH sample experience report.

4.7 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 that was screened in person.[35] 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 E.

In addition to checking for missed DUs at each sampled DU, interviewers were instructed to call their supervisors if they noticed large differences between 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. They are described in more detail in Appendix E. 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 2020 NSDUH sample experiences.

4.8 Quarter-by-Quarter Deviations

This section describes corrections and/or modifications that were implemented in the quarterly sampling procedures after quarter 1 of 2020. (Quarter 1 followed the methods and procedures described above.) In what follows, "design" refers to deviations from the original proposed plan of design, and "procedural" refers to changes made in the calculation methodologies. Finally, "dwelling unit selection" refers to changes that occurred after sample size derivations, specifically revisions implemented during fielding of the sample (i.e., sample partitioning as described in Section 4.6).

[35] The check for missed DUs was not implemented for DUs screened via web.

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 2019 NSDUH were used to compute average yields, DU eligibility, screening response, and interviewer response rates.

After the initial quarter 2 sample was selected, the decision was made to select and release additional quarter 2 sample to partially compensate for the negative impact of the COVID-19 pandemic on data collection and response rates. All remaining DUs (i.e., those not previously selected) in the quarter 2 panel "G" segments (i.e., those retiring from use after 2020) were selected and released to web data collection during the multimode data collection period. The quarter 2 sample weights will be adjusted to reflect the selection of the additional sample.

Dwelling Unit Selection: The initial 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).

To allow for greater flexibility, the additional quarter 2 sample was partitioned into the following release percentages:

Release 1: 33 percent of additional sample;
Release 2: 33 percent of additional sample;
Release 3: 17 percent of additional sample;
Release 4: 8 percent of additional sample;
Release 5: 4 percent of additional sample; and
Release 6: 4 percent of additional sample.

Quarter 3

Design:	The quarter 3 sample sizes were adjusted to compensate for the suspension of data collection due to COVID-19. Using the completed cases from quarter 1, each state's midyear shortfall was computed. This shortfall then was added to the half-year target and divided by 2 to obtain the new quarter 3 and 4 sample targets. Finally, the adjusted quarter 3 sample sizes were multiplied by a factor of 1.02,[36] and an additional 20 percent sample was included.
Procedural:	Data from quarters 2 through 4 of the 2019 NSDUH and quarter 1 of the 2020 NSDUH were used to compute state-specific average yields, DU eligibility, screening response, and interviewer response rates. To support the larger quarter 3 sample size, SAMHSA approved increasing the maximum number of dwelling units that could be selected from a segment from 120 to 150. Half of the listed units in each segment still were reserved for the following year's sample. After the initial quarter 3 sample was selected, the decision was made to select and release additional quarter 3 sample to partially compensate for the negative impact of the pandemic on data collection and response rates. All remaining DUs (i.e., those not previously selected) in the quarter 3 panel "G" segments (i.e., those retiring from use after 2020) were selected and released to web data collection during the multimode data collection period. The quarter 3 sample weights will be adjusted to reflect the selection of the additional sample.

[36] The quarterly state by age group sample sizes are allocated to quarters 3 and 4 in ratios of 102 and 98 percent, respectively.

Dwelling Unit Selection: To allow for greater flexibility, the initial quarter 3 D_{hj} sample and the additional quarter 3 sample were partitioned into the following release percentages:

Release 1: 33 percent of entire sample (40/120, main sample + 20 percent reserve) or additional sample;
Release 2: 33 percent of entire sample (40/120, main sample + 20 percent reserve) or additional sample;
Release 3: 17 percent of entire sample (20/120, main sample + 20 percent reserve) or additional sample;
Release 4: 8 percent of entire sample (10/120, main sample + 20 percent reserve) or additional sample;
Release 5: 4 percent of entire sample (5/120, main sample + 20 percent reserve) or additional sample; and
Release 6: 4 percent of entire sample (5/120, main sample + 20 percent reserve) or additional sample.

Quarter 4

Design: Unlike the previous quarters, the quarter 4 sample was designed for web data collection to prepare for the worst-case scenario: in-person data collection would not resume in 2020. Web data collection was expected to result in a much lower overall response rate than in-person data collection; therefore, a much larger sample of DUs would be required. Because the NSDUH segments were not designed to support as large of a DU sample as would be required to make up the annual deficit through web data collection, the "maximum" number of DUs was selected in most segments.

Procedural: The coefficient adjustment of 1.004 was removed from the calculation of the estimated number of persons in the population in state h and age group a, y_{ha}, because no added DUs were expected to be found on the premises of sampled DUs through web data collection.

Data from quarters 2 through 4 of the 2019 NSDUH and quarter 1 of the 2020 NSDUH were used to compute state-specific average yields, DU eligibility, screening response, and interviewer response rates. The state-level response rates were then adjusted so that the overall expected response rate would be 20 percent, the expected web response rate. Finally, a 95 percent adjustment was applied for the estimated percentage of mailable addresses on the DU sample frame.

To maximize the DU sample in the quarter 4 segments, the requirement to select no more than 150 DUs per segment was lifted. However, the number of DUs that could be selected in a segment was still limited to half of the listed DUs.

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

Release 1: 33 percent of entire sample;
Release 2: 33 percent of entire sample;
Release 3: 17 percent of entire sample;
Release 4: 8 percent of entire sample;
Release 5: 4 percent of entire sample; and
Release 6: 4 percent of entire sample.

4.9 Impact of COVID-19 on Dwelling Unit and Person Samples

As mentioned throughout this chapter, COVID-19 greatly impacted the 2020 NSDUH DU and person samples. Although data collection was suspended at the end of quarter 1 and did not resume until quarter 4, the hope was that field data collection could resume after a relatively short period of time. Consequently, the DU samples still were selected on a quarterly basis. However, the ongoing pandemic impacted the way the initial quarter 3 and 4 samples were selected and required additional quarter 2 and 3 samples to be selected to minimize the impact of the pandemic on national estimates.

When determining the number of DUs to select for the initial quarter 3 sample, the target person sample sizes were adjusted to compensate for the suspension of data collection due to COVID-19. Approximately half of the interviews not collected in quarter 2 were added to the quarter 3 sample target with the other half of the deficit expected to be made up in quarter 4. To support the larger sample size, SAMHSA approved increasing the maximum number of DUs that

could be selected from a segment from 120 to 150, subject to half of the listed units in each segment being reserved for the following year's sample.

As in the quarter 3 sample, the target person sample sizes for the initial quarter 4 sample were adjusted to compensate for the suspension of data collection due to COVID-19. However, unlike previous quarters, the initial quarter 4 sample was designed for web data collection. Historical state-level response rates were adjusted so that the overall expected response rate would be 20 percent (the expected web response rate), and these adjusted rates were used in the DU sample calculations. In addition, a 95 percent adjustment was applied for the estimated percentage of mailable addresses on the DU sample frame. As a result of these changes, a much larger sample of DUs would be required. Because the NSDUH segments were not designed to support as large of a DU sample as would be required to make up the annual deficit through web data collection, the "maximum" number of DUs was selected in most segments.

In 2020, in-person data collection was limited to quarter 1, a small number of counties during a 1-week data collection in July 2020, and areas deemed safe according to state- and county-level health metrics in quarter 4. The addition of web data collection in quarter 4 partially compensated for the negative impact of COVID-19 on NSDUH data collection. During this multimode data collection period, all selected dwelling units (SDUs) originally selected for the quarter 2, 3, and 4 samples in areas deemed safe for in-person data collection were released to the field. Quarter 2, 3, and 4 SDUs in the remaining areas (except in segments that were worked by FIs during the 1-week, small-scale data collection) were mailed an invitation to participate via web. In addition, as described in Section 4.8, all remaining DUs (i.e., those not previously selected) in the quarter 2 and 3 panel "G" segments (i.e., those retiring from use after 2020) were selected and released to web data collection to increase the final respondent sample size.

This page intentionally left blank

5. Sample Weighting and Variance Estimation

5.1 Sample Weighting Procedures

At the conclusion of data collection, design weights are constructed reflecting the various stages of sampling. This section describes the computation of design weights and briefly describes the planned procedures for computing the final person-level weights for the 2020 National Survey on Drug Use and Health (NSDUH). Full details of the sample weighting procedures will be described in the report on the person-level sampling weight calibration for the 2020 NSDUH (CBHSQ, in press a). Details on how pair weights are computed will be described in the report on the 2020 NSDUH questionnaire dwelling-unit level and person pair-level sampling weight calibration (Center for Behavioral Health Statistics and Quality [CBHSQ], in press b).

5.1.1 Main Study Design Weights

The calculation of the design weights will be based on the stratified, five-stage design of the study. Specifically, the person-level design 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 quarterly samples being examined (equal to 1 if all four quarters are included); and

(2) subsegmentation inflation: by-product of counting and listing (includes up to two levels of subsegmenting).

Stage 4: Selection of dwelling unit (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.

5.1.2 Main Study Weight Adjustments

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 (Folsom and Singh, 2000). 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.

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. This adjustment is useful for providing more stable control totals for subsequent adjustments and pair weights.
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.
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.
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.
6. *Person-Level Poststratification.* This step is to adjust the final person sample weights to the census population estimates for 2020 derived from the 2010 census.
7. *Extreme Weight Treatment at the Person Level.* This adjustment is similar to adjustment 3, except that the weights reflect the fifth stage of selection.

5.1.3 Main Study Analysis Weights

The final adjusted weight, which is the product of 16 weight components (5 design weight components, 4 design weight adjustments, and 7 model-based weight adjustments via GEM), is the analysis weight used in estimation. Exhibit 3 displays all of the individual weight components.

As mentioned previously, quarter 1 data collection was conducted via in-person mode as in prior NSDUH years. Due to the COVID-19 pandemic, NSDUH data collection was suspended on March 16, 2020. No data were collected in quarter 2, and extremely limited in-person data collection was carried out in some counties in July of quarter 3. In quarter 4, multimode (in-person and web) data collection was used to field samples that had originally been drawn for quarters 2 through 4. Additional quarter 2 and 3 samples were also selected and released to web data collection during quarter 4. Because in-person data collection in quarter 4 was limited to segments where COVID-19 infection rates were below a safety threshold level, the majority of screenings and interviews were conducted in web mode. To reflect the data collection changes, the process of developing person-level analysis weights for the 2020 NSDUH will be adjusted from previous NSDUHs. The similarities and differences are as follows:

- Three sets of person-level analysis weights will be developed for different analytical objectives: analysis weights for quarter 1 data, analysis weights for quarter 4 data, and analysis weights for combined quarter 1 and 4 data.

- Quarter 1 and 4 analysis weights will be developed separately following the same weighting procedure (16 weight components, as described in Exhibit 3).
- A poststratification adjustment (#15 weight component in Exhibit 3) will be performed for combined quarter 1 and 4 data to develop analysis weights for quarter 1 and 4 combined data.

Exhibit 3 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

5.1.4 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 subsegmented segments, check that the subsegmenting adjustment factor is greater than 1 (i.e., the count for the entire segment is greater than the count for the subsegment). This check is also performed for segments that are subsegmented twice.
- Compare the DU eligibility indicator with the completed screener indicator. Make sure all screener-complete DUs are eligible.
- Compare the final screening code with the DU eligibility and completed screener indicators to ensure these variables are defined correctly.
- Check the subsampling rate for added DUs that are subsampled. Review the frequency distribution of the DU subsampling rates to check values and ensure 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 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 all completed cases are selected and all selected cases are eligible.
- Compare the final interview code with the person-level eligibility indicator to make sure this variable is defined correctly.
- Make sure the probability of selection is nonmissing for all selected persons.
- Check the maximum-of-two selected persons' adjustment to make sure the maximum value is 2.
- Check the person-level weight to ensure there are no missing values and the sum is close to the expected value.

5.2 Creation of Variance Strata and Variance Primary Sampling Units for Person-Level Estimates

The nature of the stratified, clustered sample design requires the design structure to be taken into consideration when computing variances of survey estimates. Variance strata and variance primary sampling units (PSUs)[37] were created to capture explicit stratification and to identify clustering. For the 2014 through 2022 NSDUHs, variance strata are defined at the state sampling region (SSR) and quarter level (each SSR appears in a different stratum every quarter). Because census tracts, block groups, and segments are nested within variance PSUs, the variance contributions of all three sampling stages are covered by the variance estimation variables. Also, because one segment is selected per sampled census tract and block group, the selection of

[37] Variance PSUs were previously called variance replicates.

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 strata for each quarter of the 2014 through 2022 NSDUHs, the 750 SSRs were first placed in random order (states were randomly sorted, and SSRs were randomly sorted within states). This list, numbered 1 to 750, defined the quarter 1 variance strata (VESTRQ1) for the 2014 through 2022 NSDUHs. For quarter 2, the variance strata, VESTRQ2, were defined as VESTRQ1 + 150 (or VESTRQ1 + 150 − 750 if VESTRQ1 is ≥ 601).[38] 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 variance stratum 451 in quarter 1 was assigned to variance stratum 601 (= 451 + 150) in quarter 2, variance stratum 1 (= 601 + 150 − 750) in quarter 3, and variance stratum 151 (= 1 + 150) in quarter 4. Finally, to make the values unique from previous years, 40,000 was added to the quarterly variance strata to create the final variance strata (e.g., VESTR = VESTRQ1 + 40,000). The resulting 750 variance strata values are the same for the 2014 through 2022 NSDUHs but have different values from variance strata defined for previous years because the 2014 through 2022 sample was selected independently from previous samples.

The 2014 through 2022 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. By combining SSRs across states to form strata, the risk of disclosure is reduced because an intruder cannot assume that respondents in the same variance stratum reside in the same state.

Similar to the 2005 through 2013 definition of variance strata, the 2014 through 2022 definition also has the effect of increasing the number of (nominal) degrees of freedom (*df*) for state-level estimates while preserving the number of degrees of freedom for national estimates (750).[39] Each of the smallest sample states is in 48 different variance strata (12 SSRs × 4 quarters); therefore, 48 degrees of freedom are available for state estimates in these states (each variance stratum contains two variance PSUs, as described in the next paragraph). At the other extreme, the largest sample state, California, is in 144 variance strata (36 SSRs × 4 quarters) and therefore has 144 degrees of freedom for estimation. As demonstrated in Appendix F, the 2014 through 2022 definition of variance 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 strata).

Two variance PSUs per year were defined within each variance stratum. Each variance PSU consists of four segments, one for each quarter of data collection. The first variance PSU consists of those segments that are "phasing out" or will not be used in the next survey year. The

[38] The interval of 150 between quarterly variance estimation strata for SSRs was arbitrarily chosen. Any number larger than the maximum number of SSRs in a state (36) could have been used to ensure that the resulting variance strata contain SSRs from four different states.

[39] This *df* is a measure of the precision of the variance estimate. The conventional (or "nominal") *df* measure for a variance estimate under a complex sample is the number of variance PSUs minus the number of variance strata. Ideally, the variance of the variance estimator is 2/*df* (the qualifier "nominal" is sometimes used because this ideal may not be realized).

second variance PSU consists of those segments that are "phasing in" or will be fielded again the following year, thus constituting the 50 percent overlap in area segments between survey years.

Table 3.2 (shown earlier) describes the assignment of segments to variance PSUs that are designed to account for any positive covariance among consecutive year change estimates. As shown in the table, the variance PSU values alternate between 1 and 2 for each panel of the 2014 through 2022 design. As a result, when combining data from multiple years, the pooling of the samples within variance PSUs provides increased precision of estimates.

In addition to variance strata and variance PSUs, a sample weight is computed for each final respondent as described in Section 5.1. The use of sample weights in analyses of NSDUH data is necessary to properly represent the target population and to account for disproportionate sampling by age group. All weighted statistical analyses for which variance estimates are needed should use the variance stratum and variance PSU variables to identify nesting. Variance estimates can be computed using a clustered data analysis software package such as SUDAAN® (RTI International, 2012b).[40] 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.[41] Using data from the 2007 and 2008 NSDUHs and examining multiple measures, the average relative change in the standard error after accounting for covariance was about 1 percent.

5.3 Creation of Variance Strata and Variance Primary Sampling Units for Pair Analyses

Alternate versions of the variance strata and variance PSUs 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 strata and variance PSUs described in Section 5.2. Very rarely, however, alternate versions must be made because not all variance strata have at least one responding pair in both variance PSUs. In this scenario, variance strata with only one variance PSU represented by responding pairs were collapsed with a neighboring variance stratum.

The algorithm for selecting the neighboring variance stratum is as follows:

- Identify the person-level variance stratum that precedes the problematic one and the person-level variance stratum that follows the problematic one. The numbering is described in Section 5.2.

[40] See the 2019 NSDUH statistical inference report for additional examples of software that may be used (CBHSQ, 2021).

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

- 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.

5.4 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 quarters.[42] 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 2022 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 2.2), county-level population density, and a rural or urban indicator.[43] Each census block is assigned a rural or urban status based on population density and/or proximity to a census-designated urbanized area or urban cluster. 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 3.3). However, the definition slightly overestimates the urban population.

Similar to the 2005 through 2013 NSDUH samples, the 2014 through 2022 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 4.3.1), only the group of blocks in which a NSDUH respondent resides is known. Beginning in 2014 and continuing through 2020, the specific census block associated with each NSDUH respondent's listing unit was assigned. Using desktop computer mapping software, census block information was recorded by manually comparing electronic segment maps to field listings as is described in Appendix G. 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 quarter 4, preliminary census block assignments were prepared so that block-level covariates could be defined and used in item imputation models. To accommodate the 12-month imputation schedule, all respondent DUs as of December 1, 2020, were identified. Next, the specific census block associated with each of these DUs was manually assigned using the procedures described in Appendix G. Addresses for all pending DUs were geocoded. Finally, block-level covariates were defined using the manually assigned or geocoded census block. If

[42] Data were obtained or derived from the Census 2010 Summary File 1 (see footnote 27) and adjusted using revised population counts from Claritas.

[43] All variables were obtained or derived from the Census 2010 Summary File 1 (see footnote 27).

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 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 1st, 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.

5.5 Impact of COVID-19 on Sample Weighting and Variance Estimation

As described in Section 5.1, the impact of COVID-19 on 2020 NSDUH data collection requires the development of three person-level weights to support various analytic goals: analysis weights for quarter 1 data, analysis weights for quarter 4 data, and analysis weights for combined quarter 1 and 4 data. Quarter 1 and 4 analysis weights will be developed separately using standard weighting procedures for three reasons. First, the change in mode from in-person to primarily web necessitates analyses that can be done separately for quarter 1 and 4. Second, new questionnaire items added in quarter 4 can be analyzed only with quarter 4 data. Third, nonresponse patterns may be quite different between quarter 1 and 4 due to different modes employed. To compute the combined quarter 1 and 4 analysis weight, a poststratification adjustment for the combined quarter 1 and 4 data will be performed.

The COVID-19 pandemic had no impact on the creation of variance strata, variance PSUs, or other sampling-related variables.

This page intentionally left blank

6. Clinical Validation Study Sample Design

6.1 Overview

The Clinical Validation Study (CVS) sample was originally planned for the first 6 months of the 2020 National Survey on Drug Use and Health (NSDUH). However, because quarter 2 data collection was suspended because of COVID-19, the CVS was limited to quarter 1 of the 2020 NSDUH. A probability subsample of 750 respondents was expected to be drawn from the quarter 1 expanded sample. Persons eligible for the CVS sample were aged 12 or older who chose to answer the NSDUH main study interview questions in English and who did not break off before beginning the substance use disorder (SUD) module, including those with no past year use of alcohol or any drug.[44] Within the NSDUH interview, eligible respondents were selected for the CVS sample based on their responses to questions on past year use of cigarettes, alcohol, and marijuana. Eight strata were constructed from all combinations of past year use of cigarettes, alcohol, and marijuana. Using the initials of these three variables, the stratification design is referred to as the cigarette, alcohol, and marijuana (CAM) design. A random sample was selected from each CAM stratum based on predetermined sampling rates. In addition, a state and age group adjustment increased efficiency by selecting a nationally representative sample of respondents aged 12 or older, with no oversampling of any specific populations (such as age group or state) other than the CAM designations.[45]

The CVS sample was administered a new SUD module that used criteria in the *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition (DSM-5) (American Psychiatric Association [APA], 2013). NSDUH respondents selected for the CVS were routed to the new SUD module instead of to the previous SUD module that used criteria in the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV) (APA, 1994). Otherwise, except for the SUD items, the NSDUH interview was virtually identical for both respondents selected for the CVS and those not selected for the CVS.[46] NSDUH respondents selected for the CVS and diverted to the new SUD items were then invited by the interviewer at the end of the NSDUH interview to participate in a follow-up clinical interview by telephone, ideally to be scheduled within 2 weeks of the NSDUH interview.

After passing through the CVS sample selection algorithm, the remaining cases not selected for the CVS were referred to as the non-CVS sample. Based on experience from past follow-up studies, it was expected that approximately 400 of the 750 CVS sample respondents would complete the CVS clinical follow-up, and these 400 cases would be referred to as the CVS

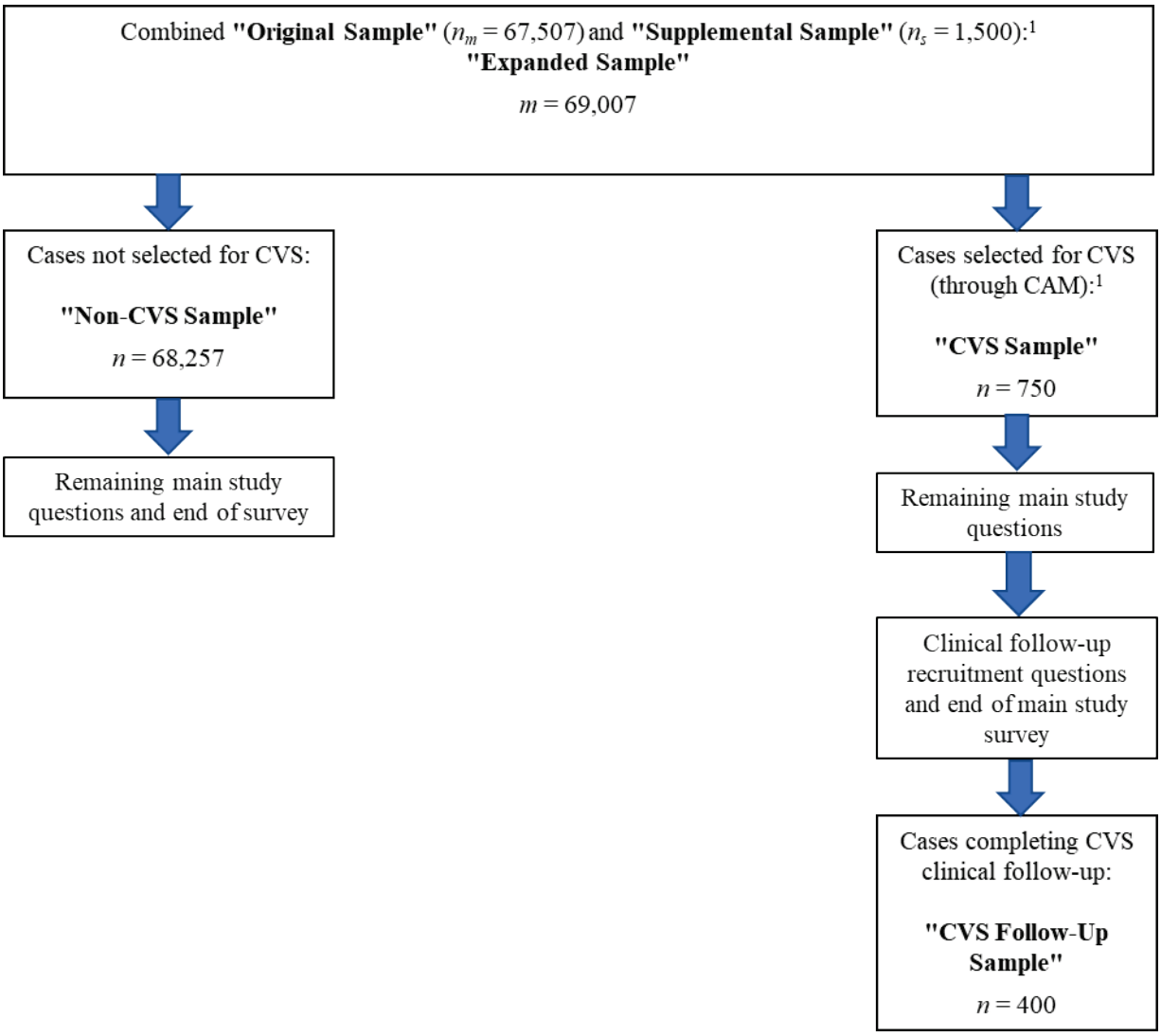
[44] The new SUD module and the clinical follow-up interview were conducted only in English. Therefore, respondents who chose to answer the NSDUH main study questions in Spanish were excluded from the CVS.

[45] The Substance Abuse and Mental Health Services Administration (SAMHSA) determined that the CVS did not need to provide state-level estimates and that it would be sufficient to provide estimates of those aged 12 or older rather than by various age groups.

[46] For CVS respondents, questions assessing DSM-IV symptoms that were not also included in the DSM-5 criteria (e.g., legal troubles) were included at the end of the DSM-5 SUD module. For non-CVS respondents, questions assessing DSM-5 symptoms that were not also included in the DSM-IV criteria (e.g., craving and cannabis withdrawal) were included at the end of the computer-assisted interviewing (CAI) interview.

follow-up sample. A diagram illustrating the structure of the planned CVS sample design is given in Exhibit 4.

Exhibit 4 Structure of the Clinical Validation Study Sample Design



CAM = cigarette, alcohol, and marijuana; CVS = Clinical Validation Study.
[1] The supplemental sample was included in quarters 1 and 2 with approximately half allocated to each quarter's sample. The CVS sample was selected from the quarter 1 expanded sample only.

6.2 CVS Sample Size Determination

The majority of respondents in the CVS follow-up sample have SUD measures from two instruments: (1) NSDUH's self-administered CAI, and (2) clinical interview measures from the *Structured Clinical Interview for DSM-5—Research Version* (SCID-5-RV) (First et al., 2015). The primary analyses of the CVS are agreement analyses between CAI and SCID measures of several SUDs. Therefore, the sample size calculation for this study was based on the precision of Cohen's kappa statistic for the agreement of two estimates. The sample size calculation required

(1) an estimate of the expected kappa, (2) an estimated prevalence of the SUD, and (3) the desired precision of the kappa estimate (as represented by the tightness of the 95 percent confidence interval [CI] surrounding the kappa). Kappa estimates were calculated from a previous analysis that assessed agreement between the CAI and SCID[47] versions of three binary SUDs based on DSM-IV (i.e., overall SUD, illicit drug use disorder [DUD], and alcohol use disorder [AUD]) from the 2008-2012 MHSS. In that nationally representative study, estimated kappa values ranged from 0.48 to 0.54. For the CVS, a kappa value of 0.50 for all disorders was assumed, and a desired lower 95 percent CI level was assumed to be 0.40. All kappa sample size calculations were based on the R program "CIBinary" in the R library "kappaSize" (Rotondi, 2018); for further details about sample size calculations for agreement analyses, see Rotondi and Donner (2012). Using the sample size calculation results, SAMHSA selected a CVS follow-up sample size of 800 respondents with half of the follow-up interviews targeted in each of quarters 1 and 2. As noted previously, the CVS was implemented in quarter 1 only; thus, the expected CVS follow-up sample size was 400.

Assuming an 85.5 percent clinical follow-up agreement rate and a 64.4 percent clinical follow-up completion rate, a CVS sample of 750 was expected to yield at least 400 completed clinical follow-up interviews in quarter 1 (see Table 6.1). Assumed outcome rates for the clinical interviews were based on consideration of the 2008-2012 MHSS outcome rates for clinical interviews and the recent National Mental Health Study (NMHS) Clinical Reappraisal Study (CRS) (Center for Behavioral Health Statistics and Quality [CBHSQ], 2014a, 2014b, 2019c). The average 2008-2012 MHSS combined agreement and completion rate was 66 percent, and the combined agreement and completion rate for the NMHS CRS was 60 percent. The projected combined agreement rate and completion rate in Table 6.1 is 55 percent. To ensure the CVS sample would yield the required number of clinical interviews, this assumption also reflects recent trends in declining response rates. The expected CVS eligibility rate reflects the proportion of interviews that were expected to be completed in English and was computed using 2016 NSDUH data.

Table 6.1 Expected Sample Sizes and Outcome Rates for the Clinical Validation Study (Quarter 1)

Item	Quarter 1 Expanded Sample[1]	Eligible for the CVS	Selected for the CVS	Agreed to Follow Up	Completed Follow-Up Interview
Sample Size	16,922	16,260	750	641	413
Outcome Rate	N/A	0.9609	0.0461	0.8550	0.6440

CVS = Clinical Validation Study; N/A = not applicable.

[1] The sample was allocated to quarters 1 and 2 in proportions of 96 and 104 percent, respectively. Thus, the quarter 1 expanded sample was 96 percent of the quarterly sample (i.e., 0.96 * [16,877 + 750]). The quarter 1 CVS sampling rates were adjusted so that 750 respondents would still be selected.

[47] For the 2008-2012 Mental Health Surveillance Study (MHSS), the following version of the SCID was used: *Structured Clinical Interview for the DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition* (SCID-I/NP) (First et al., 2002).

6.3 CVS Sample Allocation

The CVS sample was drawn from the expanded NSDUH sample by means of a stratified random sample design. The stratified design consisted of eight strata constructed from all combinations of three binary variables measuring the presence or absence of past year cigarette use (CIGPY_R), past year alcohol use (ALCPY_R), and past year marijuana use (MRJPY_R) because these variables are highly correlated with the corresponding SUDs of interest. These three variables are recodes representing raw versions of past year use of cigarettes, alcohol, and marijuana obtained early in the NSDUH interview. Using the initials of the three variables, this stratification design is referred to as the CAM design.

Similar to the 2008-2012 MHSS, Neyman optimal allocation (Lohr, 1999) was applied to the strata (i.e., stratum selection probabilities are proportional to the standard deviation of the measure in question) using 2015 NSDUH data. Let N_h equal the population size in stratum h, W_h equal the population proportion per stratum (i.e., stratum weight for population or $N_h / \sum_{h=1}^{8} N_h$), P_h equal the expected population prevalence of the outcome variable (weighted prevalence estimate). Then the Neyman optimal sampling proportion, f_h, is defined as

$$f_h = W_h * \sqrt{P_h(1-P_h)}.$$

The Neyman f_h are then scaled to sum to 1 to obtain the subsample proportions. Finally, the scaled subsample proportions are applied to the CVS sample size to obtain the stratum sample sizes. Neyman allocation results in increased precision of the key outcome variables (e.g., SUD, AUD, and DUD) compared with other allocations, assuming a fixed sample size and equal costs across strata.

The stratified random sample based on the CAM design minimizes the sample size required for the CVS in two ways: (1) by increasing the yield of DSM-5 SUD cases within the CVS sample (i.e., the closer the yield is to 50 percent, the smaller the required sample size will be for kappa, given all else is equal), and (2) by reducing the design effect within the CVS sample due to the Neyman allocation.

Because the representation of AUD is large in the SUD outcome variable, optimization of the stratified design based on SUD favors AUD over most other disorders associated with illicit drug use. Therefore, the DUD[48] outcome variable was used in the Neyman allocation to increase the number of respondents with disorders associated with illicit drug use in the CVS sample. Table 6.2 displays quarter 1 CVS sample sizes by stratum with Neyman allocation applied to strata based on DUD.

[48] DUD refers to any disorder associated with illicit drug use; that is, it refers to any SUD except AUD.

Table 6.2 Quarter 1 CVS Sample Allocation and Sampling Rates, by CAM Stratum

CAM Stratum	CIGPY_R	ALCPY_R	MRJPY_R	N_h	W_h	P_h	SD_h	Neyman f_h	O_h	CVS Sample Size	Q1 2020 Projected Sample	Unadjusted CVS Sampling Rate
1	0	0	0	78,468,278	0.293	0.4%	0.059	0.017	0.148	111	5,891	0.0188
2	0	0	1	1,616,739	0.006	11.4%	0.317	0.002	0.016	13	189	0.0688
3	0	1	0	111,463,672	0.416	0.4%	0.063	0.026	0.225	168	5,664	0.0297
4	0	1	1	14,418,702	0.054	9.9%	0.298	0.016	0.137	103	1,411	0.0730
5	1	0	0	10,547,272	0.039	3.3%	0.177	0.007	0.060	45	515	0.0874
6	1	0	1	1,803,846	0.007	22.1%	0.415	0.003	0.024	18	138	0.1304
7	1	1	0	31,443,682	0.117	2.4%	0.152	0.018	0.153	115	1,637	0.0703
8	1	1	1	17,932,298	0.067	21.8%	0.413	0.028	0.237	177	1,477	0.1198
Total	N/A	N/A	N/A	267,694,489	1.000	2.9%	N/A	0.117	1.000	750	16,922	N/A

CAM = cigarette, alcohol, and marijuana; CVS = Clinical Validation Study; N/A = not applicable; Q1 = quarter 1.

Notes:

N_h = Population size per stratum (aged 12 or older).

W_h = Population proportion per stratum (i.e., stratum weight for population).

P_h = Expected population prevalence of illicit drug use disorder (weighted prevalence estimates used).

SD_h = Standard deviation of P_h (i.e., $\sqrt{P_h(1-P_h)}$).

Neyman f_h = Neyman optimal allocation to strata (i.e., $f_h = W_h * SD_h$).

O_h = Neyman f_h scaled to sum to 1 (i.e., each f_h is divided by $\sum_{h=1}^{8} W_h * SD_h$ to obtain the optimal stratum weight for subsample; notice how it differs from W_h after application of the Neyman allocation).

CVS sample size = 750 * O_h. The unrounded values of O_h were used in the CVS sample size calculations. Using the rounded values displayed in the table results in slightly different sample sizes.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Surveys on Drug Use and Health (NSDUH), 2015 and 2017.

Using unweighted 2017 NSDUH data, the expected number of completed interviews by stratum were projected for quarter 1 of the 2020 NSDUH. These numbers were then used as the denominators in the unadjusted CVS sampling rates. Table 6.2 includes the quarter 1 projected NSDUH completed interviews and unadjusted CVS sampling rates.

6.4 CVS Sample Selection

Because estimates at the state and age group levels were not required, the impact of unequal sampling by state and age group in the NSDUH design was removed from the CVS probabilities of selection in order to increase efficiency for national estimates. The probability of selection for individual k was computed as follows:

$$p_{CVS-k} = \min(0.5, p_{CAM-k} * F_{h,age}) ,$$

where p_{CAM-k} is the unadjusted CAM stratum sampling rate based on Neyman allocation (see Table 6.2) and $F_{h,age}$ is the state (h) and age group adjustment. To compute $F_{h,age}$, the NSDUH probability of selection was first calculated as the state and age group target sample size divided by its population estimated using 2017 NSDUH data. Then the inverse of the NSDUH probability of selection, or the mean weight for the state and age group, was standardized by dividing by the overall mean weight. These standardized weights were the state and age group adjustment factors, $F_{h,age}$. Setting the maximum value of p_{CVS-k} to 0.5 made it almost certain that the non-CVS sample would be represented in each CAM stratum. Note that because $F_{h,age}$ is a state and age group adjustment, the p_{CVS-k} can vary within strata.

Choosing a NSDUH respondent for the CVS sample during the NSDUH interview limited how the probability proportional to size sample could be drawn. Like the MHSS, a value v_k randomly selected from the unit interval [0, 1) was attached to every NSDUH respondent. Then Poisson sampling was used to select eligible respondents for the CVS sample. Individual k was selected for the CVS sample if v_k is less than or equal to p_{CVS-k}.

6.5 CVS Sample Results

Throughout quarter 1 data collection, the CVS sample and outcome rates were monitored. Although some outcome rates differed from their expected values, the overall quarter 1 sample performed as expected, and no adjustments to the CVS sampling rates were required. As shown in Table 6.3, a total of 699 quarter 1 respondents were selected for the CVS. Of those selected, 539 agreed to participate in the CVS (77.1 percent), and 424 completed the follow-up interview (78.7 percent).

Table 6.3 Achieved Sample Sizes and Outcome Rates for the Clinical Validation Study

Item	Quarter 1 Expanded Sample	Eligible for the CVS	Selected for the CVS	Agreed to Follow Up	Completed Follow-Up Interview
Sample Size	15,628	15,001	699	539	424
Outcome Rate	N/A	0.9599	0.0466	0.7711	0.7866

CVS = Clinical Validation Study; N/A = not applicable.

6.6 CVS Weighting and Variance Estimation

Analysis weights for the CVS sample, the CVS follow-up sample, and the non-CVS sample will be separate from the NSDUH expanded sample analysis weight. To compute the CVS analysis weight, the expanded sample analysis weight (ANALWT) will be multiplied by the CVS design weight (the inverse of the probability of selecting the person for the CVS or $1/p_{CVS\text{-}k}$) and a poststratification adjustment. Similarly, for the CVS follow-up sample, the expanded sample analysis weight will be multiplied by the CVS design weight, adjusted for nonresponse, and poststratified. Finally, the non-CVS sample weight will be computed as the expanded sample analysis weight multiplied by the inverse of the probability of *not* being selected for the CVS and a poststratification adjustment factor. For more detail on CVS weighting, see the 2020 CVS analysis plan (CBHSQ, 2019a).

In addition to analysis weights, CVS-specific variance estimation strata and replicates were created to account for the design when computing variances of estimates. To create the CVS variance strata, groups of 25 adjacent NSDUH variance strata were collapsed. Thus, a total of 30 CVS variance strata were formed. The variance replicates were retained from the main study.

6.7 Impact of COVID-19 on the CVS Sample Design

As noted previously, the CVS was originally planned to be conducted in quarters 1 and 2 of the 2020 NSDUH. Because data collection was suspended at the end of quarter 1, the CVS was implemented in quarter 1 only. Thus, because of COVID-19, the CVS resulted in approximately half of the completed follow-up interviews that were originally planned.

This page intentionally left blank

References

American Association for Public Opinion Research. (2016). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (9th ed.). Retrieved from https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (DSM-IV) (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (DSM-5) (5th ed.). Arlington, VA: Author.

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. (2014a). *2012 National Survey on Drug Use and Health: Methodological resource book: Section 16a, Mental Health Surveillance Study operations report, 2008-2012*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2014b). *2012 National Survey on Drug Use and Health: Methodological resource book: Section 16a, 2012 Mental Health Surveillance Study design and estimation report*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2016). *2015 National Survey on Drug Use and Health methodological resource book: Section 2, sample design report.* Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2019a). *2020 Clinical Validation Study test analysis plan in the National Survey on Drug Use and Health (NSDUH): NSDUH methodological report* (unpublished internal documentation). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2019b). *2020 National Survey on Drug Use and Health: Sample design plan* (unpublished internal documentation). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2019c). *National Mental Health Study-Field Test (NMHS-FT): Final report*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2021). *2019 National Survey on Drug Use and Health methodological resource book: Section 13, statistical inference report*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (in press a). *2020 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). *2020 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.

Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section*, Indianapolis, IN (pp. 598-603). Alexandria, VA: American Statistical Association.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition* (SCID-I/NP). New York, NY: New York State Psychiatric Institute, Biometrics Research.

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). *Structured Clinical Interview for DSM-5—Research Version* (SCID-5 for DSM-5, Research Version; SCID-5-RV). Arlington, VA: American Psychiatric Association.

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.

Lohr, S. L. (1999). *Sampling: Design and analysis*. Belmont, CA: Duxbury Press.

Morton, K., Asman, K., Brown, K., Currivan, D., Peytchev, A., & Vaish, A. (2020). *2020 NSDUH minimum sample size and alternative plan memo* (unpublished internal documentation). Research Triangle Park, NC: RTI International.

Office of Management and Budget. (2009, December 1). *OMB Bulletin No. 10-02: Update of statistical area definitions and guidance on their uses*. Retrieved from https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/bulletins/2010/b10-02.pdf

Rotondi, M. A. (2018, November 26). *Package 'kappaSize' (Version 1.2, sample size estimation functions for studies of interobserver agreement)*. Retrieved from https://cran.r-project.org/web/packages/kappaSize/kappaSize.pdf

Rotondi, M. A., & Donner, A. (2012). A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. *Journal of Clinical Epidemiology*, *65*, 778-784. https://doi.org/10.1016/j.jclinepi.2011.10.019

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. (2019). *2020 National Survey on Drug Use and Health: Counting and listing general manual* (RTI/0215638, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283201700002C). Research Triangle Park, NC: Author.

U.S. Census Bureau. (2001). Appendix A: Geographic definitions. In *Census 2000 Redistricting data (Public Law 94-171) summary file*. Retrieved from https://www.census.gov/programs-surveys/decennial-census/about/rdo/summary-files.2000.html

This page intentionally left blank

List of Contributors

This National Survey on Drug Use and Health methodological report was prepared by the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality, and RTI International. Work by RTI was performed under Contract No. HHSS283201700002C. Rong Cai served as the government project officer and as the contracting officer representative, and David Hunter served as the RTI project director.

Significant contributors to this report at SAMHSA include Qiyuan Pan.

Significant contributors to this report at RTI include Katherine B. Morton, Peilan C. Martin, Jeremy Aldworth, Erica L. Hirsch, and Charlotte B. Looby. Other contributors to and/or reviewers of this report at RTI include (in alphabetical order) James C. Cajka, Bernadette M. Chasteen, Patrick P. Chen, Rachel M. Harter, Andrew Haynes, Phillip S. Kott, Susan K. Myers, and Jesse Pegg.

Also at RTI, Margaret Johnson copyedited the report, and Teresa F. Bass coordinated its web production.

This page intentionally left blank

Appendix A:
Comparison of the 2014 through 2022 NSDUH Sample Design with the 2005 through 2013 NSDUH Sample Design

Similar to the sample design for the 2005 through 2013 National Surveys on Drug Use and Health (NSDUHs), the 2014 through 2022 design is a stratified, multistage area probability sample. Although the 2005 through 2013 NSDUHs used data from the 2000 census and Claritas, dwelling unit (DU) counts and other geographic data were obtained from the 2010 census, Claritas, and the 2006-2010 American Community Survey for the 2014 through 2022 NSDUHs. Table A.1 summarizes the data sources used for the 2014 through 2022 and earlier NSDUH sample designs.

Table A.1 Sources of Geographic Data Used in NSDUH Sample Designs: 1999 through 2022

Data Source	1999-2004	2005-2013	2014-2022
1990 Census, 1999 Claritas projections	States, FI Regions, Tracts, Segments	N/A	N/A
2000 Census, 2007 Claritas projections	N/A	States, SSRs, Tracts, Segments	N/A
2010 Census, 2006-2010 ACS, 2013 Claritas projections	N/A	N/A	States, SSRs, Tracts, Block Groups, Segments

ACS = American Community Survey; FI = field interviewer; N/A = not applicable; SSR = state sampling region.
Sources: U.S. Census Bureau, Decennial Census, 1990, 2000, 2010.
U.S. Census Bureau, American Community Survey, 2006-2010.

Both the 2005 through 2013 and the 2014 through 2022 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 A.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 A.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; IL = Illinois; MI = Michigan; NY = New York; OH = Ohio; PA = Pennsylvania; SSR = state sampling region; TX = Texas.

The 2014 through 2022 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 A.3, the 2014 through 2022 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.

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 A.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 2022 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 2022 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 population. For the 2014 through 2022 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 A.3 Sample Sizes and Projected Respondents, by State and Age Group: 2014 through 2022

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.

Table A.4 Population and Sample Percentages, by State: 2013 and 2014 through 2022

Variable	2010 CNI Population	Percentage of Population	2014-2022 Annual Sample Size	Percentage of 2014-2022 Annual Sample	2013 Sample Size	Percentage of 2013 Sample
Total United States	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%

See notes at end of table. (continued)

Table A.4 Population and Sample Percentages, by State: 2013 and 2014 through 2022 (continued)

Variable	2010 CNI Population	Percentage of Population	2014-2022 Annual Sample Size	Percentage of 2014-2022 Annual Sample	2013 Sample Size	Percentage of 2013 Sample
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%
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%

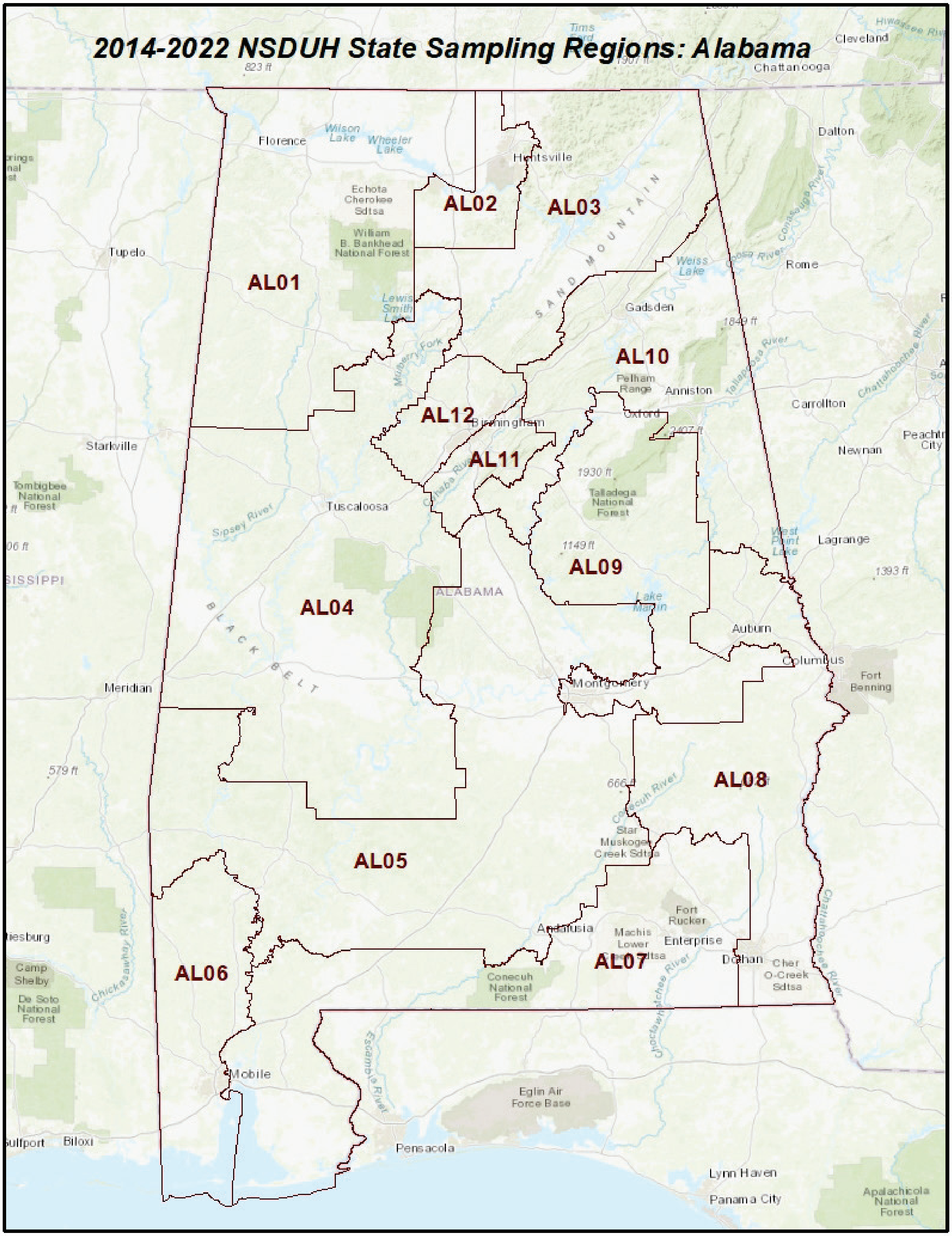
CNI = civilian, noninstitutionalized population.

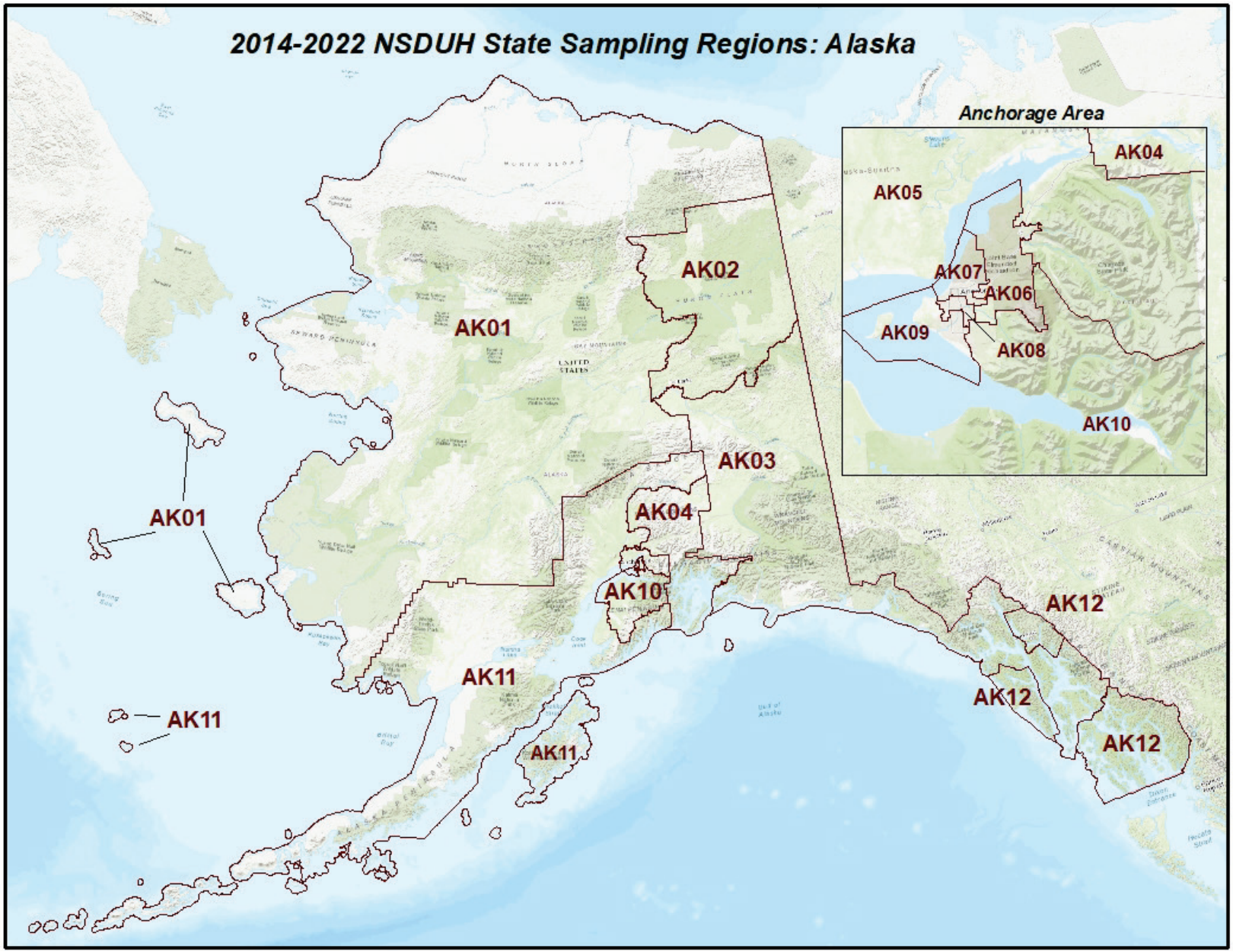
The 2014 through 2022 design includes the selection of census block groups at the second stage of selection. As mentioned in Section 2.2, this additional stage of selection was included to facilitate moving to an address-based sampling (ABS) design, if desired. Compared with geocoding at the census block 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.

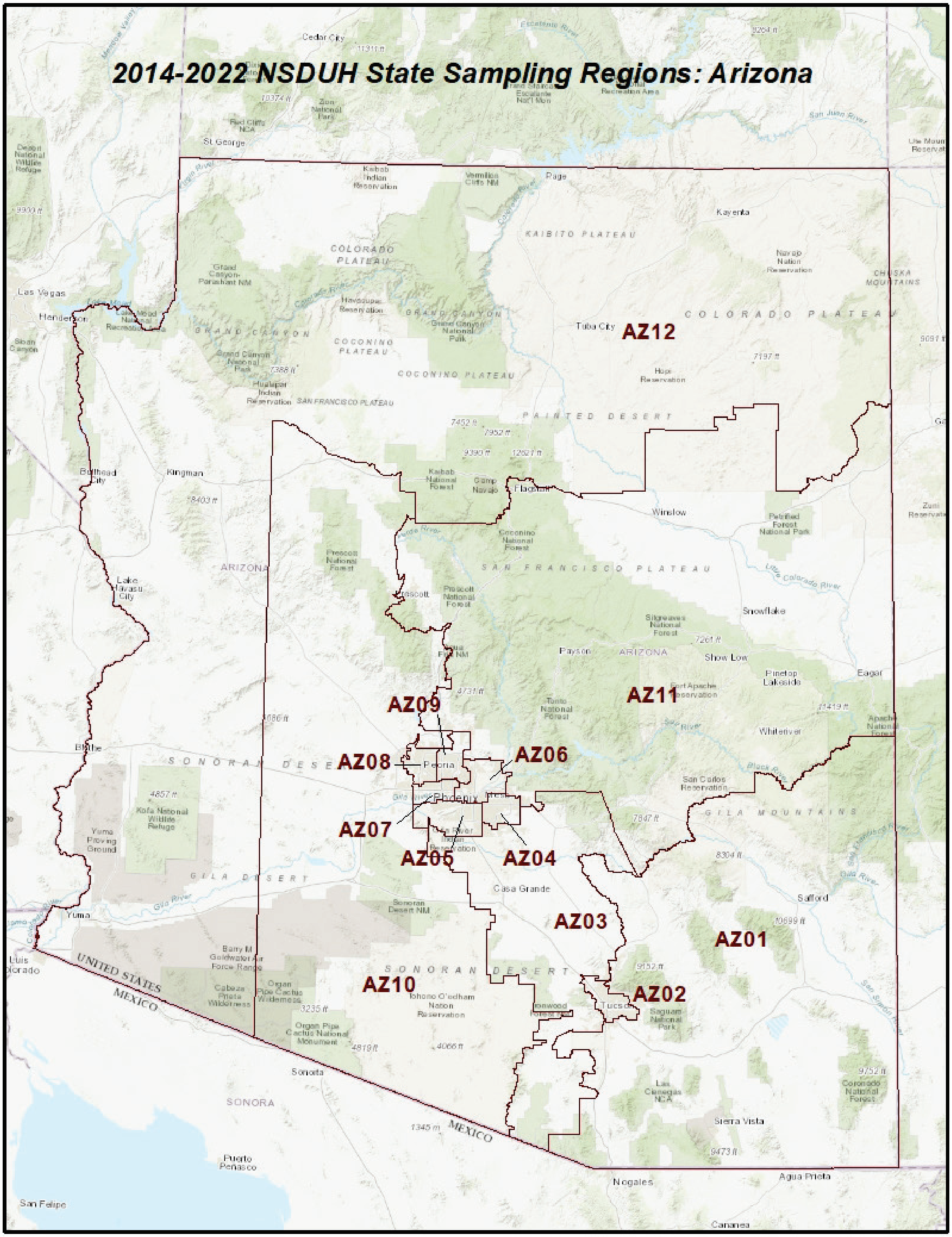
Finally, as mentioned in Section 2.3, the 2014 through 2022 NSDUH fourth-stage sampling frames are supplemented with new DUs on the premises of sampled DUs that were missed during the original counting and listing activities (e.g., garage apartments). However, the half-open interval procedure is no longer being implemented.

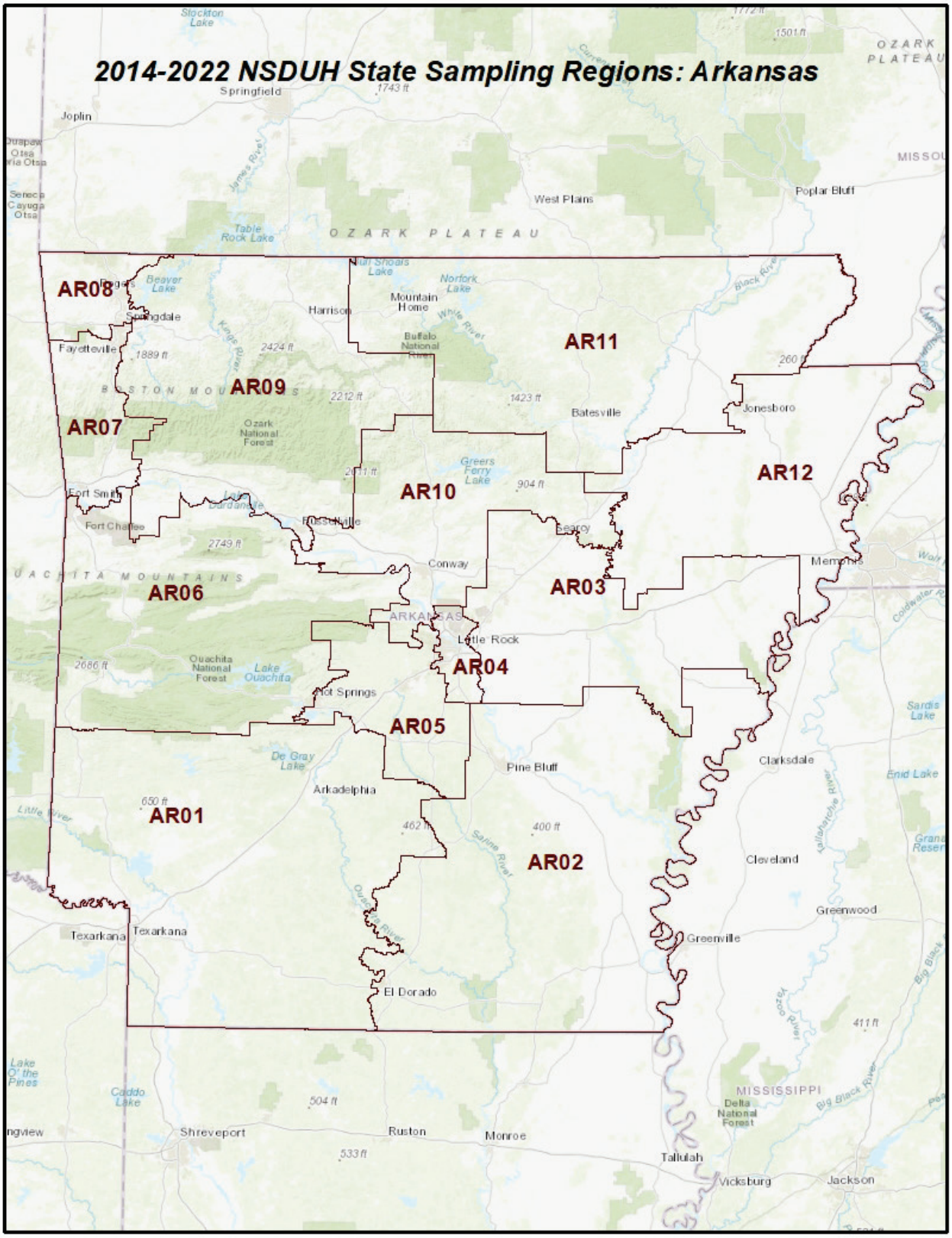
This page intentionally left blank

Appendix B:
2014 through 2022 NSDUH State Sampling Regions

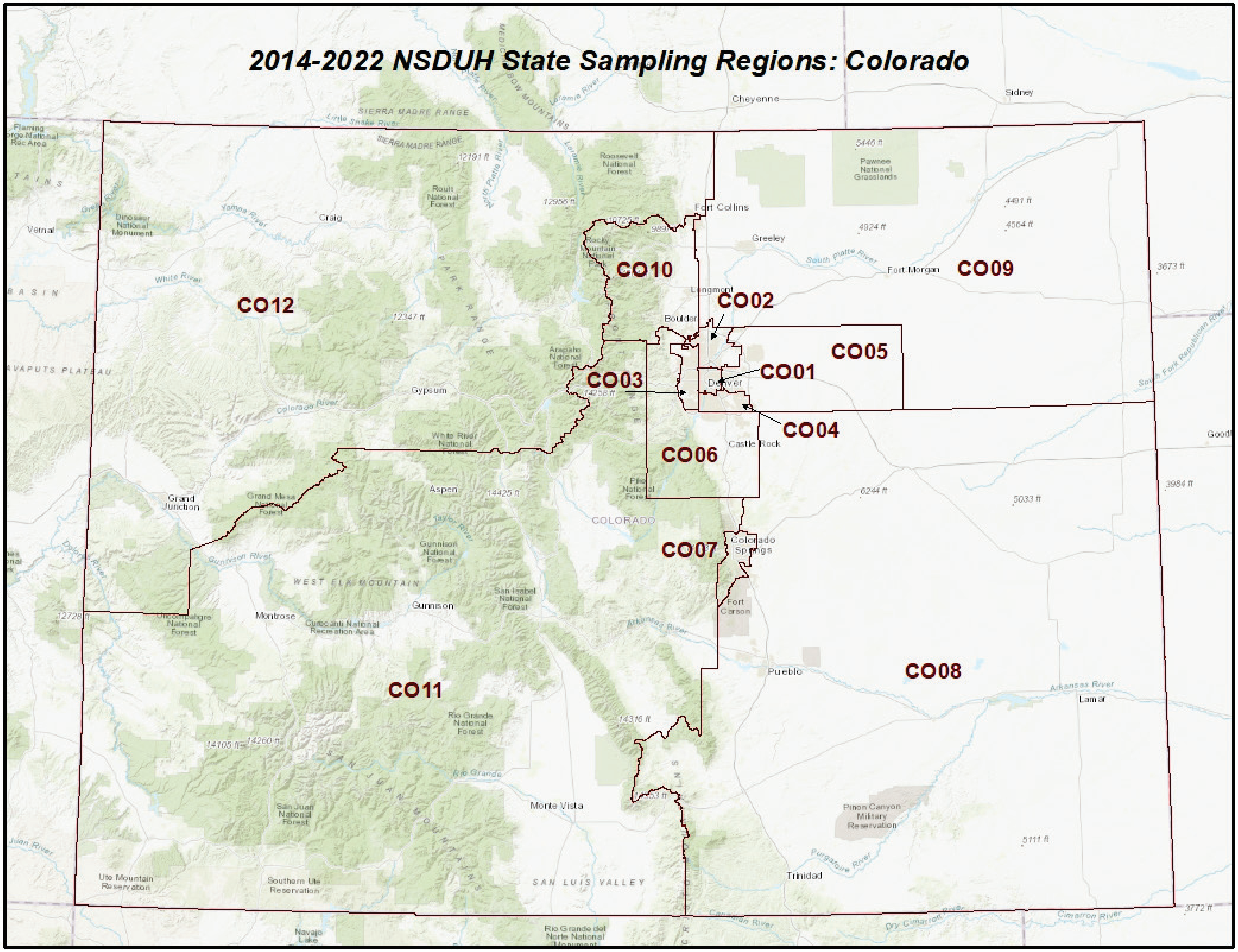

2014-2022 NSDUH State Sampling Regions: Alabama
AL01
AL02
AL03
AL04
AL05
AL06
AL07
AL08
AL09
AL10
AL11
AL12
Florence
Huntsville
Tupelo
Gadsden
Anniston
Oxford
Birmingham
Tuscaloosa
Starkville
Meridian
Montgomery
Auburn
Columbus
Fort Benning
Andalusia
Enterprise
Dothan
Fort Rucker
Mobile
Pensacola
Biloxi
Gulfport
Lynn Haven
Panama City
Chattanooga
Cleveland
Dalton
Rome
Carrollton
Newnan
Lagrange
Eglin Air Force Base
Conecuh National Forest
Talladega National Forest
William B. Bankhead National Forest
Tombigbee National Forest
De Soto National Forest
Camp Shelby
Apalachicola National Forest
Wilson Lake
Wheeler Lake
Weiss Lake
Lake Martin
West Point Lake
Lewis Smith Lake
SAND MOUNTAIN
BLACK BELT
ALABAMA
Sipsey River
Conecuh River
Chattahoochee River
Escambia River
Chickasawhay River
Tallapoosa River
Coosa River
Echota Cherokee Sdtsa
Star Muskogee Creek Sdtsa
Cher O-Creek Sdtsa
Pelham Range

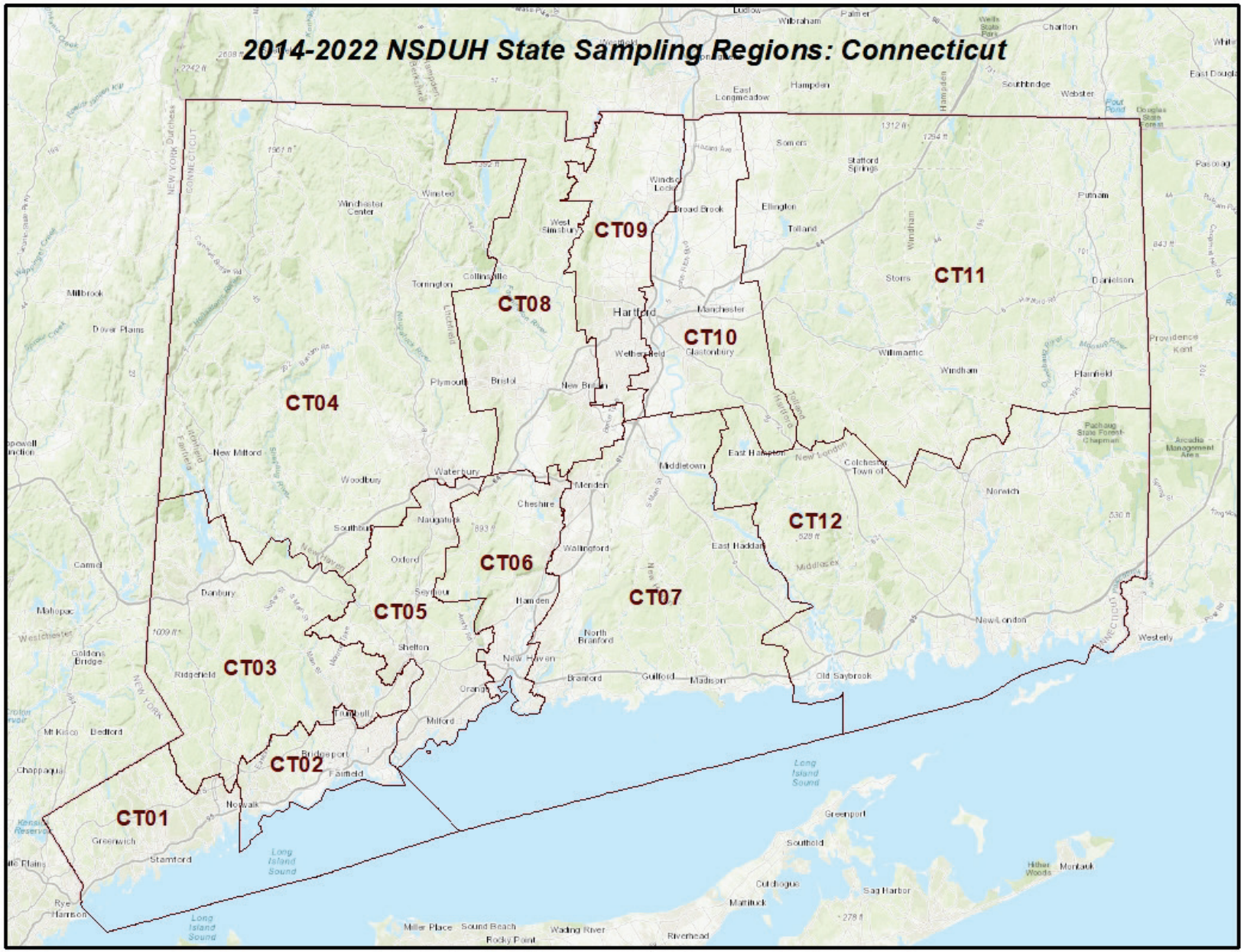

2014-2022 NSDUH State Sampling Regions: Alaska
Anchorage Area
AK01
AK02
AK03
AK04
AK05
AK06
AK07
AK08
AK09
AK10
AK11
AK12

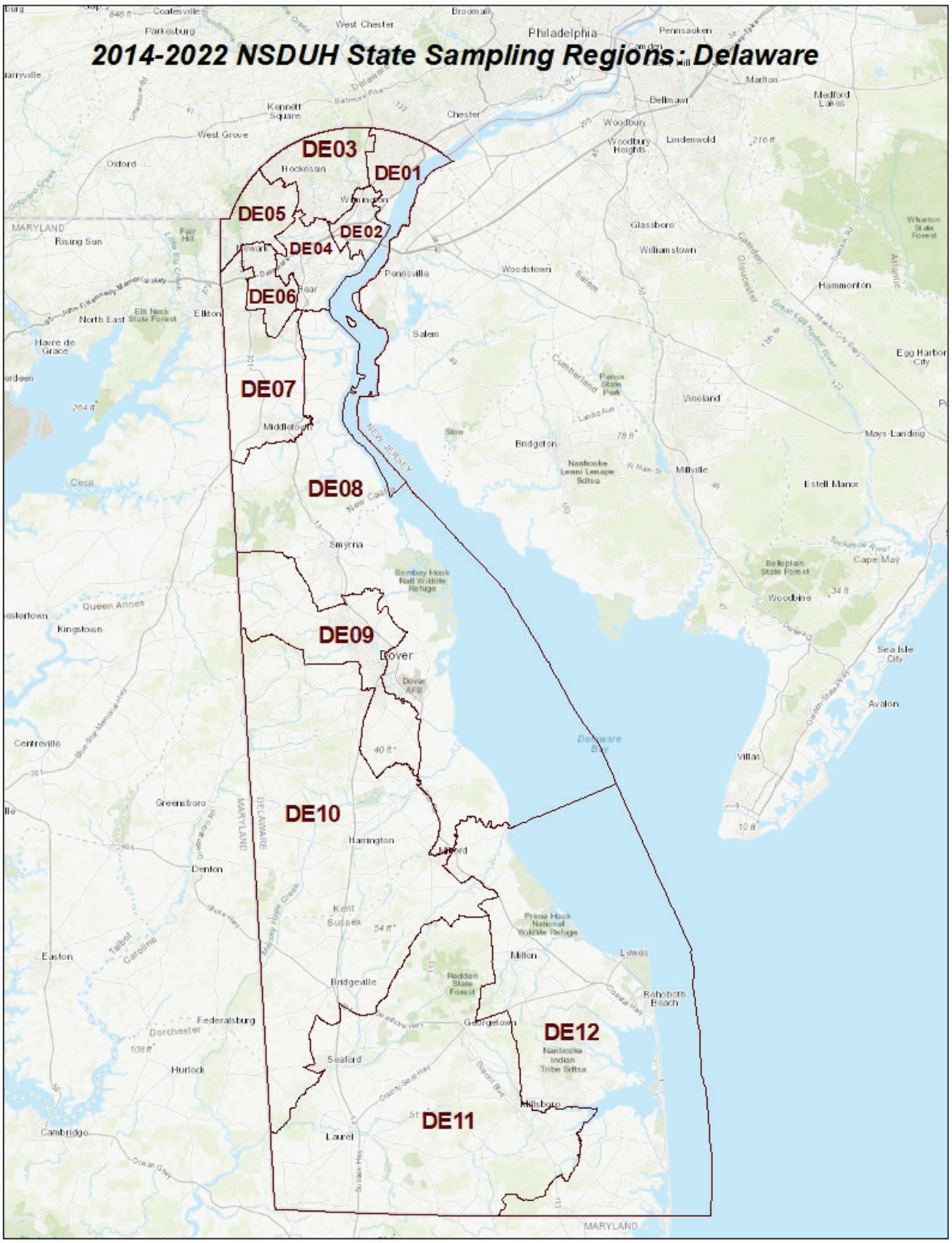

2014-2022 NSDUH State Sampling Regions: Arizona
AZ01
AZ02
AZ03
AZ04
AZ05
AZ06
AZ07
AZ08
AZ09
AZ10
AZ11
AZ12
Kingman
Flagstaff
Winslow
Snowflake
Payson
Show Low
Peoria
Phoenix
Casa Grande
Safford
Sierra Vista
Nogales
Yuma
Tuba City
Kayenta
Page
Las Vegas
Henderson
Lake Havasu City
COLORADO PLATEAU
SAN FRANCISCO PLATEAU
PAINTED DESERT
SONORAN DESERT
GILA DESERT
UNITED STATES
MEXICO
SONORA

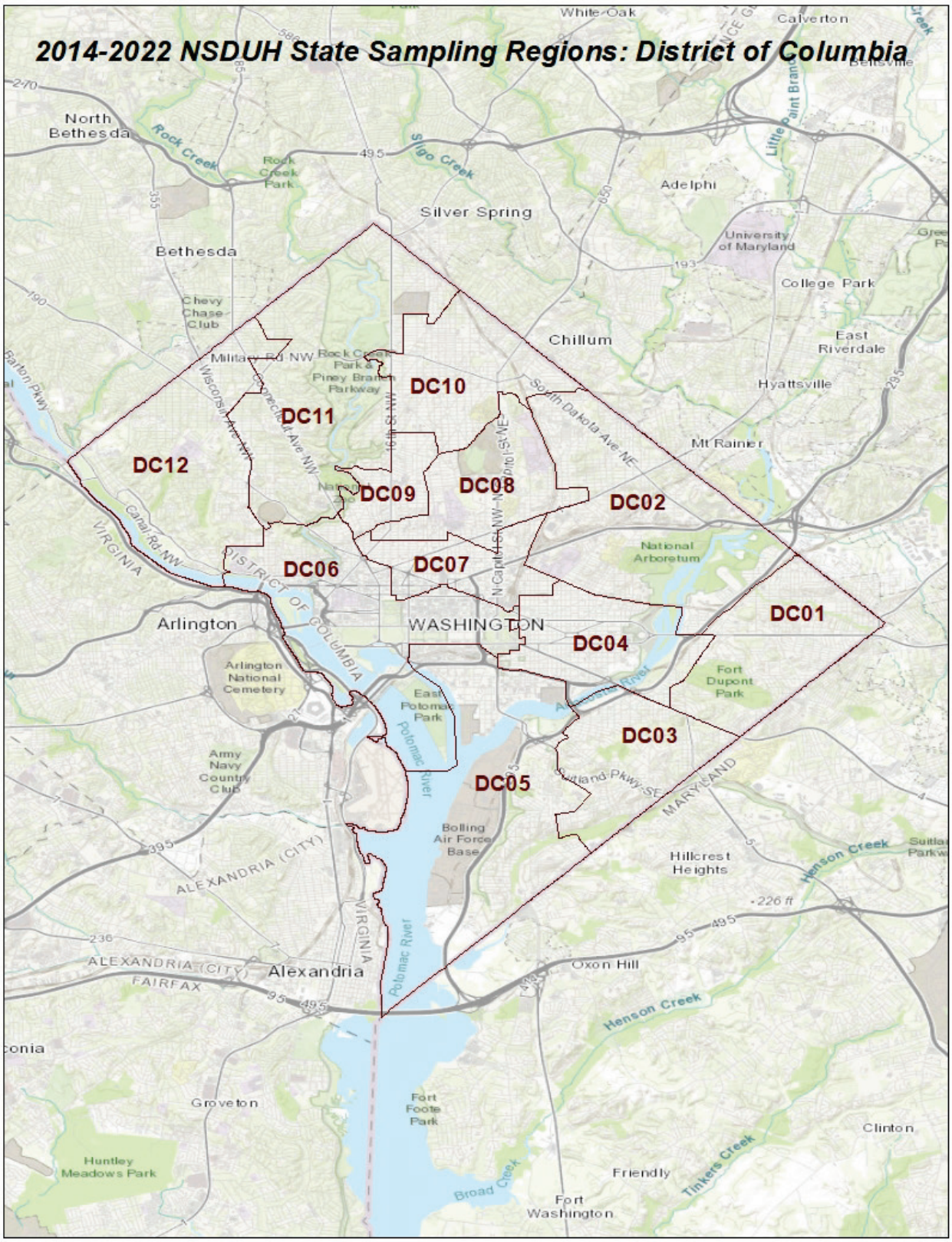

2014-2022 NSDUH State Sampling Regions: Arkansas
AR01
AR02
AR03
AR04
AR05
AR06
AR07
AR08
AR09
AR10
AR11
AR12
OZARK PLATEAU
OUACHITA MOUNTAINS
ARKANSAS
MISSISSIPPI
Springfield
Joplin
West Plains
Poplar Bluff
Harrison
Mountain Home
Fayetteville
Springdale
Batesville
Jonesboro
Fort Smith
Fort Chaffee
Russellville
Conway
Memphis
Little Rock
Hot Springs
Pine Bluff
Arkadelphia
Clarksdale
Cleveland
Greenwood
Greenville
Texarkana
El Dorado
Shreveport
Ruston
Monroe
Tallulah
Vicksburg
Jackson
Table Rock Lake
Beaver Lake
Norfork Lake
Greers Ferry Lake
Lake Ouachita
Ouachita National Forest
Ozark National Forest
Buffalo National River
De Gray Lake
Sardis Lake
Enid Lake
Lake O' the Pines
Caddo Lake
Delta National Forest


2014-2022 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

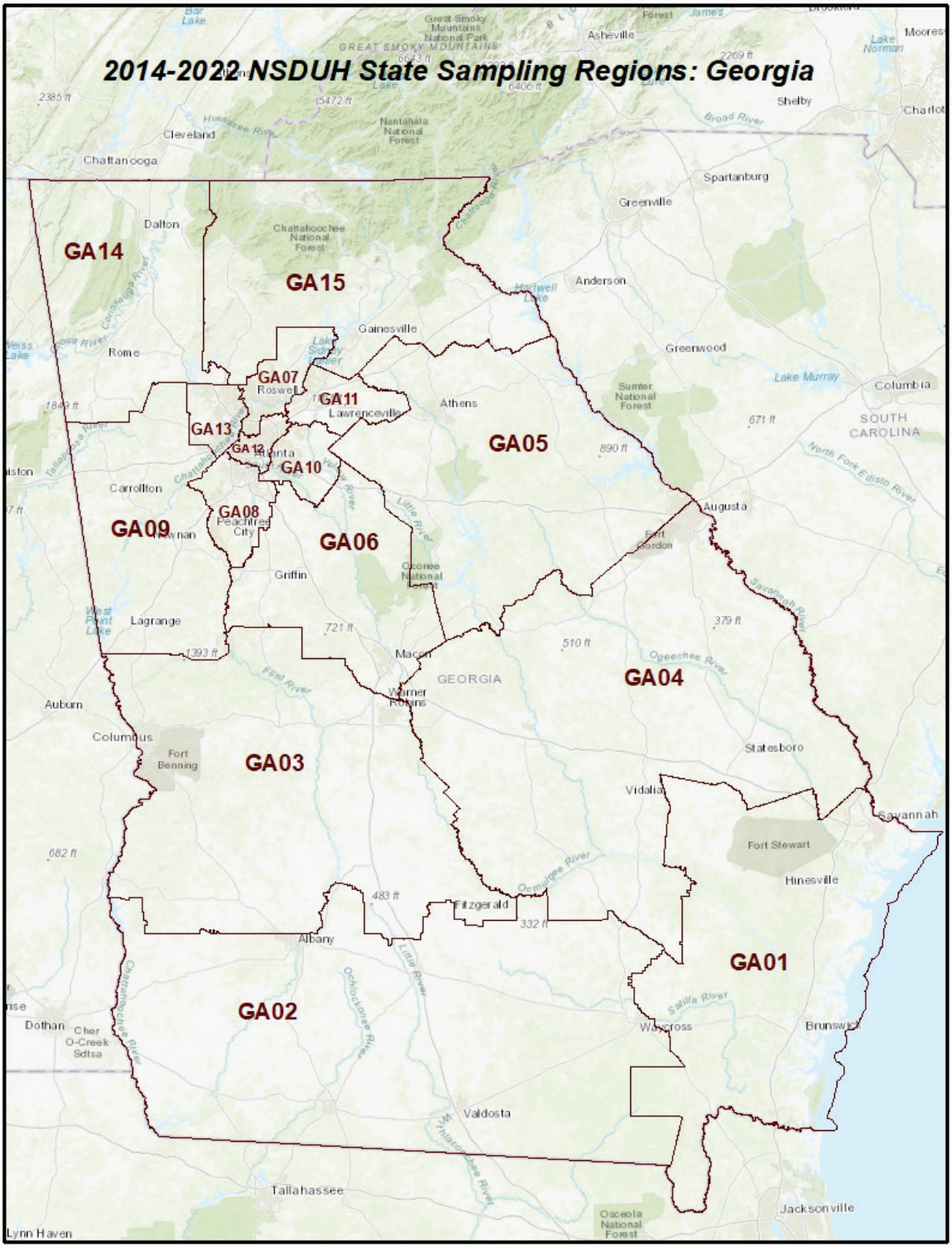

2014-2022 NSDUH State Sampling Regions: Colorado
CO01
CO02
CO03
CO04
CO05
CO06
CO07
CO08
CO09
CO10
CO11
CO12
Cheyenne
Sidney
Fort Collins
Greeley
Fort Morgan
Longmont
Boulder
Denver
Castle Rock
Colorado Springs
Fort Carson
Pueblo
Lamar
Trinidad
Craig
Vernal
Gypsum
Aspen
Grand Junction
Montrose
Gunnison
Monte Vista
COLORADO
SAN LUIS VALLEY
Pawnee National Grasslands
Pinon Canyon Military Reservation
Ute Mountain Reservation
Southern Ute Reservation

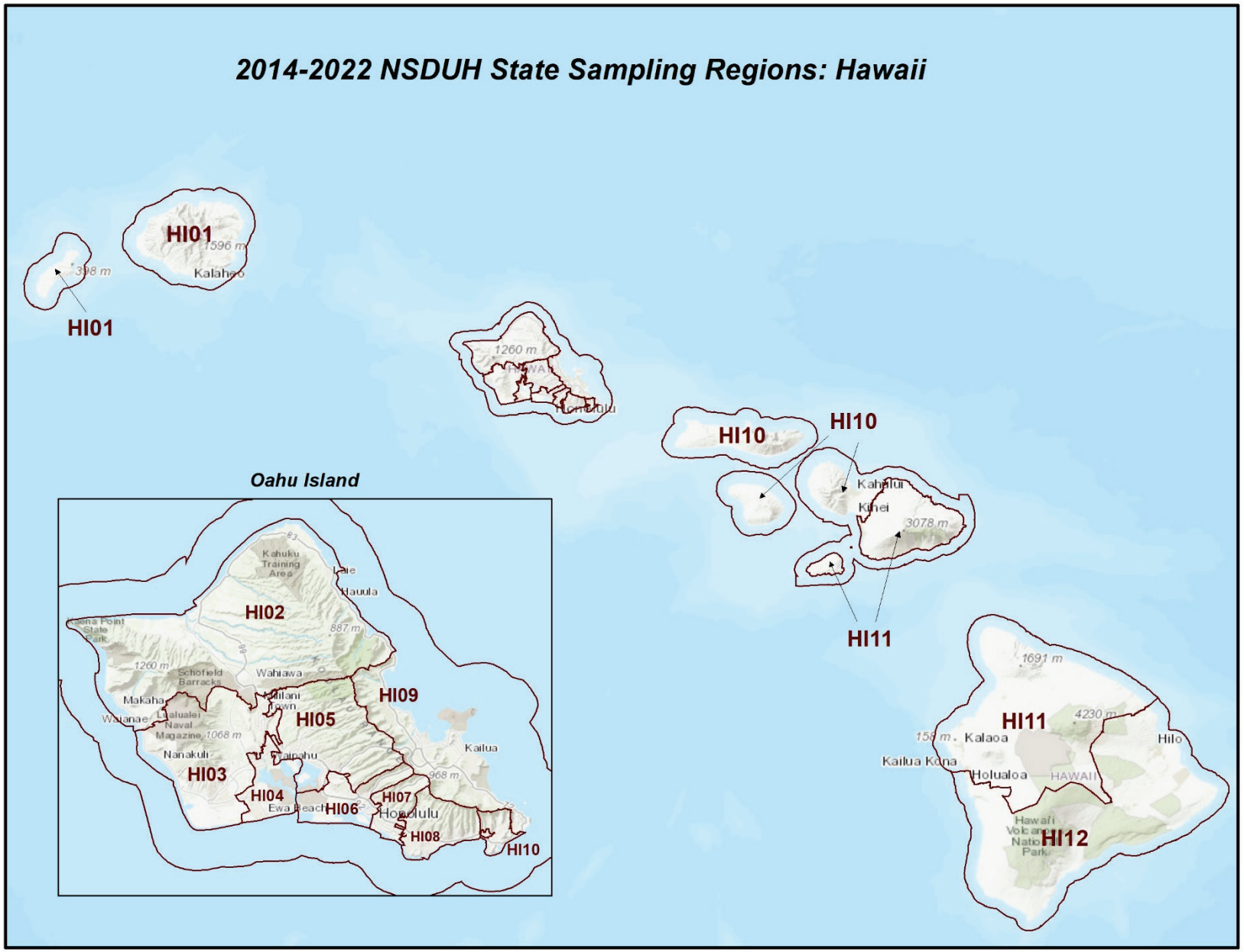

2014-2022 NSDUH State Sampling Regions: Connecticut
CT01
CT02
CT03
CT04
CT05
CT06
CT07
CT08
CT09
CT10
CT11
CT12
Hartford
Waterbury
Bridgeport
Stamford
New Haven
Danbury
New Britain
Norwalk
Long Island Sound

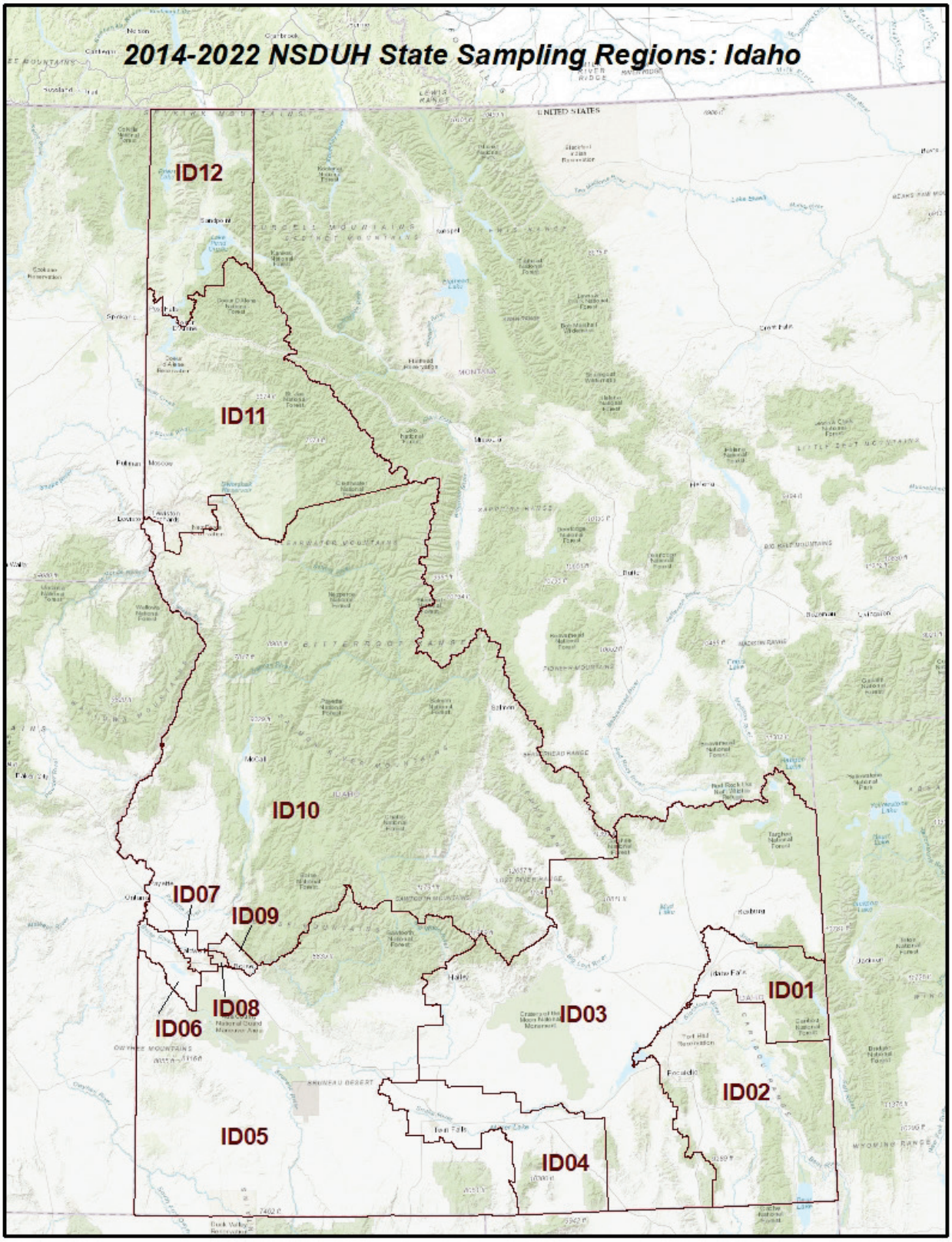

2014-2022 NSDUH State Sampling Regions: Delaware
DE01
DE02
DE03
DE04
DE05
DE06
DE07
DE08
DE09
DE10
DE11
DE12
Philadelphia
Wilmington
Dover
Salisbury
Cape May
Atlantic City
MARYLAND
DELAWARE
NEW JERSEY

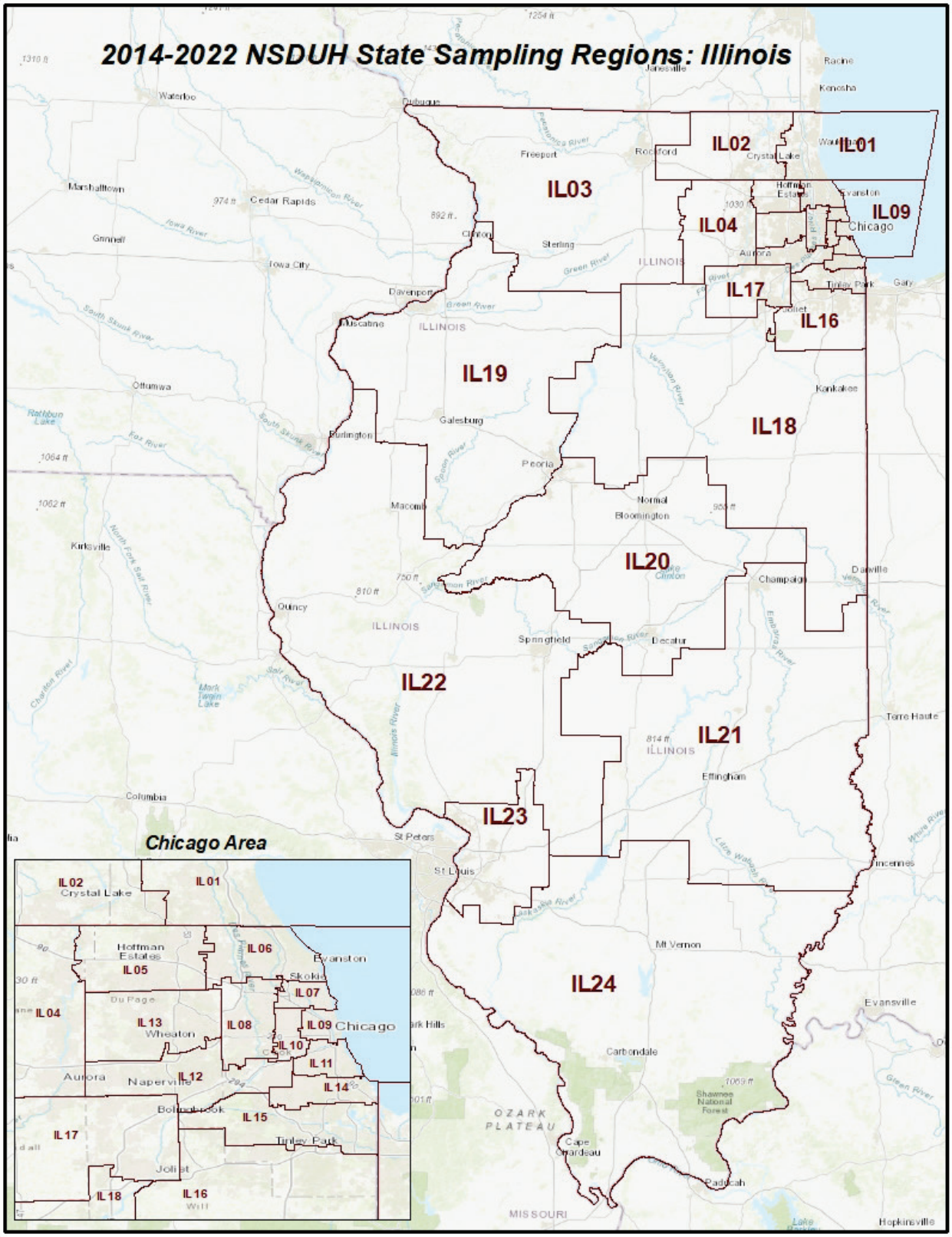

2014-2022 NSDUH State Sampling Regions: District of Columbia
DC01
DC02
DC03
DC04
DC05
DC06
DC07
DC08
DC09
DC10
DC11
DC12
WASHINGTON
Arlington
Alexandria
Silver Spring
Bethesda
North Bethesda
Chevy Chase Club
Chillum
Adelphi
College Park
East Riverdale
Hyattsville
Mt Rainier
University of Maryland
National Arboretum
Fort Dupont Park
Arlington National Cemetery
East Potomac Park
Army Navy Country Club
Bolling Air Force Base
Hillcrest Heights
Oxon Hill
Fort Foote Park
Groveton
Huntley Meadows Park
Friendly
Fort Washington
Clinton
Potomac River
Rock Creek
Sligo Creek
Henson Creek
Tinkers Creek
Broad Creek
VIRGINIA
MARYLAND
DISTRICT OF COLUMBIA
ALEXANDRIA (CITY)
FAIRFAX

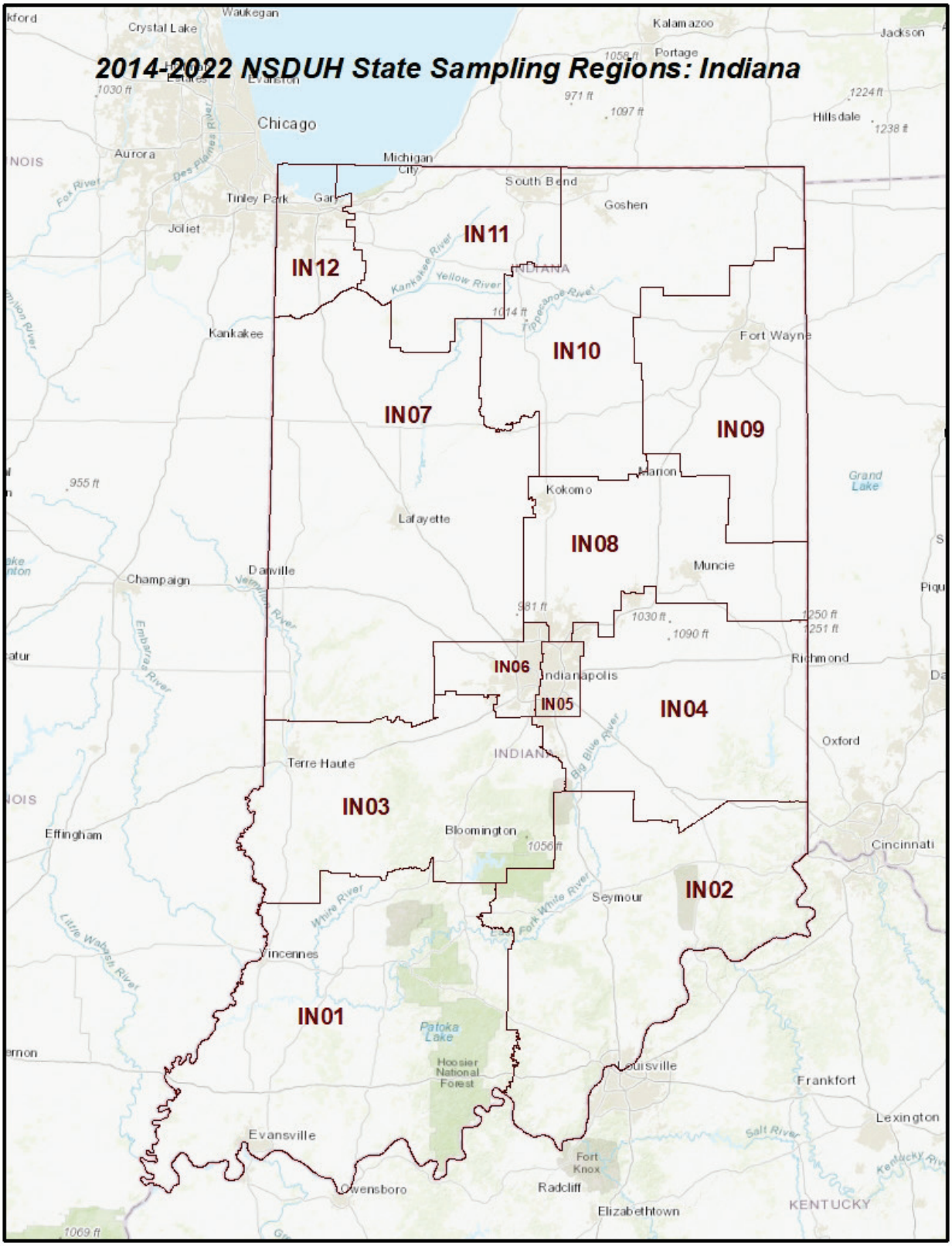

2014-2022 NSDUH State Sampling Regions: Florida
Tampa Area
Miami Area
Orlando Area
FL01
FL02
FL03
FL04
FL05
FL06
FL07
FL08
FL09
FL10
FL11
FL12
FL13
FL14
FL15
FL16
FL17
FL18
FL19
FL20
FL21
FL22
FL23
FL24
FL25
FL26
FL27
FL28
FL29
FL30
ALABAMA
GEORGIA
Macon
Warner Robins
Montgomery
Columbus
Albany
Dothan
Valdosta
Charleston
Savannah
Pensacola
Tallahassee
Jacksonville
Gainesville
Palm Coast
Melbourne
Palm Bay
Port St Lucie
West Palm Beach
Boca Raton
Pompano Beach
Miami
Cape Coral
Tampa
Lakeland
Clearwater
St. Petersburg
Sarasota
Orlando
Daytona Beach
Port Orange
Titusville
Homestead
Coral Springs
Pembroke Pines
Hollywood
Gulf of Mexico
Straits of Florida
Havana
Everglades National Park
Big Cypress National Preserve
Apalachicola National Forest
Ocala National Forest

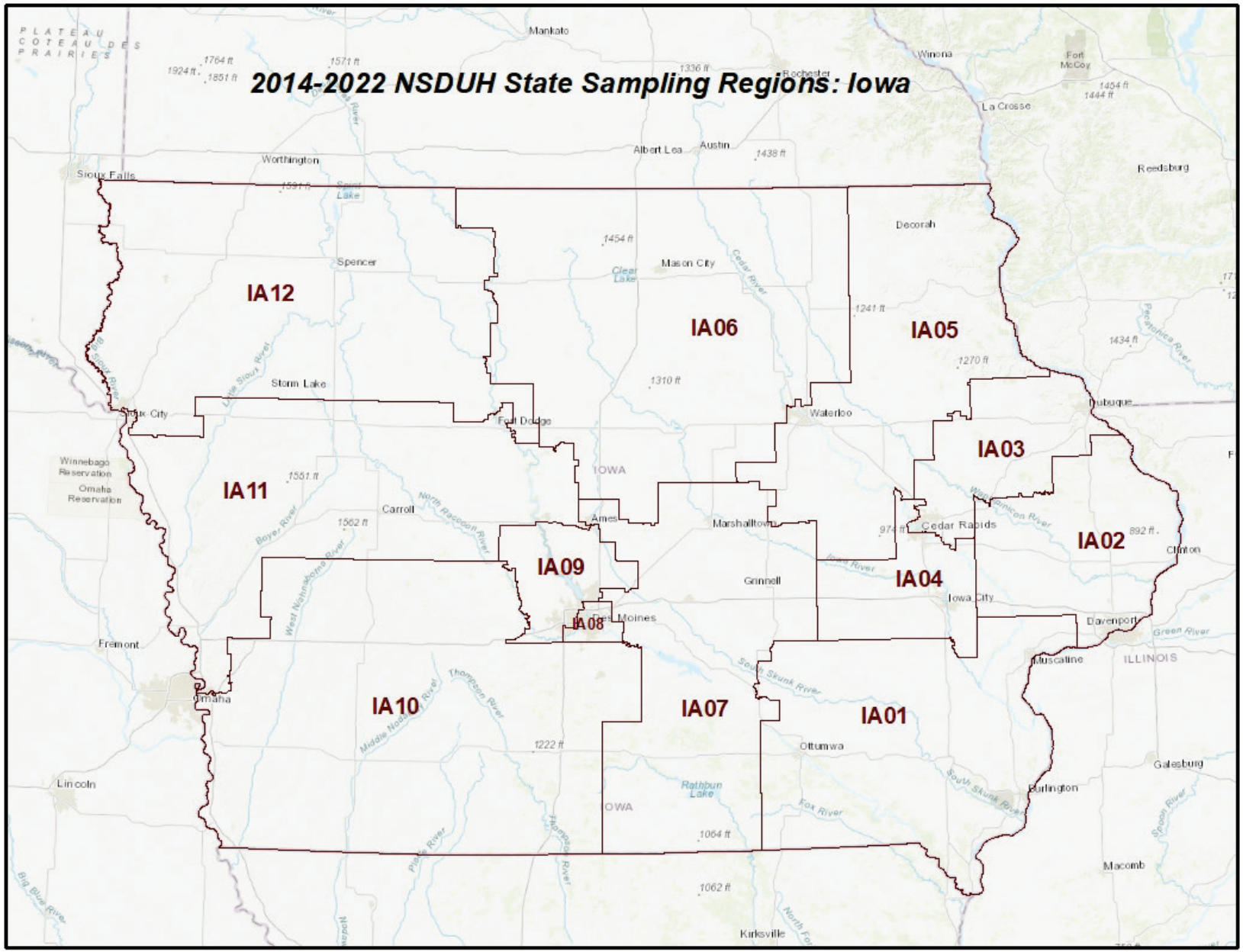

2014-2022 NSDUH State Sampling Regions: Georgia
GA14
GA15
GA07
GA11
GA13
GA12
GA10
GA05
GA08
GA09
GA06
GA04
GA03
GA01
GA02
GEORGIA
Chattanooga
Cleveland
Dalton
Rome
Gainesville
Roswell
Lawrenceville
Athens
Atlanta
Carrollton
Peachtree City
Griffin
Lagrange
Macon
Warner Robins
Auburn
Columbus
Fort Benning
Albany
Dothan
Valdosta
Tallahassee
Lynn Haven
Asheville
Greenville
Spartanburg
Anderson
Greenwood
Augusta
Fort Gordon
Statesboro
Vidalia
Savannah
Fort Stewart
Hinesville
Fitzgerald
Waycross
Brunswick
Jacksonville
Columbia
Shelby
SOUTH CAROLINA

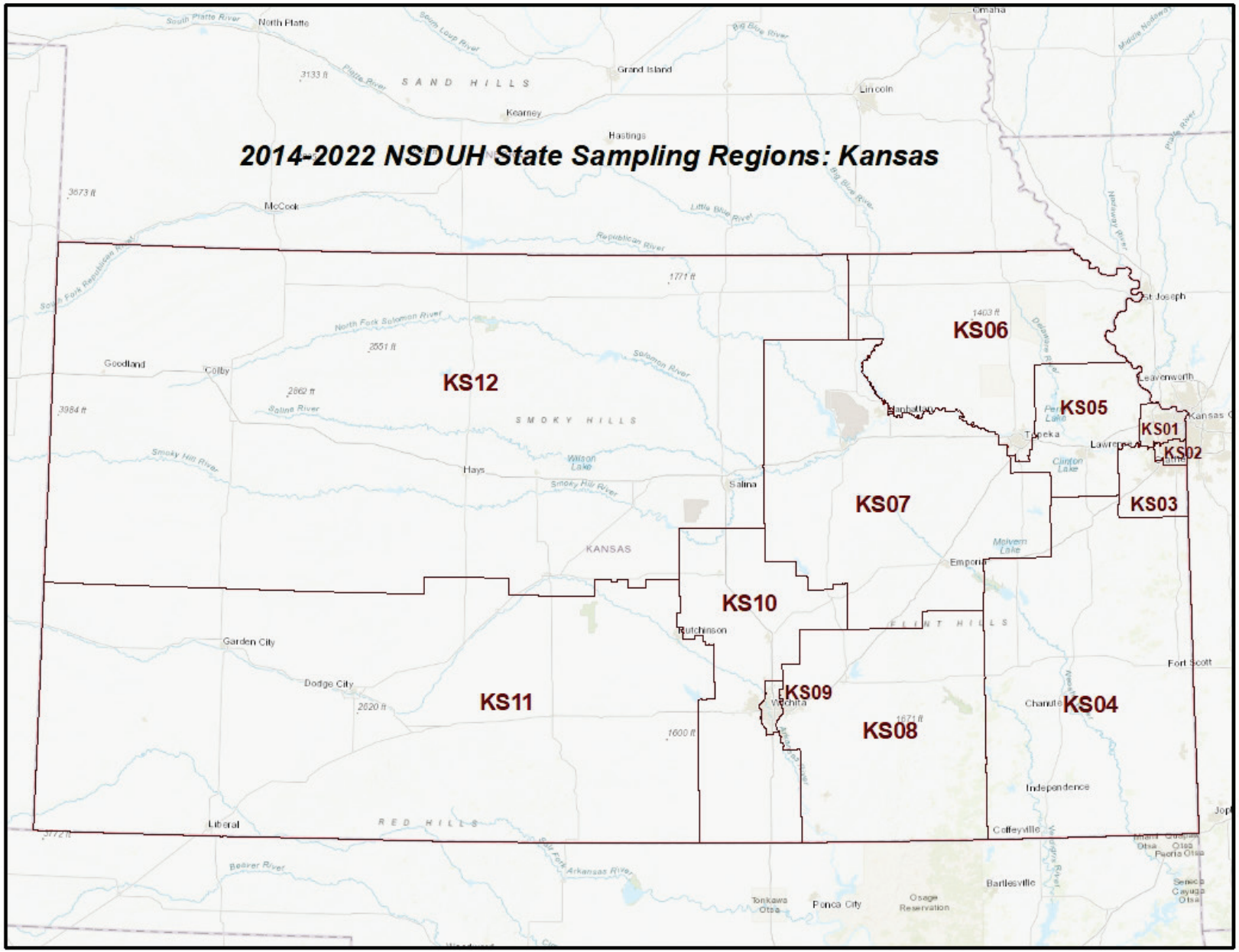

2014-2022 NSDUH State Sampling Regions: Hawaii
HI01
1596 m
Kalaheo
398 m
HI01
1260 m
HI10
HI10
Kahului
Kihei
3078 m
HI11
Oahu Island
Kahuku Training Area
Laie
Hauula
HI02
887 m
Kaena Point State Park
1260 m
Schofield Barracks
Wahiawa
Makaha
Waianae
Lualualei Naval Magazine
1068 m
Mililani Town
HI09
HI05
Nanakuli
Waipahu
Kailua
968 m
HI03
HI04
Ewa Beach
HI06
HI07
Honolulu
HI08
HI10
1691 m
HI11
4230 m
156 m
Kalaoa
Hilo
Kailua Kona
Holualoa
HAWAII
Hawai'i Volcanoes National Park
HI12

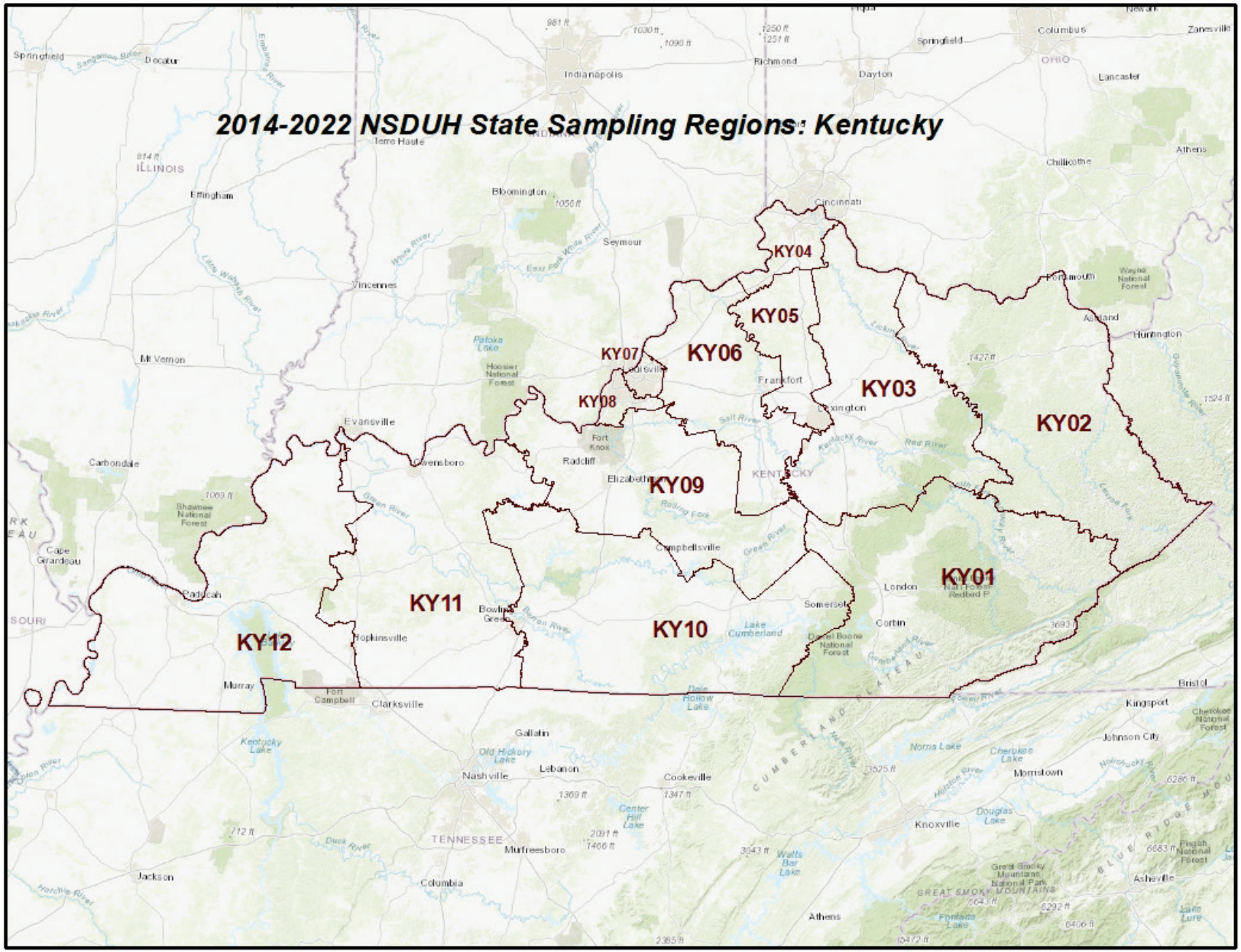

2014-2022 NSDUH State Sampling Regions: Idaho
ID12
ID11
ID10
ID07
ID09
ID08
ID06
ID03
ID01
ID02
ID05
ID04

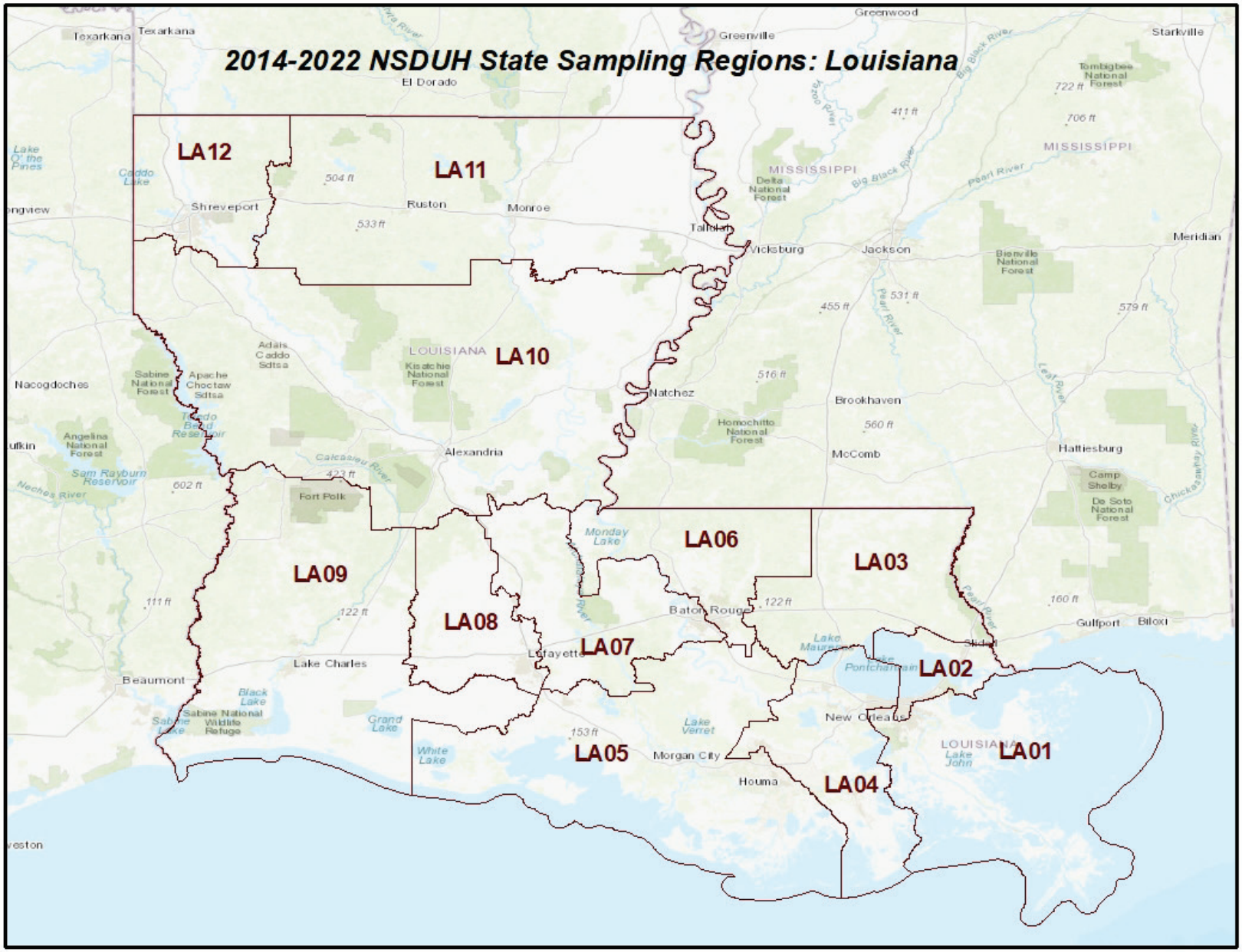

2014-2022 NSDUH State Sampling Regions: Illinois
IL01
IL02
IL03
IL04
IL09
IL16
IL17
IL18
IL19
IL20
IL21
IL22
IL23
IL24
Chicago Area
Crystal Lake
Hoffman Estates
Evanston
Skokie
Chicago
Wheaton
Naperville
Aurora
Bolingbrook
Tinley Park
Joliet
Will
Rockford
Freeport
Dubuque
Waterloo
Cedar Rapids
Marshalltown
Grinnell
Iowa City
Davenport
Muscatine
Clinton
Sterling
Aurora
Kankakee
Ottumwa
Galesburg
Burlington
Peoria
Macomb
Normal
Bloomington
Champaign
Danville
Kirksville
Quincy
Springfield
Decatur
Effingham
Terre Haute
Columbia
St Peters
St Louis
Mt Vernon
Evansville
Carbondale
Cape Girardeau
Paducah
Hopkinsville
OZARK PLATEAU
ILLINOIS
MISSOURI


2014-2022 NSDUH State Sampling Regions: Indiana
IN01
IN02
IN03
IN04
IN05
IN06
IN07
IN08
IN09
IN10
IN11
IN12
Waukegan
Crystal Lake
Kalamazoo
Jackson
Portage
Chicago
Hillsdale
Aurora
Michigan City
South Bend
Goshen
Tinley Park
Gary
Joliet
Kankakee
Fort Wayne
Kokomo
Marion
Lafayette
Muncie
Champaign
Danville
Richmond
Indianapolis
Terre Haute
Oxford
Bloomington
Effingham
Cincinnati
Seymour
Vincennes
Patoka Lake
Hoosier National Forest
Louisville
Frankfort
Lexington
Evansville
Fort Knox
Radcliff
Owensboro
Elizabethtown
KENTUCKY
INDIANA
Grand Lake
Kankakee River
Yellow River
Tippecanoe River
Fox River
Des Plaines River
Vermilion River
Embarras River
Big Blue River
White River
Fork White River
Little Wabash River
Salt River
Kentucky River

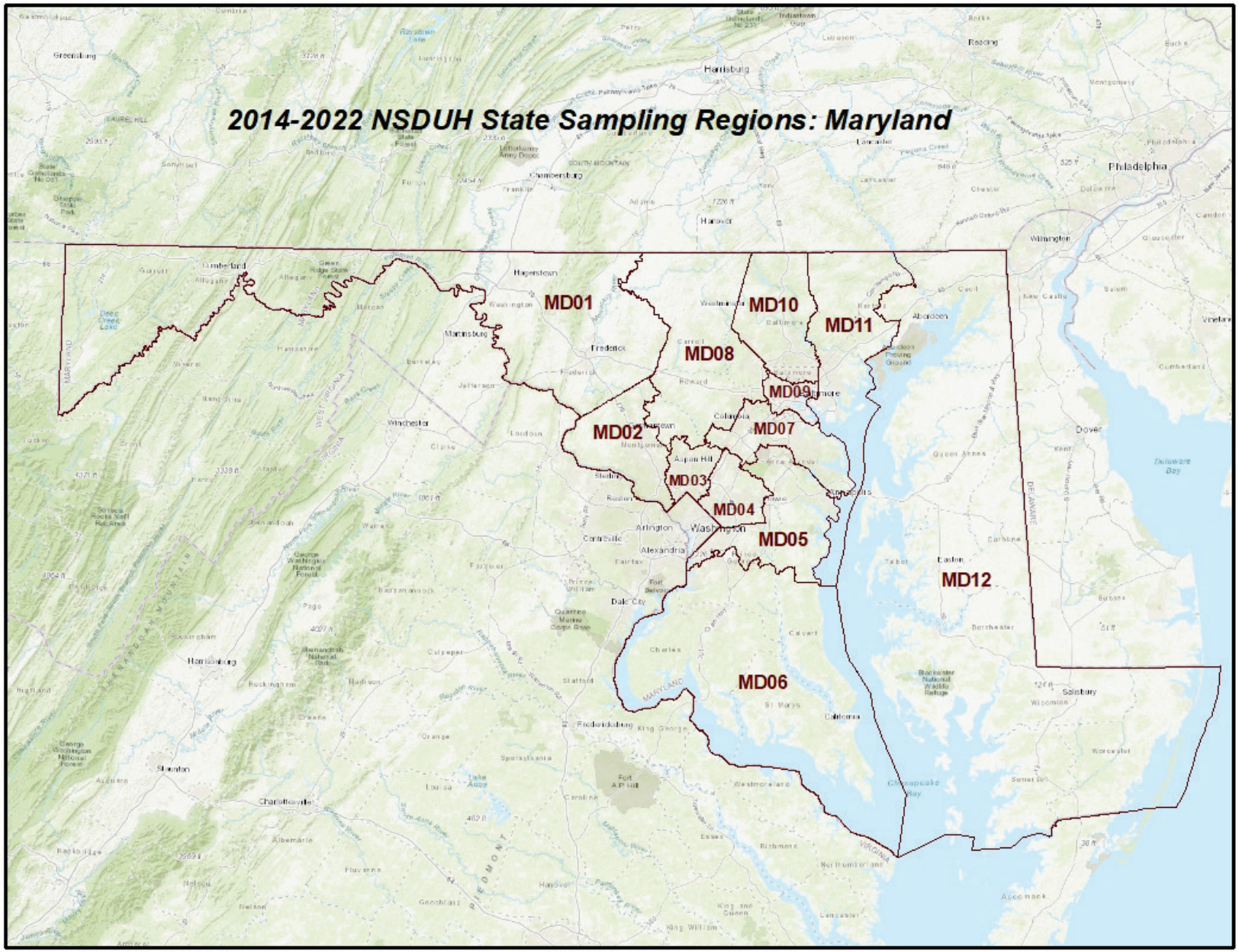

2014-2022 NSDUH State Sampling Regions: Iowa
IA12
IA06
IA05
IA11
IA03
IA09
IA02
IA04
IA08
IA10
IA07
IA01
Mankato
Winona
La Crosse
Albert Lea
Austin
Rochester
Worthington
Sioux Falls
Reedsburg
Spencer
Mason City
Decorah
Storm Lake
Sioux City
Fort Dodge
Waterloo
Dubuque
Winnebago Reservation
Omaha Reservation
IOWA
Carroll
Ames
Marshalltown
Cedar Rapids
Clinton
Grinnell
Iowa City
Davenport
Des Moines
Fremont
Omaha
Muscatine
ILLINOIS
Ottumwa
Galesburg
Lincoln
Burlington
Macomb
Kirksville

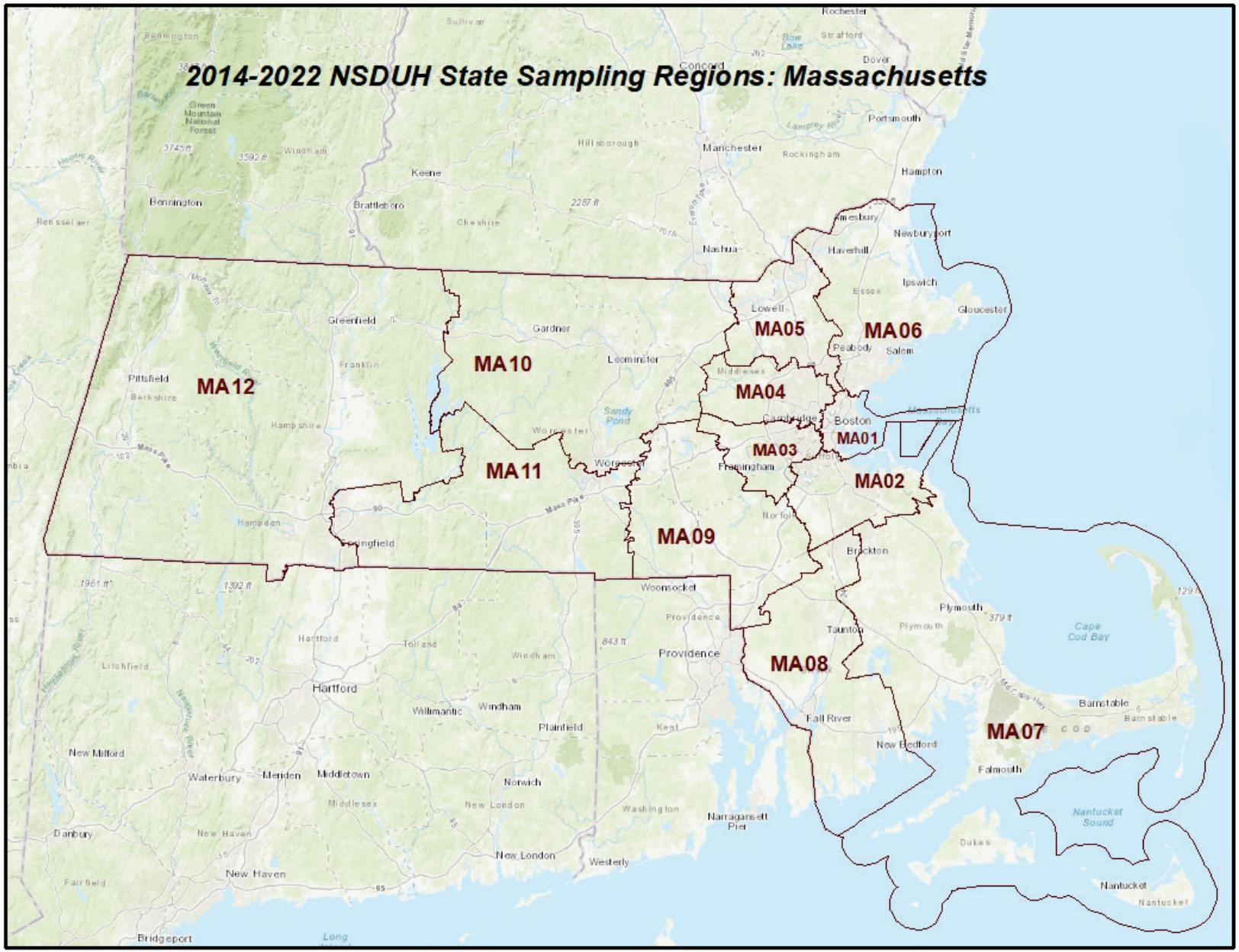

2014-2022 NSDUH State Sampling Regions: Kansas
KS01
KS02
KS03
KS04
KS05
KS06
KS07
KS08
KS09
KS10
KS11
KS12
North Platte
Grand Island
Lincoln
Omaha
Kearney
Hastings
McCook
Goodland
Colby
Hays
Salina
Topeka
Lawrence
Leavenworth
St Joseph
Manhattan
Emporia
Hutchinson
Garden City
Dodge City
Liberal
Wichita
Chanute
Fort Scott
Independence
Coffeyville
Bartlesville
Ponca City
SAND HILLS
SMOKY HILLS
FLINT HILLS
RED HILLS
KANSAS

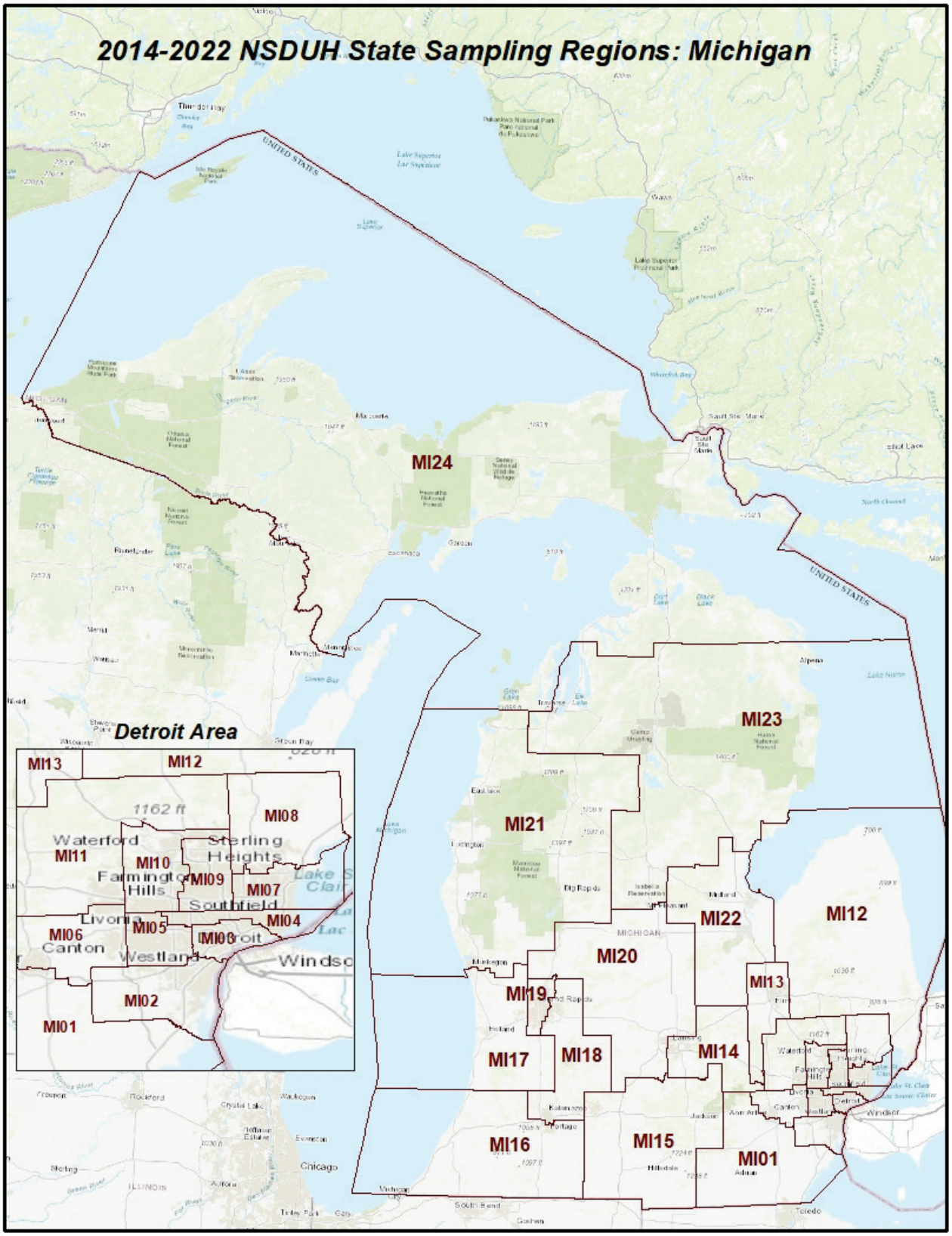

2014-2022 NSDUH State Sampling Regions: Kentucky
KY01
KY02
KY03
KY04
KY05
KY06
KY07
KY08
KY09
KY10
KY11
KY12
Springfield
Decatur
Indianapolis
Richmond
Dayton
Columbus
Springfield
Zanesville
OHIO
Lancaster
Terre Haute
ILLINOIS
Effingham
Bloomington
Cincinnati
Seymour
Chillicothe
Athens
Vincennes
Portsmouth
Ashland
Huntington
Mt Vernon
Patoka Lake
Hoosier National Forest
Louisville
Frankfort
Lexington
Evansville
Fort Knox
Radcliff
Elizabethtown
Owensboro
Carbondale
KENTUCKY
Shawnee National Forest
Cape Girardeau
Paducah
Bowling Green
Somerset
London
Corbin
Daniel Boone National Forest
Hopkinsville
Murray
Fort Campbell
Clarksville
Kentucky Lake
Gallatin
Old Hickory Lake
Nashville
Lebanon
Cookeville
Center Hill Lake
TENNESSEE
Murfreesboro
Jackson
Columbia
Athens
Lake Cumberland
Dale Hollow Lake
Norris Lake
Douglas Lake
Knoxville
Morristown
Johnson City
Kingsport
Bristol
Asheville
Great Smoky Mountains National Park
GREAT SMOKY MOUNTAINS
Watts Bar Lake
Fontana Lake
Pisgah National Forest
Cherokee National Forest
CUMBERLAND PLATEAU
BLUE RIDGE

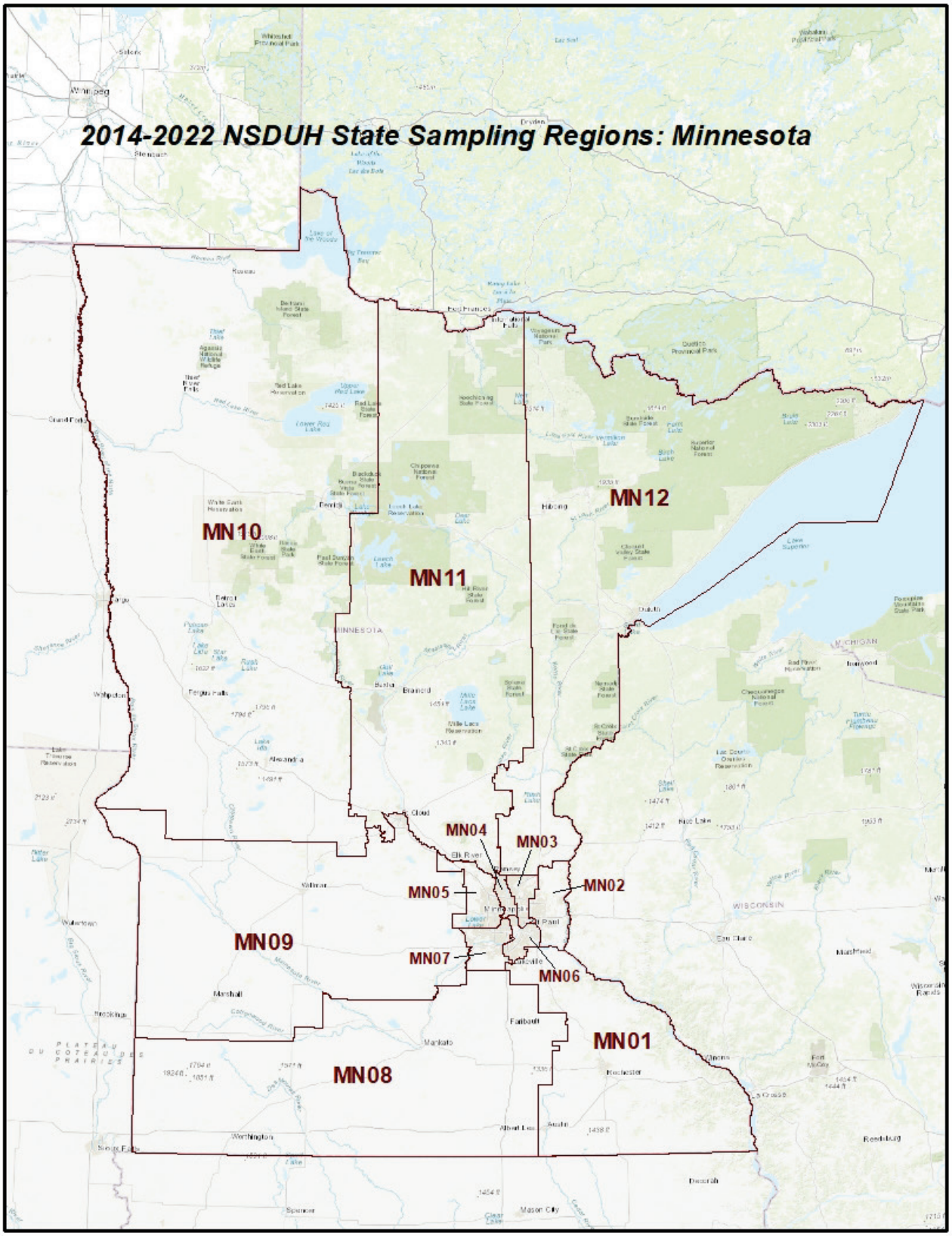

2014-2022 NSDUH State Sampling Regions: Louisiana
LA12
LA11
LA10
LA09
LA08
LA07
LA06
LA05
LA04
LA03
LA02
LA01
Texarkana
El Dorado
Shreveport
Ruston
Monroe
Greenville
Greenwood
Vicksburg
Jackson
Meridian
Starkville
Natchez
Brookhaven
McComb
Hattiesburg
Alexandria
Fort Polk
Lake Charles
Beaumont
Baton Rouge
Lafayette
Gulfport
Biloxi
New Orleans
Houma
Morgan City
Nacogdoches
MISSISSIPPI
LOUISIANA

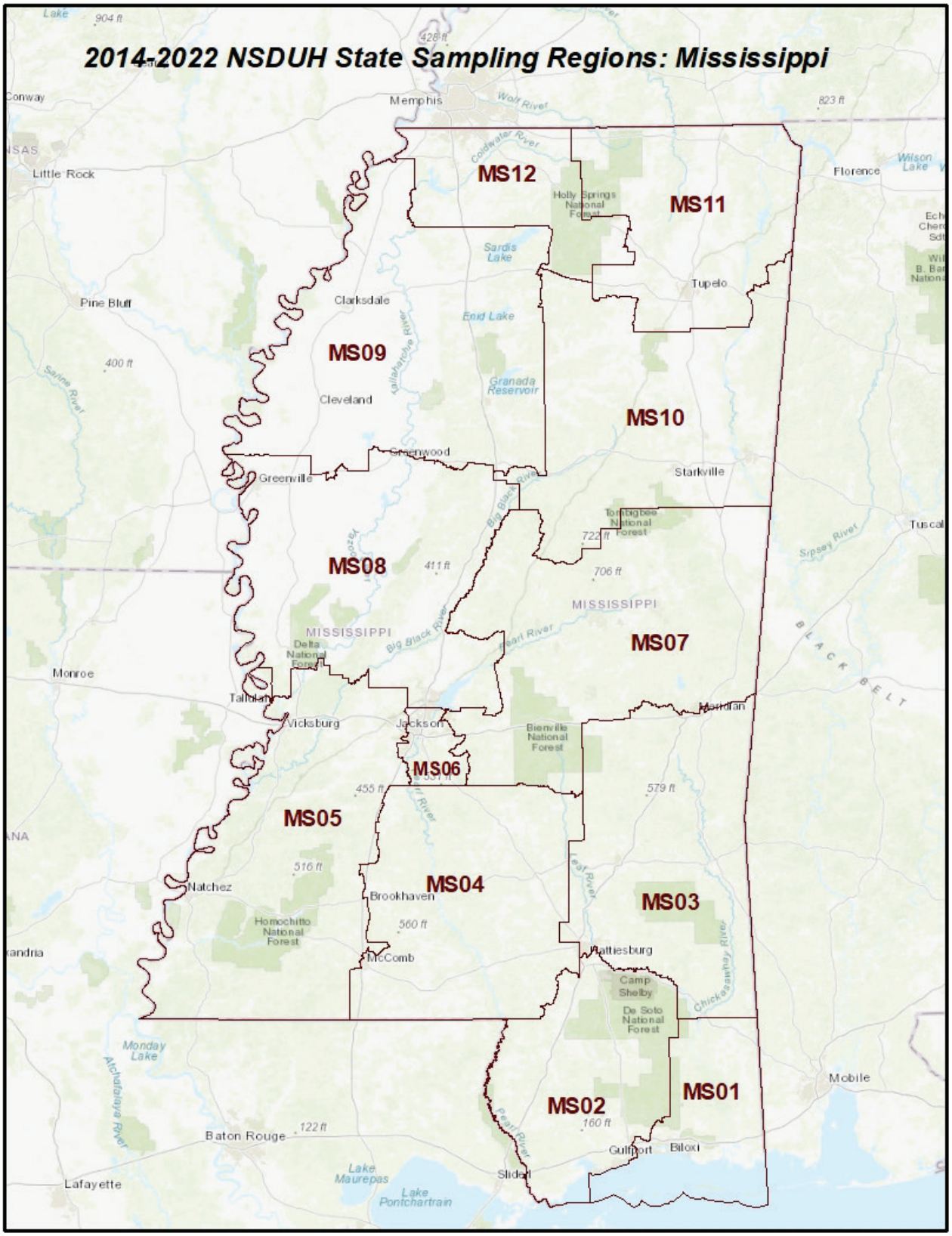

2014-2022 NSDUH State Sampling Regions: Maine
ME12
ME11
ME10
ME09
ME08
ME07
ME06
ME05
ME04
ME03
ME02
ME01
Edmundston
Caribou
NEW BR
NOUVEAU-
Québec
Levis
Thetford Mines
Saint-Georges
herbrooke
Baxter State Park
Moosehead Lake
Millinocket
Lincoln
MAINE
Sebec Lake
Bangor
Berlin
Augusta
Lewiston
Auburn
Brunswick
Conway
Sebago Lake
Portland
Sanford
Dover
Portsmouth
Concord
Manchester
NEW HAMPSHIRE
BOUNDARY MOUNTAINS
BLUE MOUNTAINS
WHITE MOUNTAINS
White Mountain National Forest
Big Lake
Eagle Lake
Long Pond
Flagstaff Lake

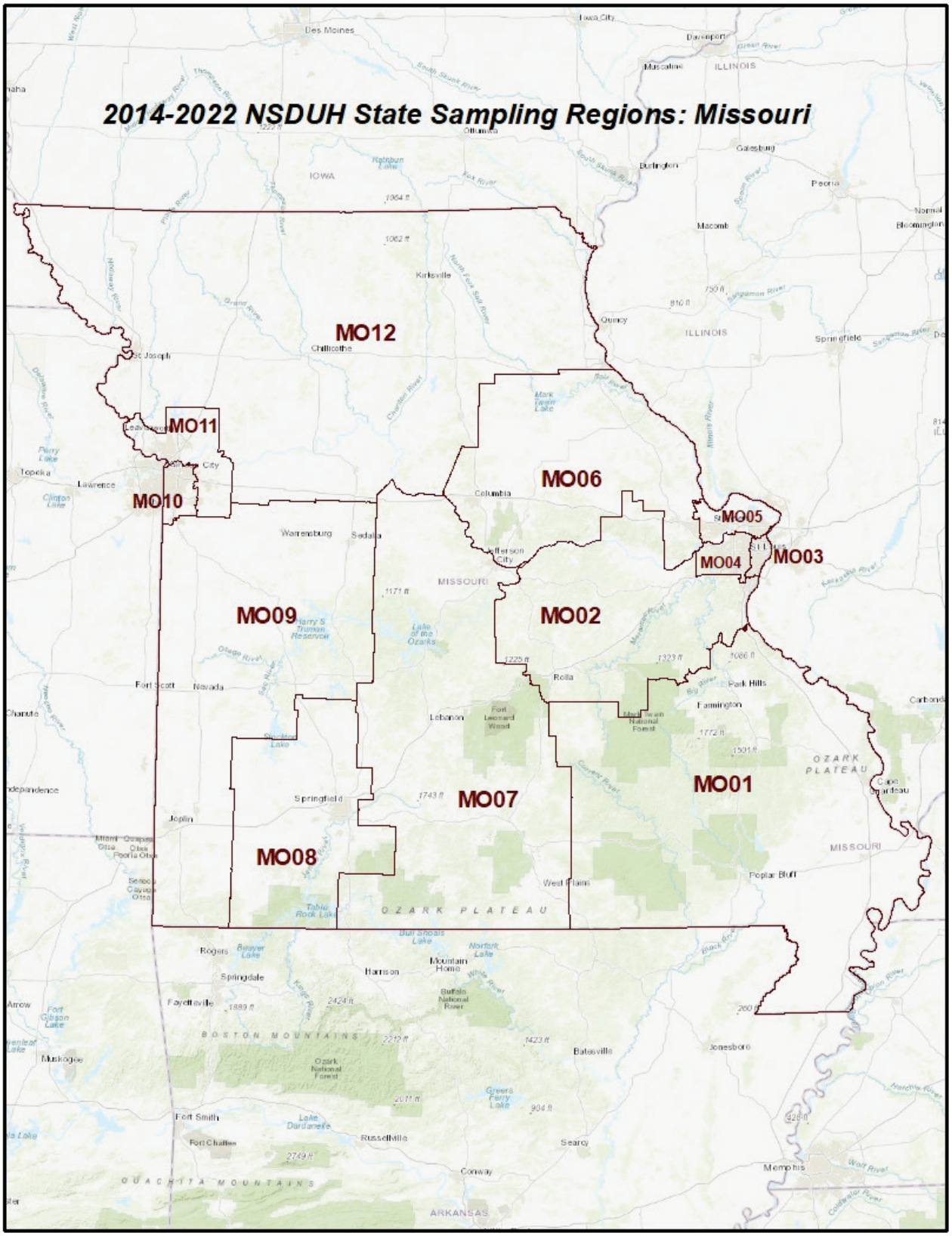

2014-2022 NSDUH State Sampling Regions: Maryland
MD01
MD02
MD03
MD04
MD05
MD06
MD07
MD08
MD09
MD10
MD11
MD12

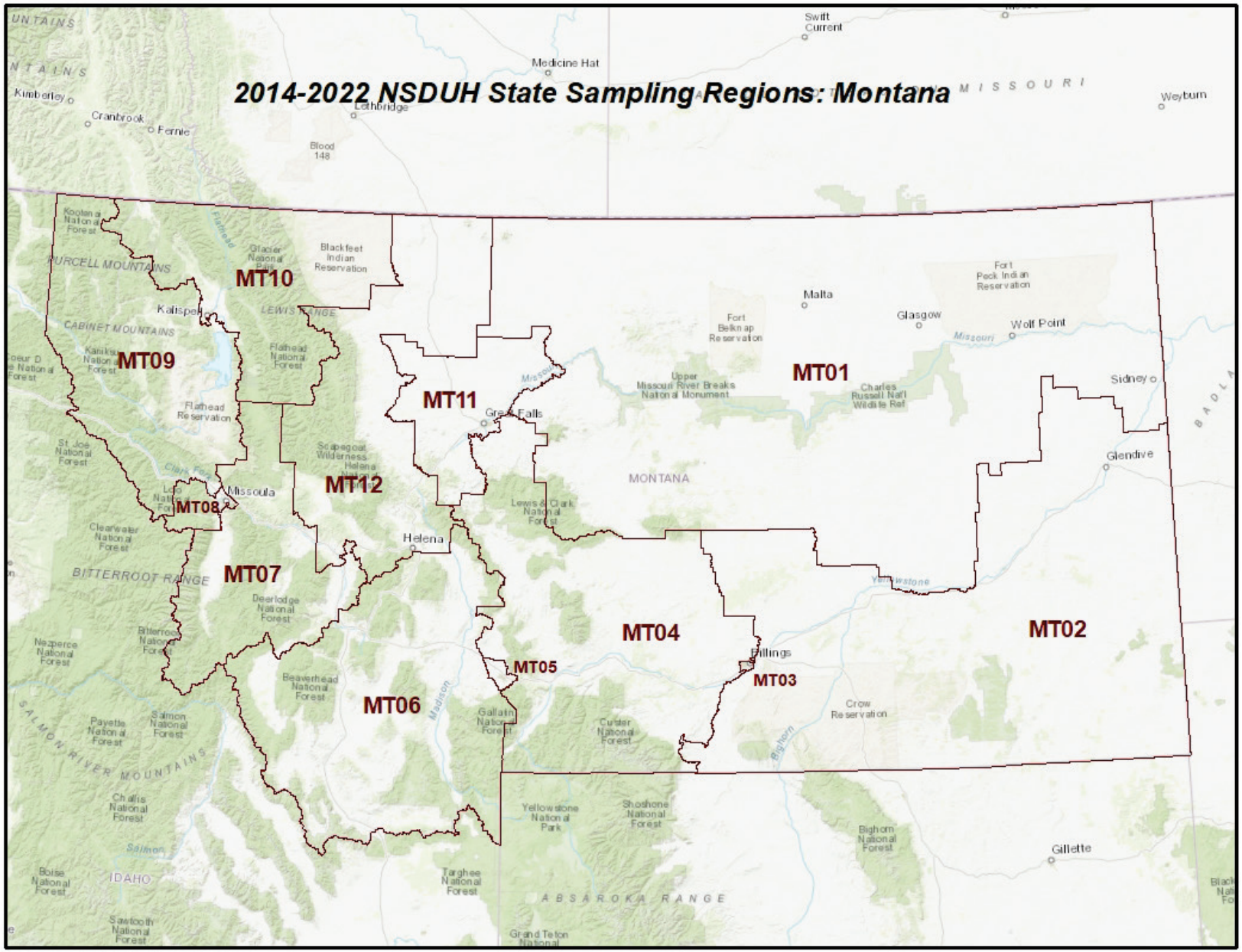

2014-2022 NSDUH State Sampling Regions: Massachusetts
MA01
MA02
MA03
MA04
MA05
MA06
MA07
MA08
MA09
MA10
MA11
MA12

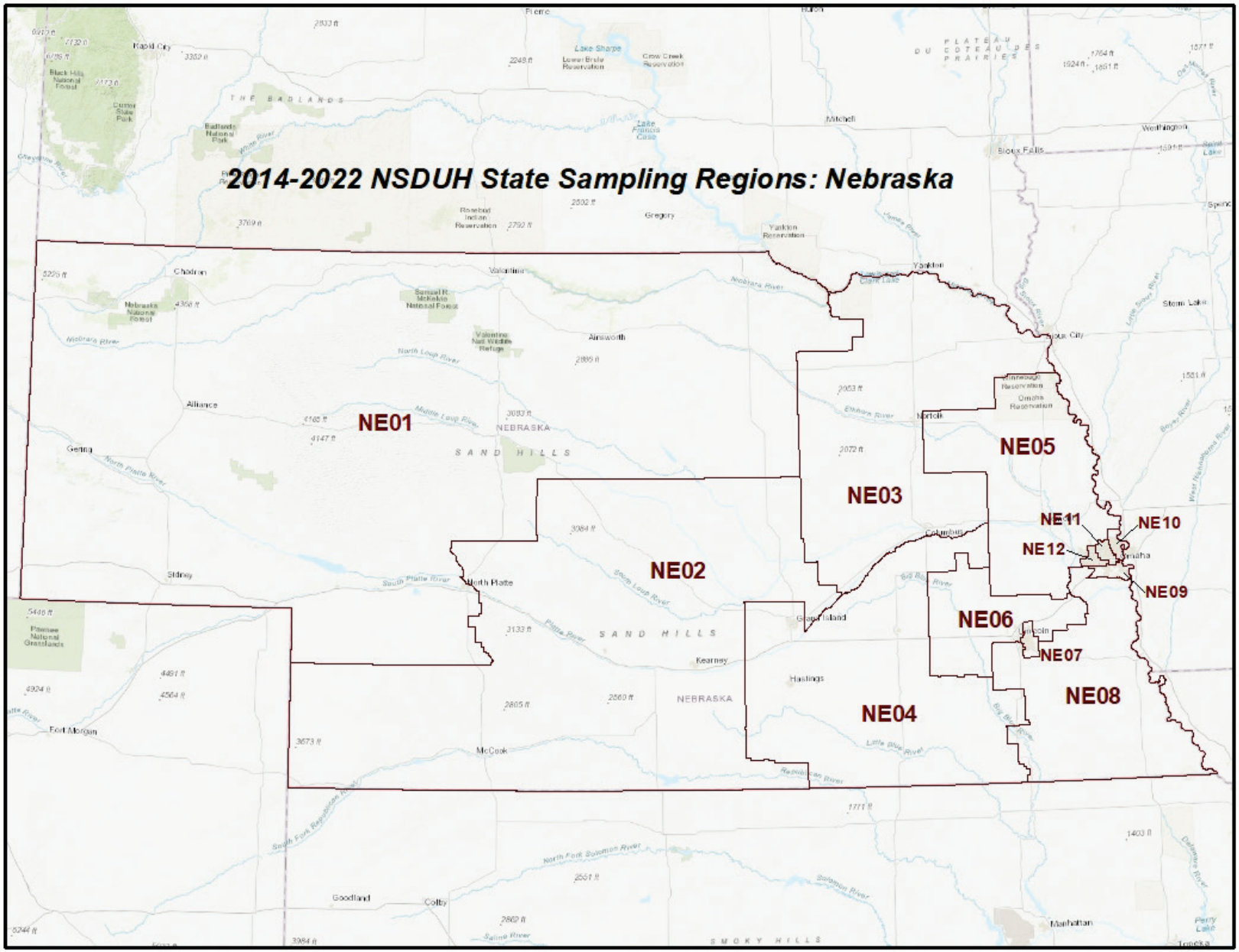

2014-2022 NSDUH State Sampling Regions: Michigan
MI24
MI23
MI21
MI22
MI12
MI20
MI13
MI19
MI14
MI17
MI18
MI16
MI15
MI01
Detroit Area
MI13
MI12
MI08
MI11
MI10
MI09
MI07
MI04
MI05
MI06
MI03
MI02
MI01
Waterford
Sterling Heights
Farmington Hills
Southfield
Livonia
Canton
Westland
Detroit
1162 ft
UNITED STATES
MICHIGAN
ILLINOIS
Chicago
Toledo
Windsor

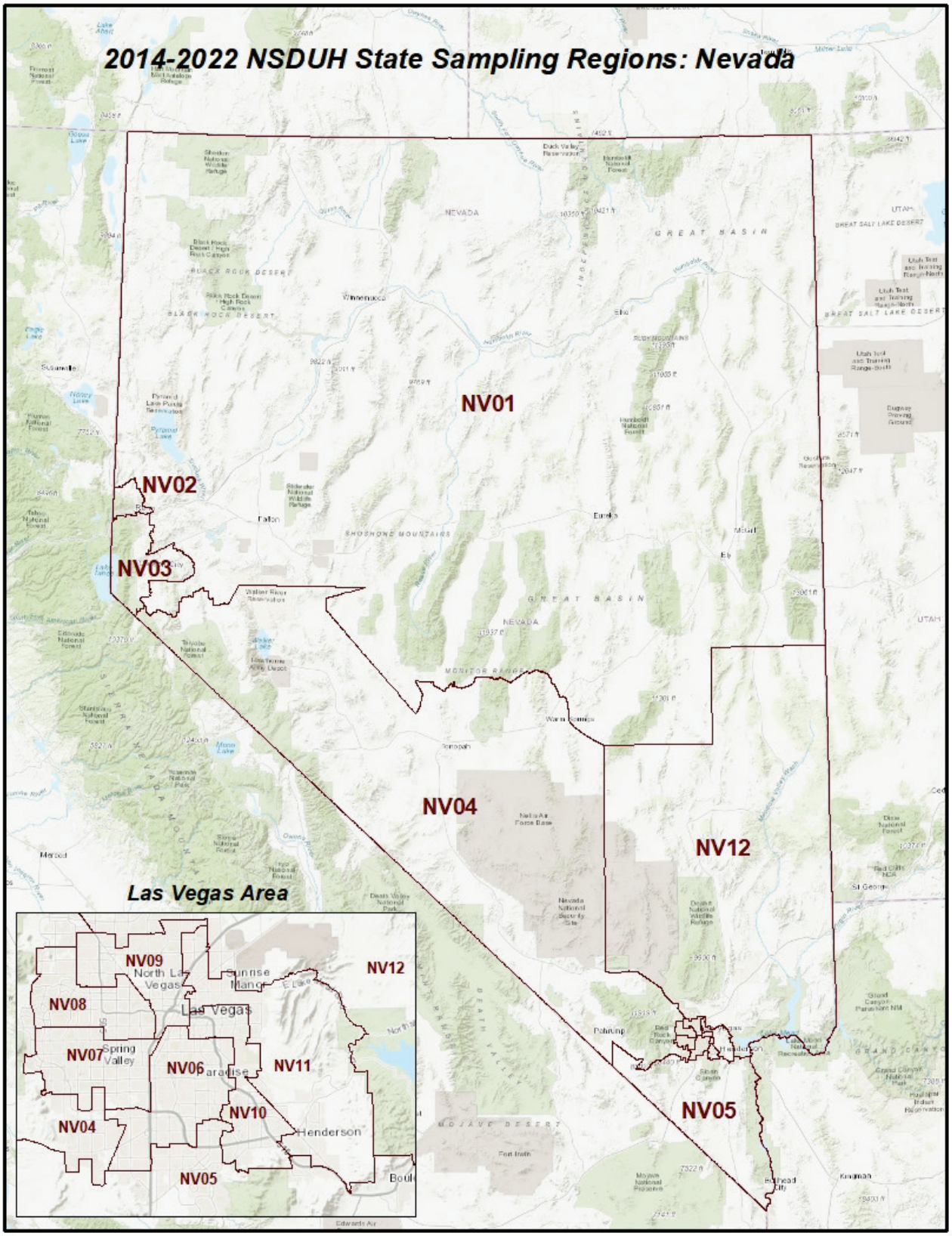

2014-2022 NSDUH State Sampling Regions: Minnesota
MN10
MN11
MN12
MN04
MN03
MN05
MN02
MN09
MN07
MN06
MN01
MN08


2014-2022 NSDUH State Sampling Regions: Mississippi
MS12
MS11
MS09
MS10
MS08
MS07
MS06
MS05
MS04
MS03
MS02
MS01
Memphis
Little Rock
Pine Bluff
Clarksdale
Cleveland
Greenwood
Greenville
Tupelo
Starkville
Florence
Monroe
Tallulah
Vicksburg
Jackson
Meridian
Natchez
Brookhaven
McComb
Hattiesburg
Baton Rouge
Lafayette
Gulfport
Biloxi
Mobile
Slidell
Holly Springs National Forest
Tombigbee National Forest
Delta National Forest
Bienville National Forest
Homochitto National Forest
De Soto National Forest
Camp Shelby
Sardis Lake
Enid Lake
Granada Reservoir
Wolf River
Coldwater River
Big Black River
Pearl River
Yazoo River
Tallahatchie River
Sunflower River
Leaf River
Chickasawhay River
Monday Lake
Lake Maurepas
Lake Pontchartrain
MISSISSIPPI
BLACK BELT
904 ft
428 ft
823 ft
400 ft
722 ft
706 ft
411 ft
455 ft
579 ft
516 ft
560 ft
122 ft
160 ft

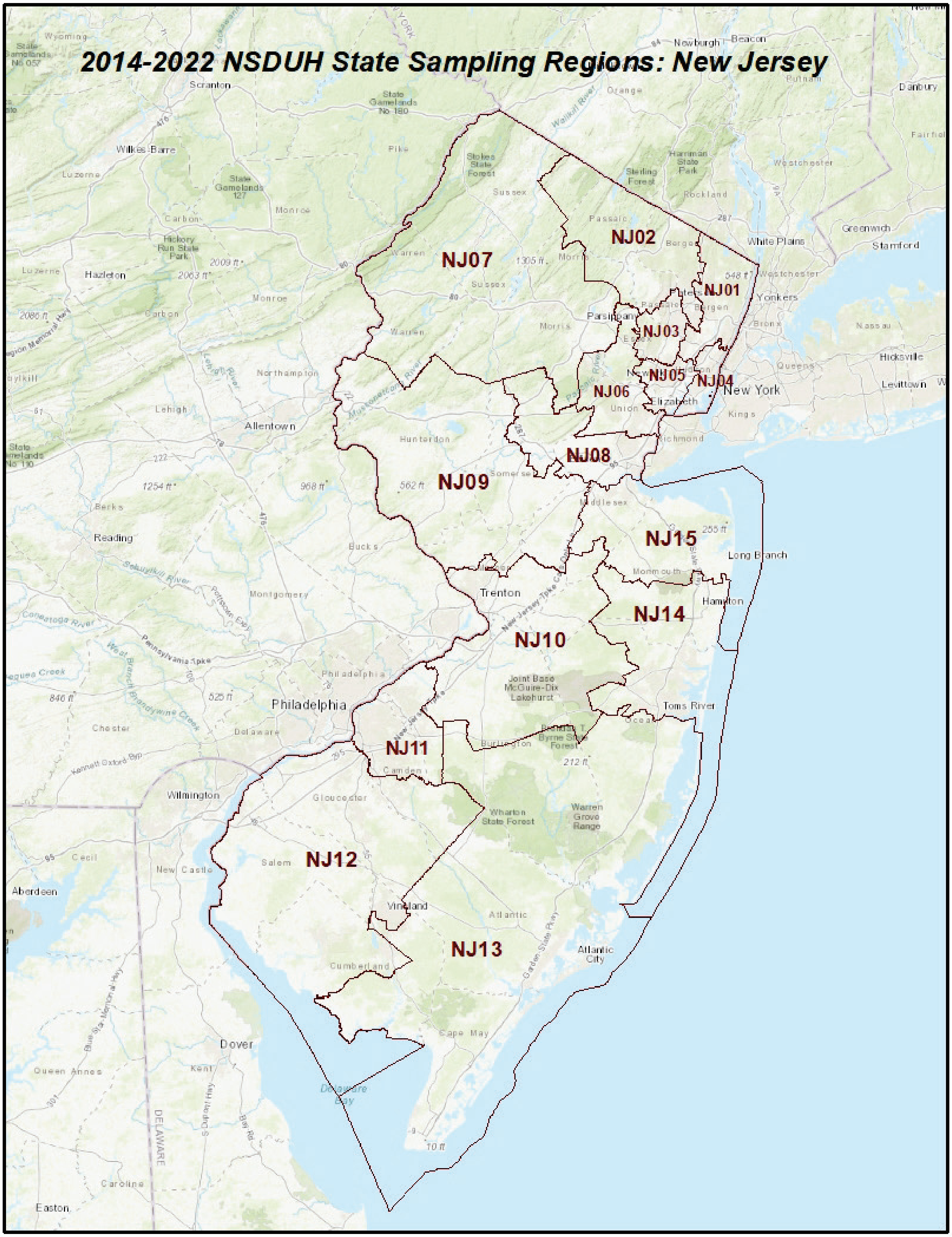

2014-2022 NSDUH State Sampling Regions: Missouri
MO12
MO11
MO10
MO06
MO05
MO04
MO03
MO09
MO02
MO07
MO01
MO08
Des Moines
Iowa City
Davenport
Muscatine
ILLINOIS
IOWA
Burlington
Galesburg
Peoria
Macomb
Bloomington
Kirksville
Quincy
Springfield
Chillicothe
St Joseph
Topeka
Lawrence
Columbia
Warrensburg
Sedalia
Jefferson City
MISSOURI
Rolla
Park Hills
Farmington
Fort Scott
Nevada
Lebanon
Fort Leonard Wood
Springfield
Joplin
Poplar Bluff
West Plains
OZARK PLATEAU
Rogers
Springdale
Harrison
Mountain Home
Fayetteville
Batesville
Jonesboro
Muskogee
Fort Smith
Russellville
Searcy
Conway
Memphis
OUACHITA MOUNTAINS
ARKANSAS

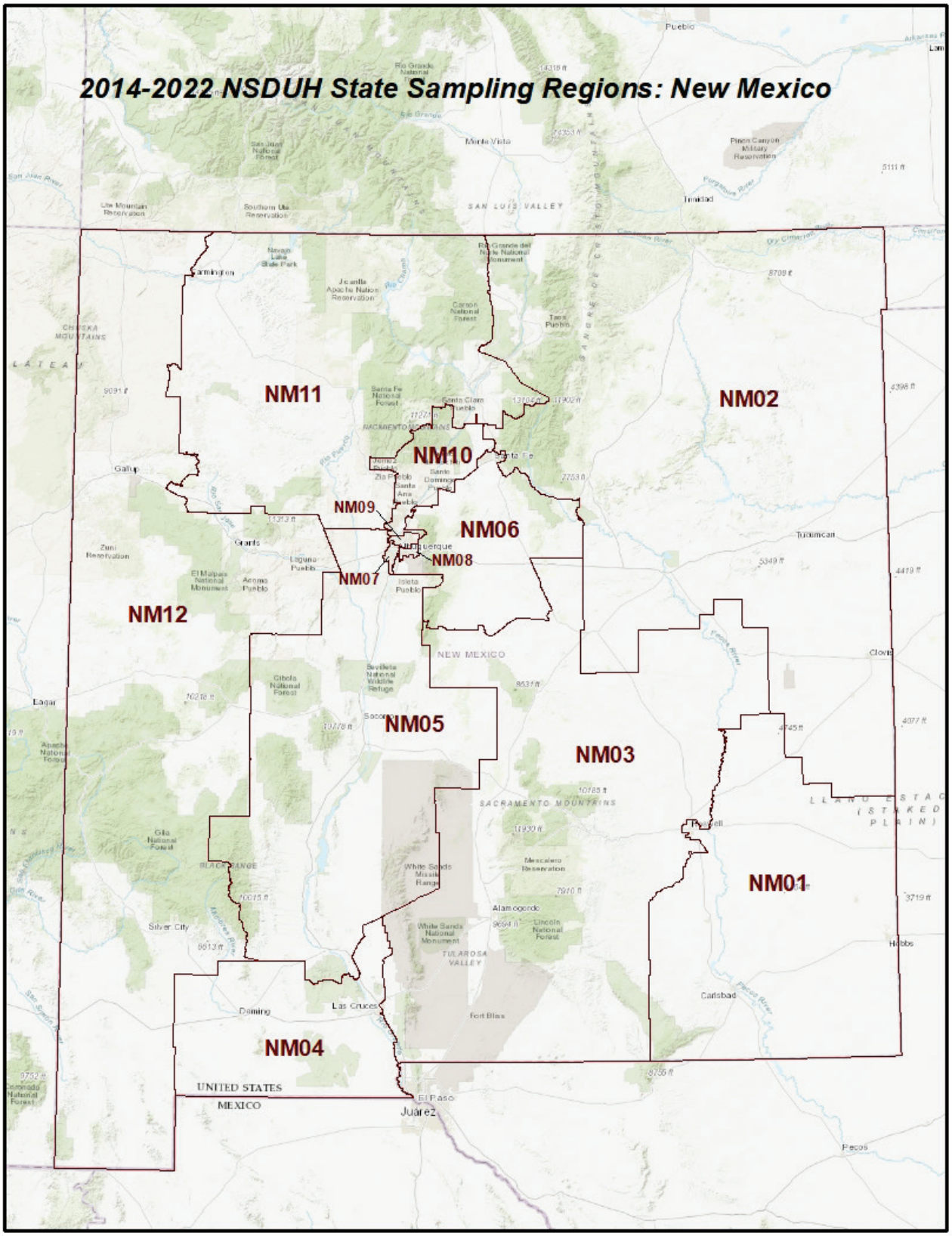

2014-2022 NSDUH State Sampling Regions: Montana
MT10
MT09
MT11
MT01
MT12
MT08
MT07
MT04
MT02
MT05
MT03
MT06
MONTANA
Kalispell
Missoula
Helena
Great Falls
Billings
Malta
Glasgow
Wolf Point
Sidney
Glendive
Gillette
Weyburn
Swift Current
Medicine Hat
Lethbridge
Cranbrook
Fernie
Kimberley
Blackfeet Indian Reservation
Fort Peck Indian Reservation
Flathead Reservation
Crow Reservation
Fort Belknap Reservation
Upper Missouri River Breaks National Monument
Charles Russell Nat'l Wildlife Ref
Glacier National Park
Yellowstone National Park
Flathead National Forest
Lewis & Clark National Forest
Deerlodge National Forest
Beaverhead National Forest
Gallatin National Forest
Custer National Forest
Bighorn National Forest
Shoshone National Forest
Targhee National Forest
Grand Teton National
Scapegoat Wilderness
Kootenai National Forest
St Joe National Forest
Clearwater National Forest
Nezperce National Forest
Bitterroot National Forest
Payette National Forest
Salmon National Forest
Challis National Forest
Boise National Forest
Sawtooth National Forest
CABINET MOUNTAINS
BITTERROOT RANGE
SALMON RIVER MOUNTAINS
ABSAROKA RANGE
IDAHO
MISSOURI

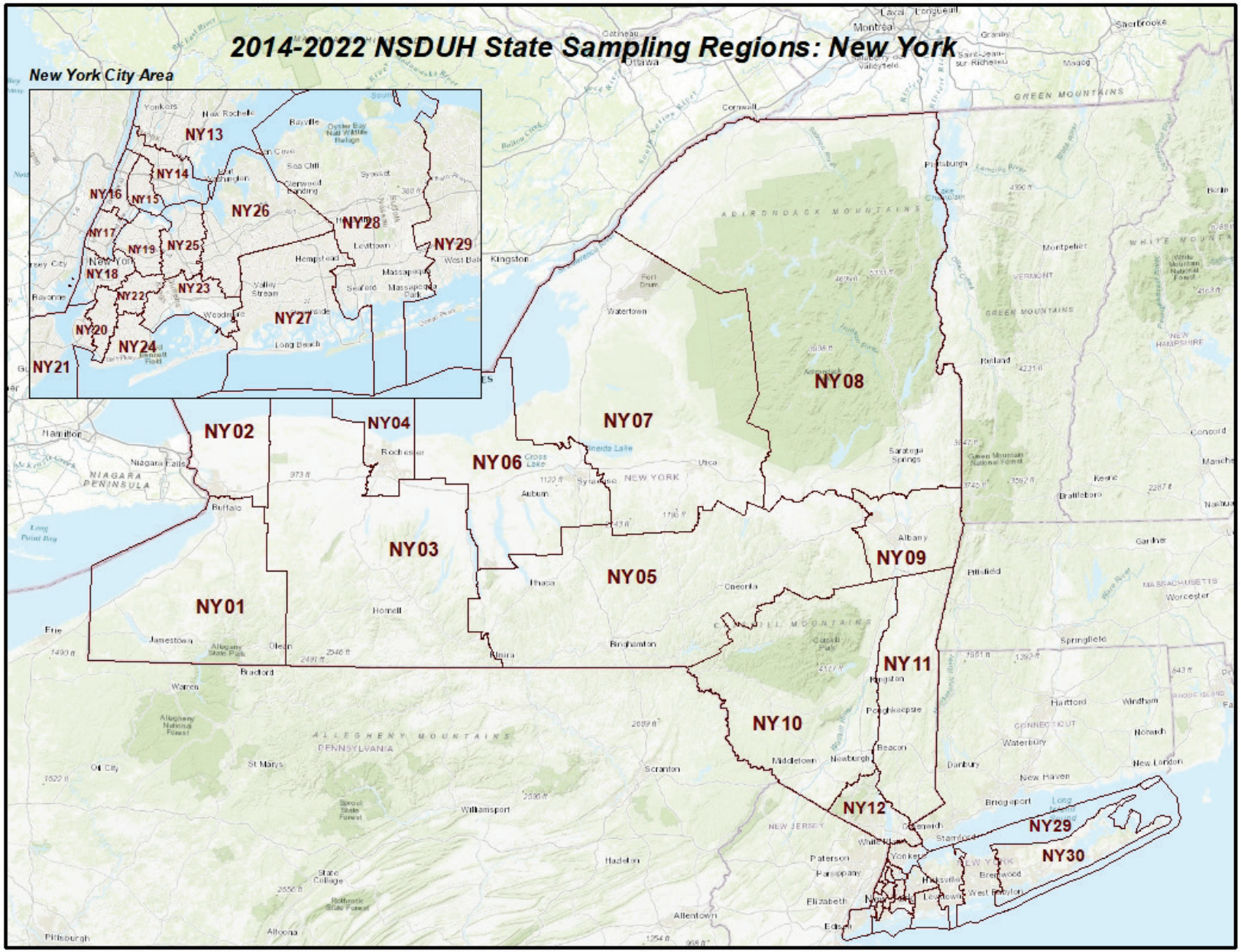

2014-2022 NSDUH State Sampling Regions: Nebraska
NE01
NE02
NE03
NE04
NE05
NE06
NE07
NE08
NE09
NE10
NE11
NE12
NEBRASKA
SAND HILLS
SMOKY HILLS
THE BADLANDS
Rapid City
Pierre
Mitchell
Sioux Falls
Yankton
Norfolk
Columbus
Grand Island
Hastings
Kearney
North Platte
McCook
Alliance
Chadron
Gering
Sidney
Fort Morgan
Goodland
Colby
Ainsworth
Valentine
Gregory
Omaha
Lincoln
Manhattan

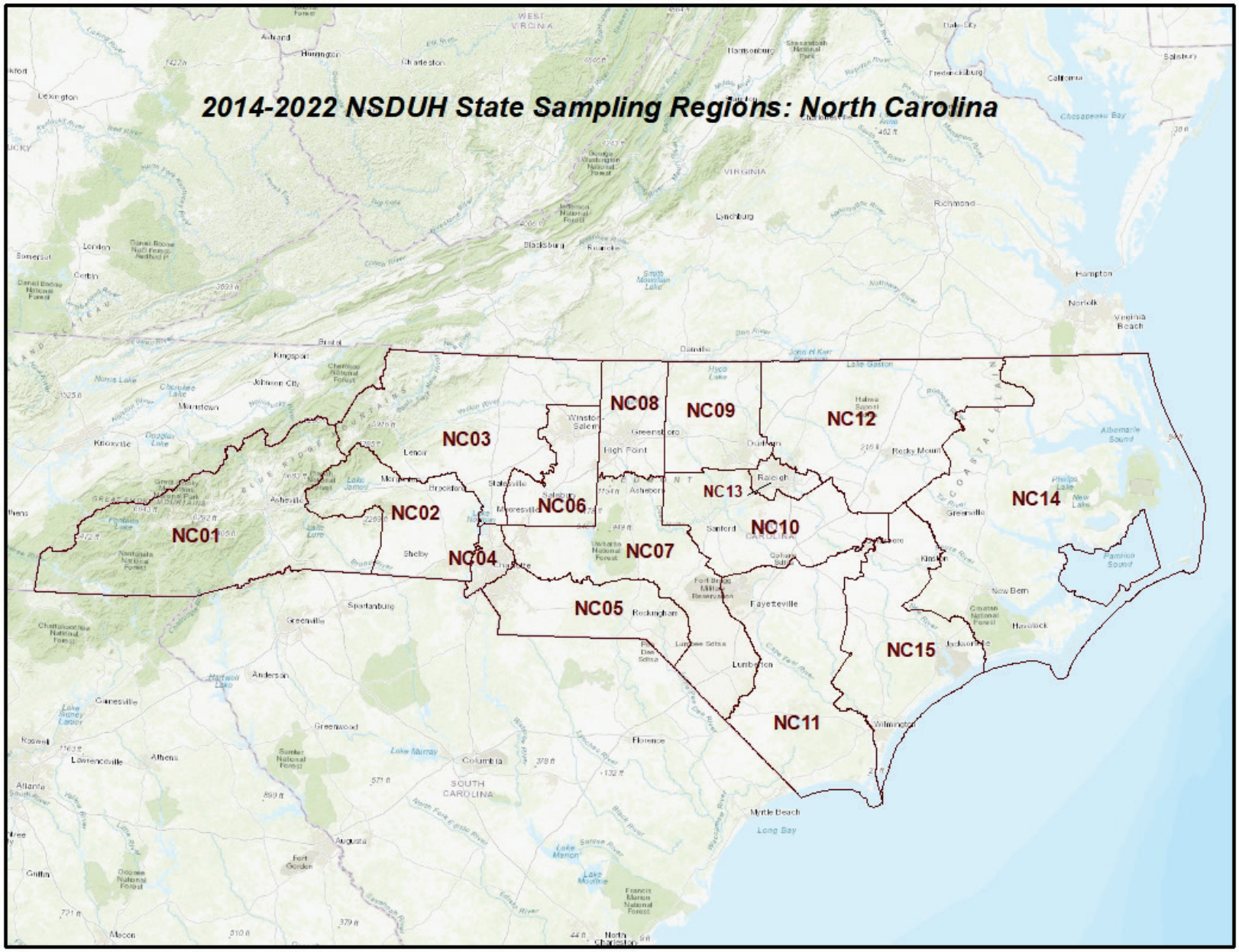

2014-2022 NSDUH State Sampling Regions: Nevada
NV01
NV02
NV03
NV04
NV05
NV12
Las Vegas Area
NV09
NV12
NV08
NV07
NV06
NV11
NV10
NV04
NV05
North Las Vegas
Las Vegas
Spring Valley
Henderson

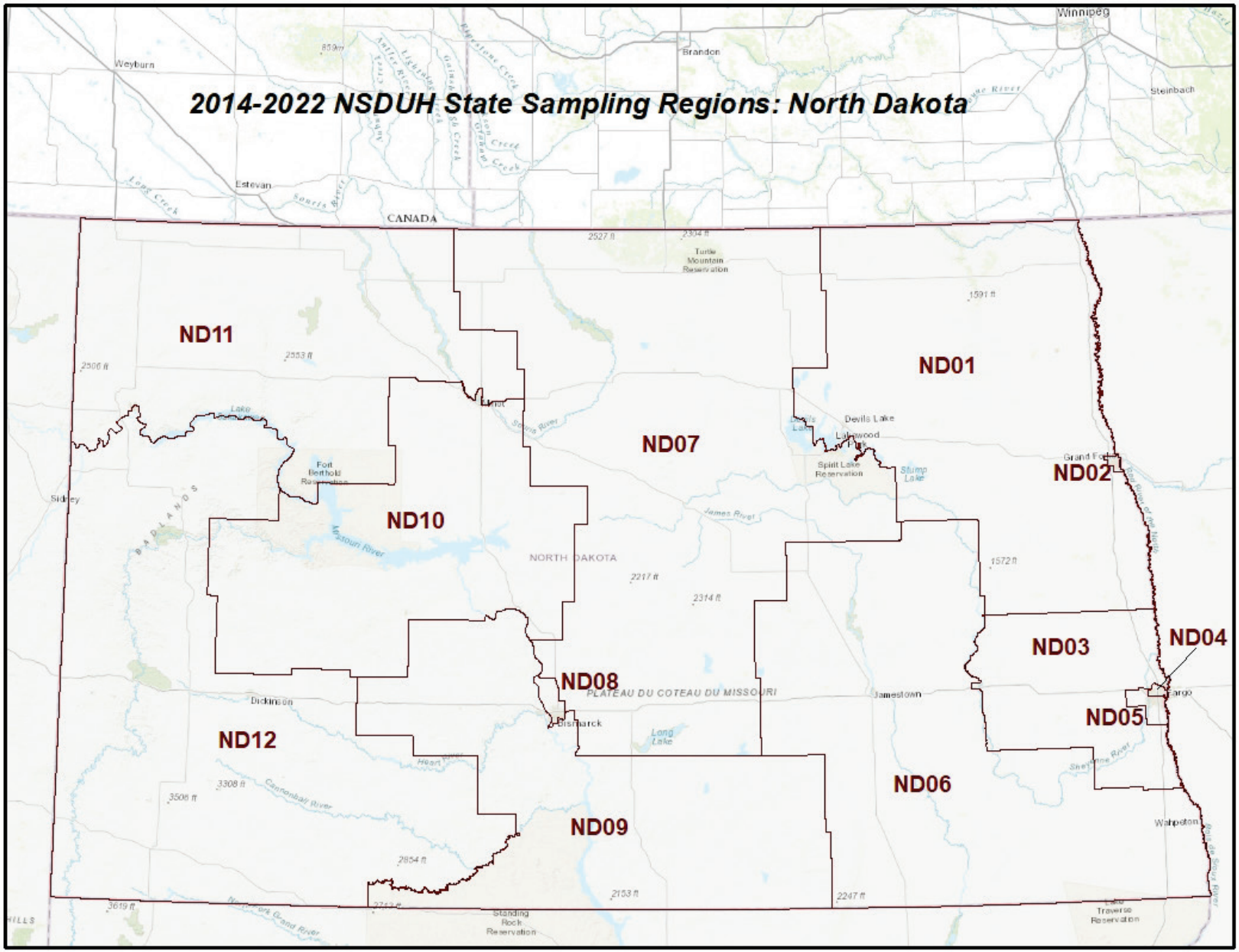

2014-2022 NSDUH State Sampling Regions: New Hampshire
NH01
NH02
NH03
NH04
NH05
NH06
NH07
NH08
NH09
NH10
NH11
NH12

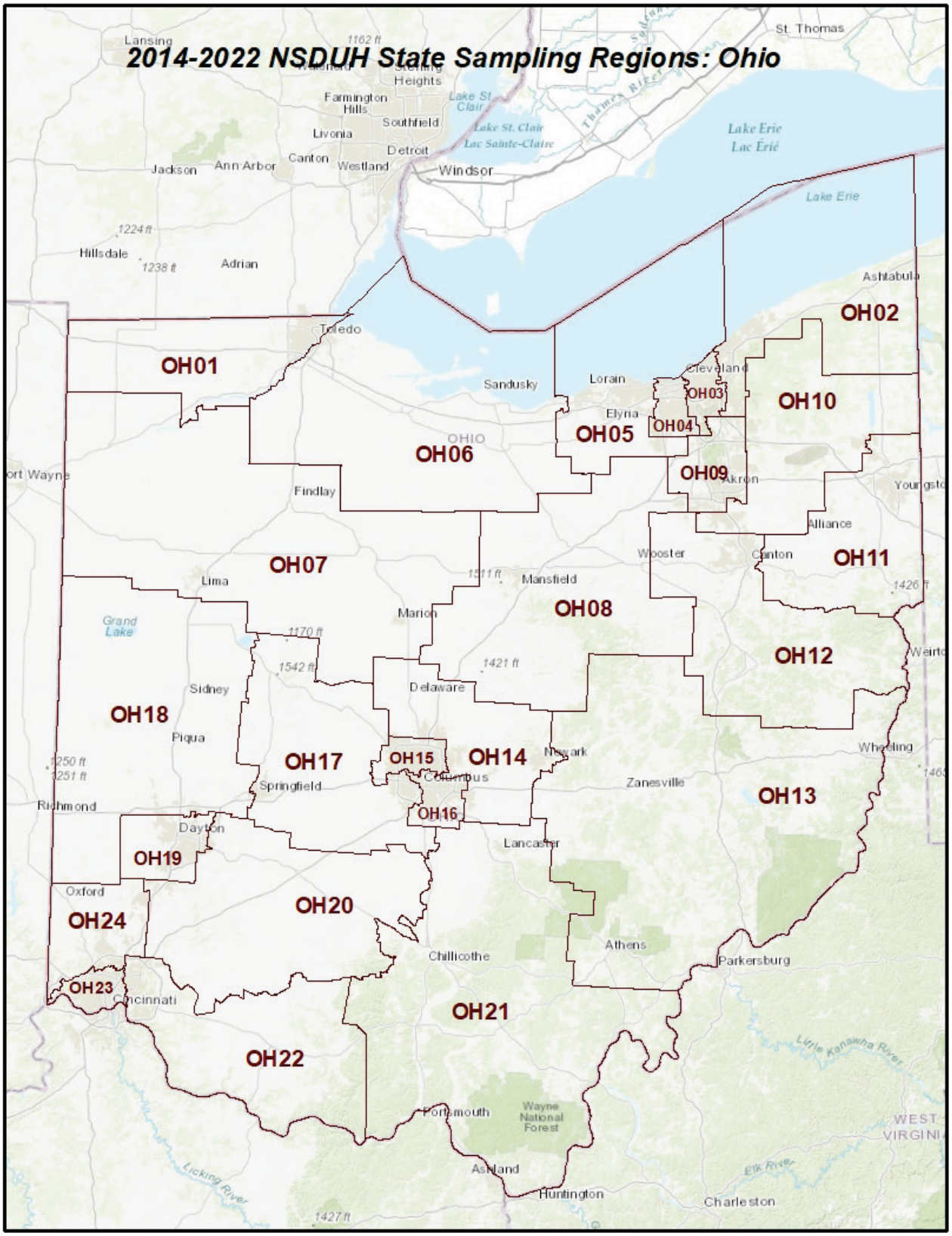

2014-2022 NSDUH State Sampling Regions: New Jersey
NJ01
NJ02
NJ03
NJ04
NJ05
NJ06
NJ07
NJ08
NJ09
NJ10
NJ11
NJ12
NJ13
NJ14
NJ15
Scranton
Wilkes-Barre
Hazleton
Allentown
Reading
Philadelphia
Wilmington
Trenton
New York
Yonkers
White Plains
Stamford
Long Branch
Toms River
Atlantic City
Vineland
Dover
Aberdeen
Easton
Newburgh
Beacon
Danbury

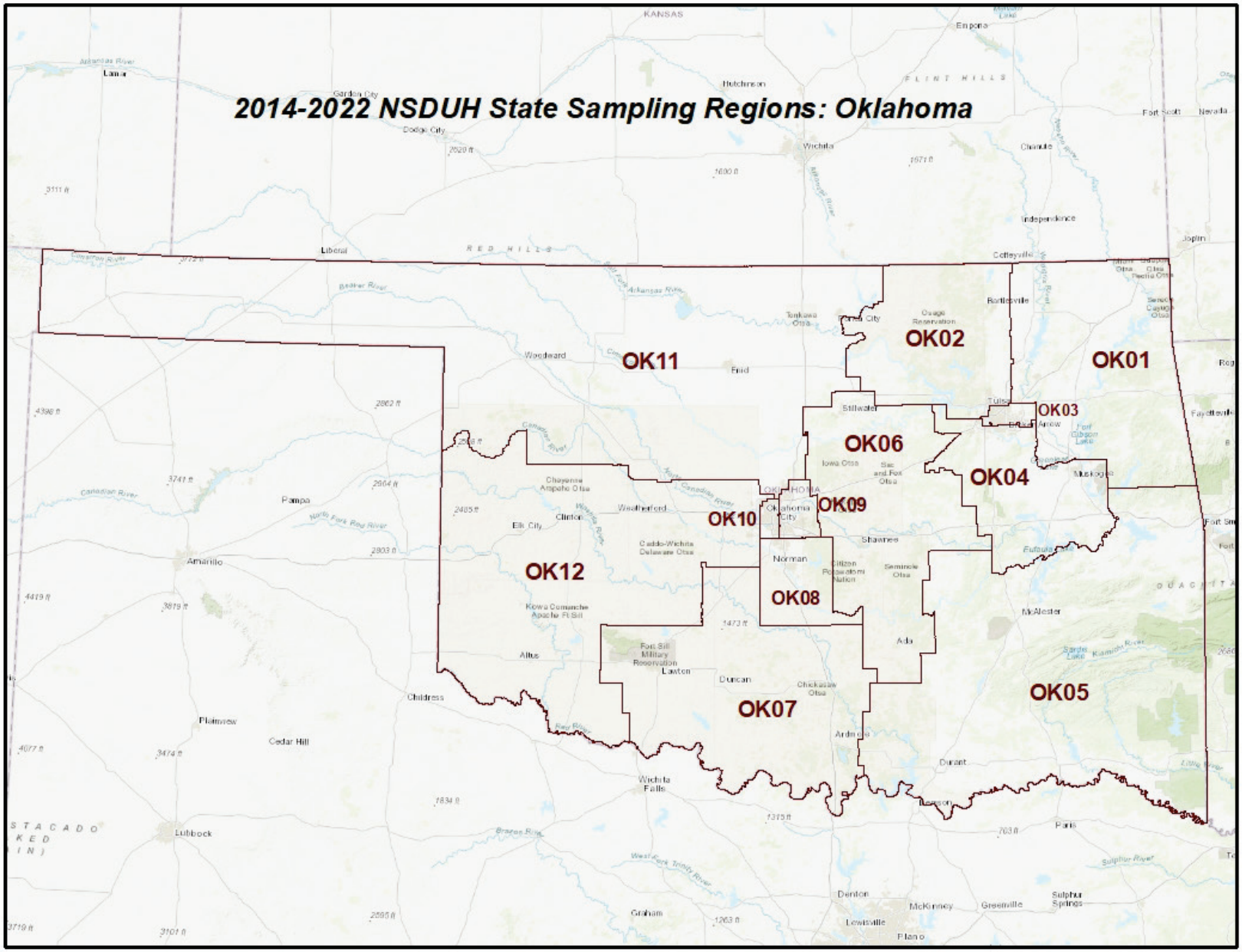

2014-2022 NSDUH State Sampling Regions: New Mexico
NM11
NM02
NM10
NM09
NM06
NM08
NM07
NM12
NM05
NM03
NM01
NM04
NEW MEXICO
SAN LUIS VALLEY
SACRAMENTO MOUNTAINS
TULAROSA VALLEY
UNITED STATES
MEXICO
Pueblo
Trinidad
Farmington
Gallup
Grants
Albuquerque
Santa Fe
Taos
Tucumcari
Clovis
Roswell
Hobbs
Carlsbad
Las Cruces
Deming
Silver City
El Paso
Juárez
Alamogordo
Ruidoso
Pecos
Eagar

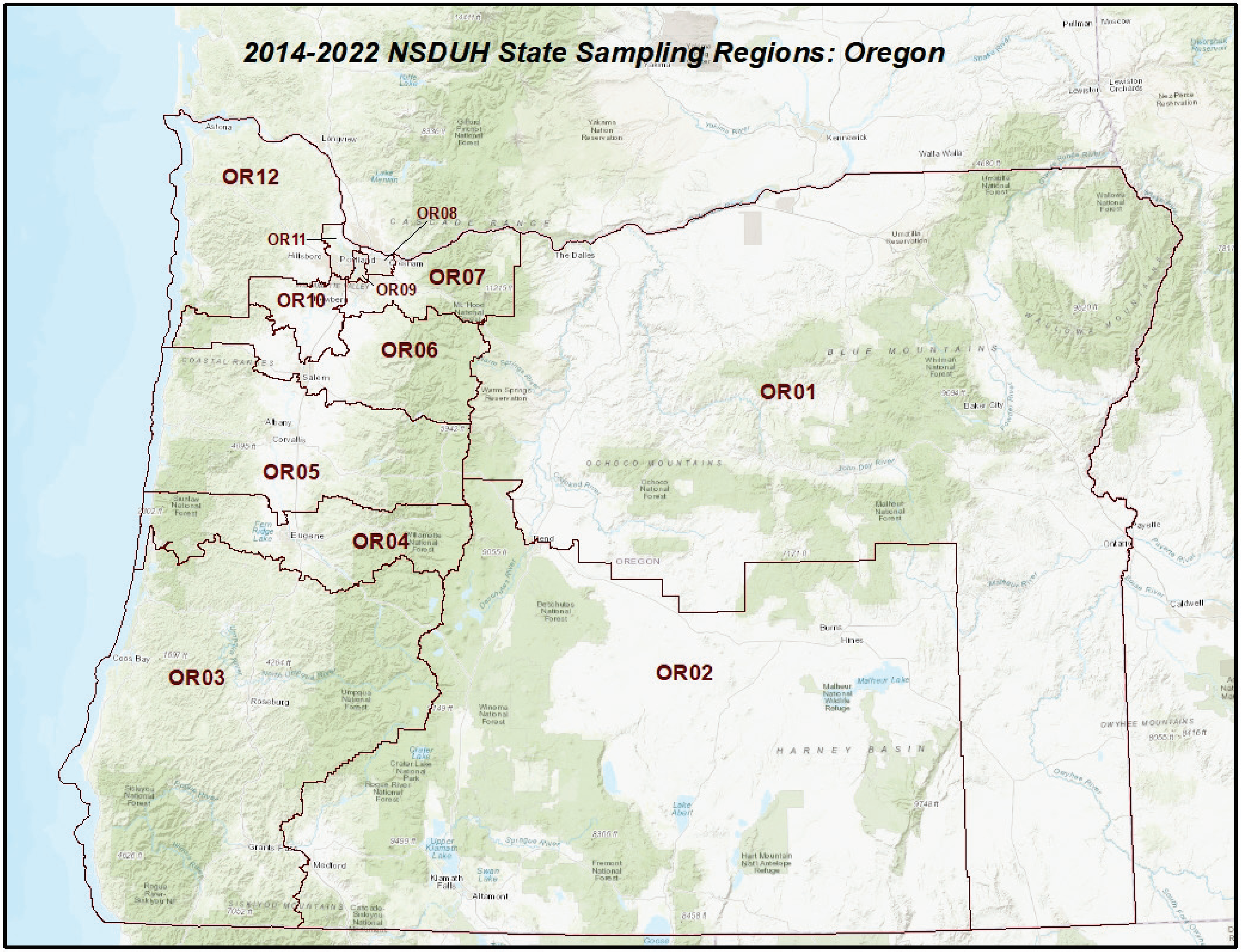

2014-2022 NSDUH State Sampling Regions: New York
New York City Area
NY01
NY02
NY03
NY04
NY05
NY06
NY07
NY08
NY09
NY10
NY11
NY12
NY13
NY14
NY15
NY16
NY17
NY18
NY19
NY20
NY21
NY22
NY23
NY24
NY25
NY26
NY27
NY28
NY29
NY30

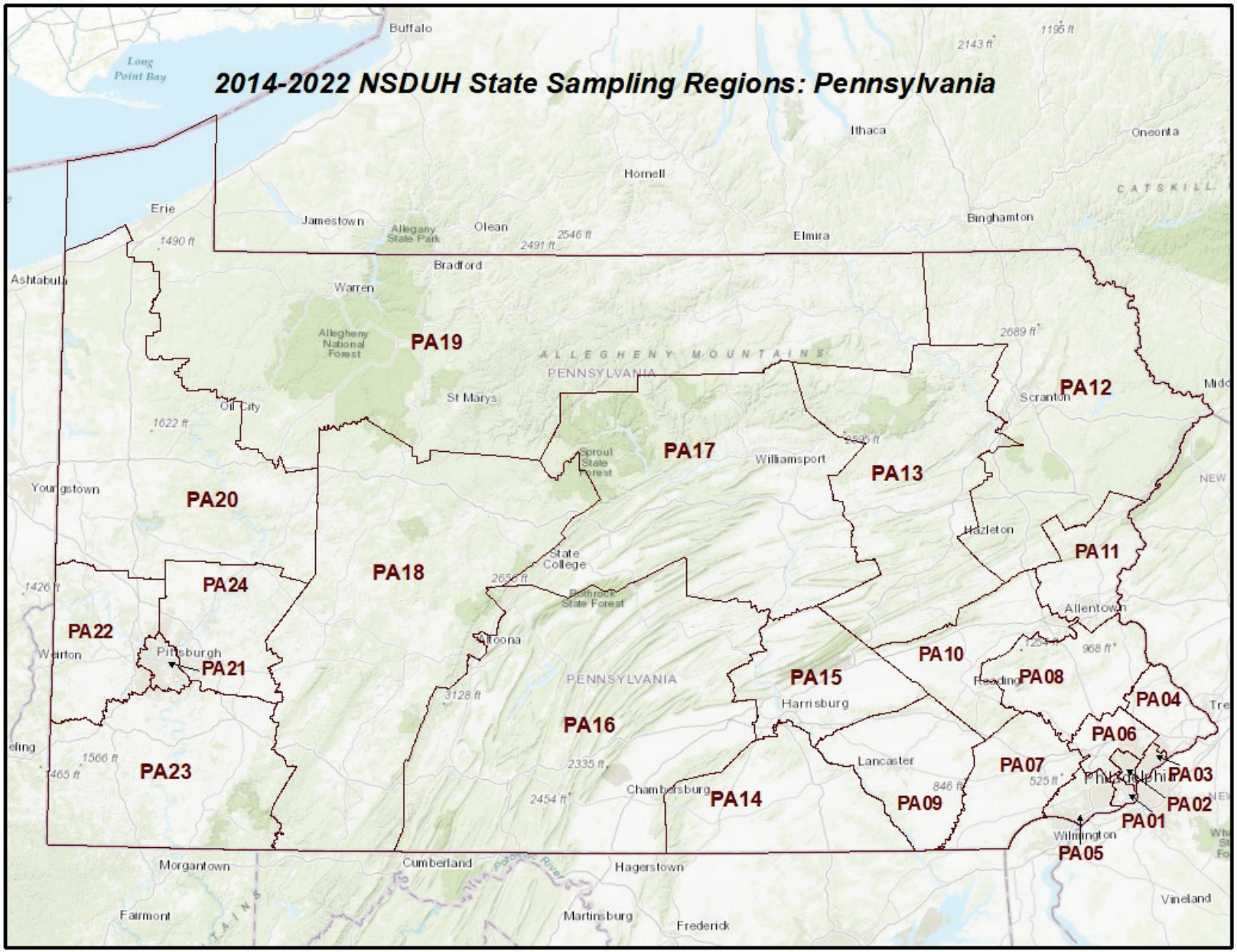

2014-2022 NSDUH State Sampling Regions: North Carolina
NC01
NC02
NC03
NC04
NC05
NC06
NC07
NC08
NC09
NC10
NC11
NC12
NC13
NC14
NC15

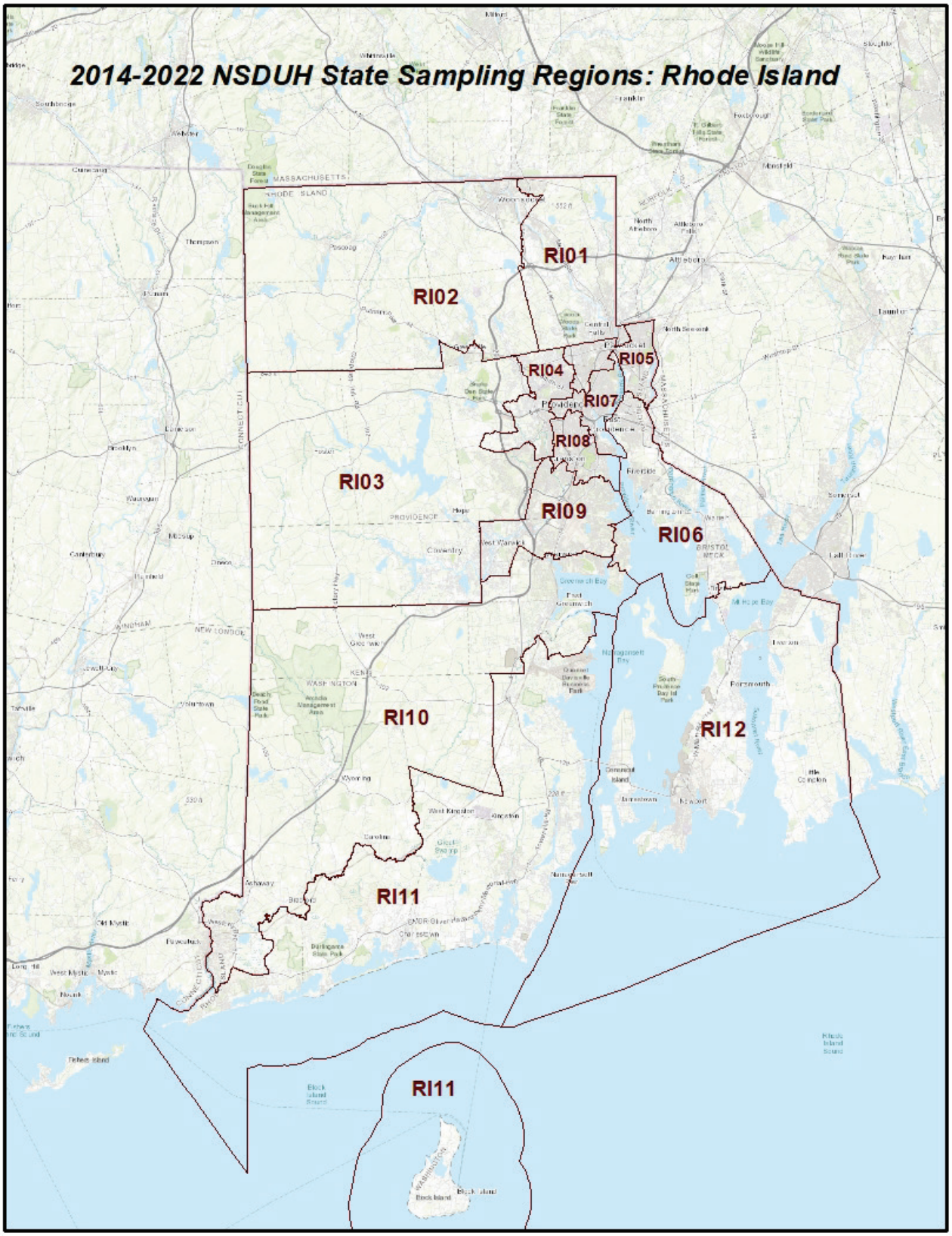

2014-2022 NSDUH State Sampling Regions: North Dakota
ND11
ND01
ND07
ND02
ND10
ND04
ND03
ND08
ND05
ND12
ND06
ND09
CANADA
NORTH DAKOTA
PLATEAU DU COTEAU DU MISSOURI
Winnipeg
Brandon
Weyburn
Estevan
Steinbach
Turtle Mountain Reservation
Devils Lake
Spirit Lake Reservation
Fort Berthold Reservation
Sidney
Dickinson
Bismarck
Jamestown
Fargo
Wahpeton
Standing Rock Reservation
Lake Traverse Reservation

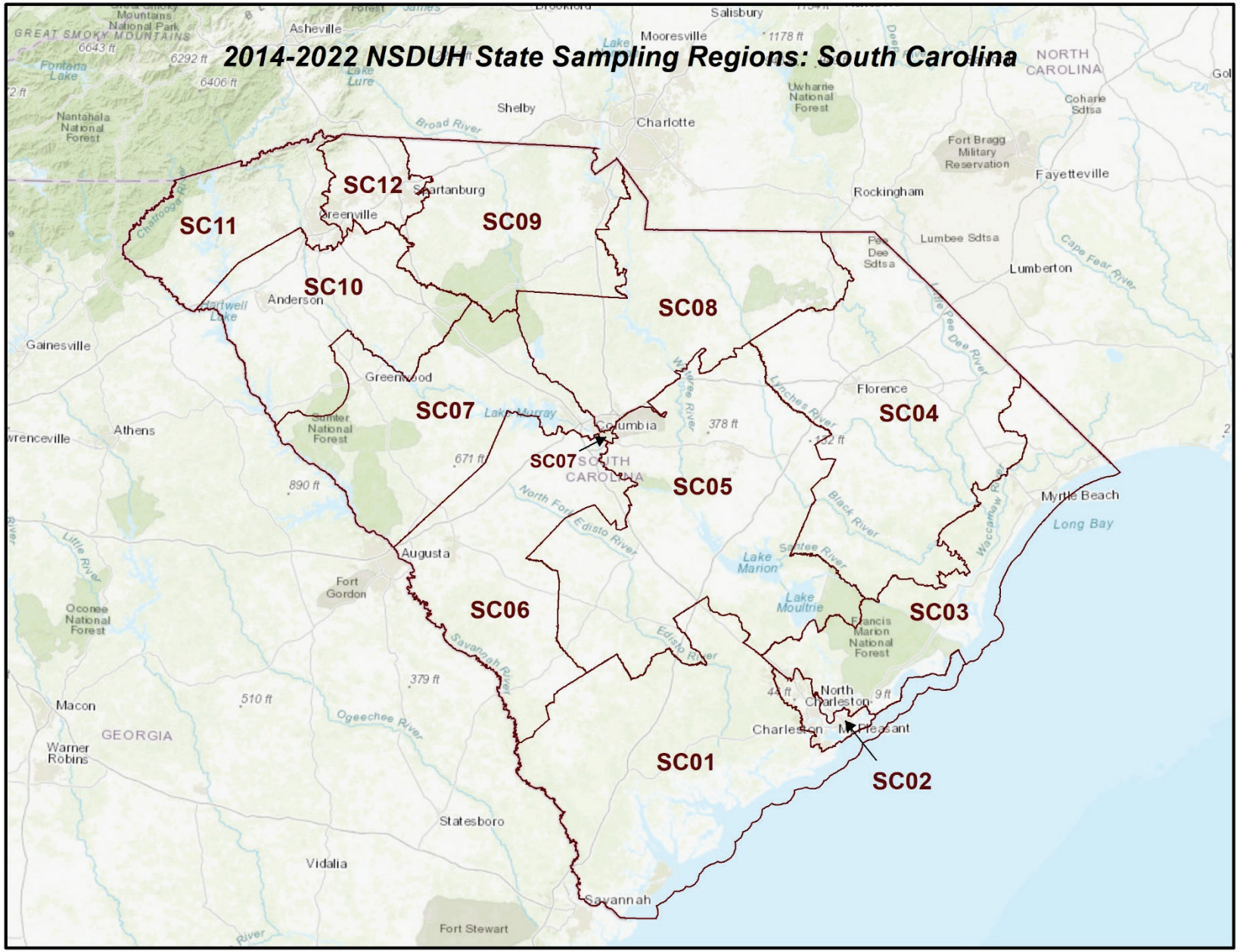

2014-2022 NSDUH State Sampling Regions: Ohio
OH01
OH02
OH03
OH04
OH05
OH06
OH07
OH08
OH09
OH10
OH11
OH12
OH13
OH14
OH15
OH16
OH17
OH18
OH19
OH20
OH21
OH22
OH23
OH24
Lansing
Farmington Hills
Southfield
Livonia
Detroit
Canton
Westland
Jackson
Ann Arbor
Windsor
St. Thomas
Lake St. Clair
Lac Sainte-Claire
Lake Erie
Lac Érié
Hillsdale
Adrian
Toledo
Sandusky
Lorain
Cleveland
Elyria
Akron
Findlay
Lima
Mansfield
Marion
Grand Lake
Wooster
Canton
Alliance
Sidney
Delaware
Piqua
Springfield
Columbus
Newark
Zanesville
Wheeling
Richmond
Dayton
Lancaster
Oxford
Chillicothe
Athens
Parkersburg
Cincinnati
Portsmouth
Wayne National Forest
Ashland
Huntington
Charleston
Licking River
Elk River
Lyrle Kanawha River
WEST VIRGINIA
OHIO

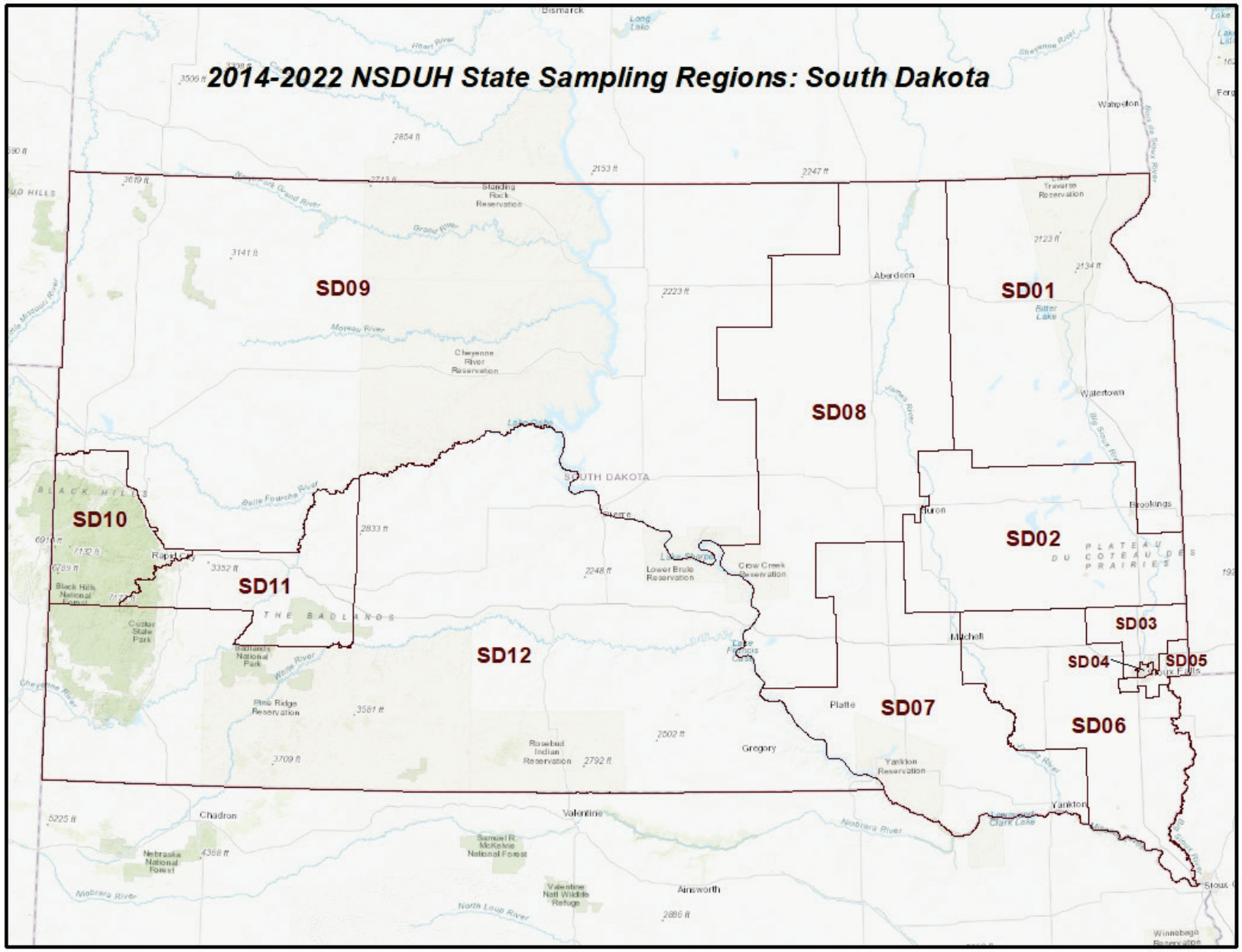

2014-2022 NSDUH State Sampling Regions: Oklahoma
OK01
OK02
OK03
OK04
OK05
OK06
OK07
OK08
OK09
OK10
OK11
OK12
KANSAS
Liberal
Dodge City
Garden City
Hutchinson
Wichita
Emporia
Independence
Coffeyville
Bartlesville
Stillwater
Enid
Woodward
Pampa
Amarillo
Clinton
Elk City
Weatherford
Oklahoma City
Norman
Shawnee
McAlester
Ada
Lawton
Altus
Duncan
Childress
Plainview
Cedar Hill
Lubbock
Wichita Falls
Durant
Paris
Denton
McKinney
Greenville
Sulphur Springs
Lewisville
Plano
Graham
Fort Scott
Joplin
Fayetteville
Muskogee
Tulsa

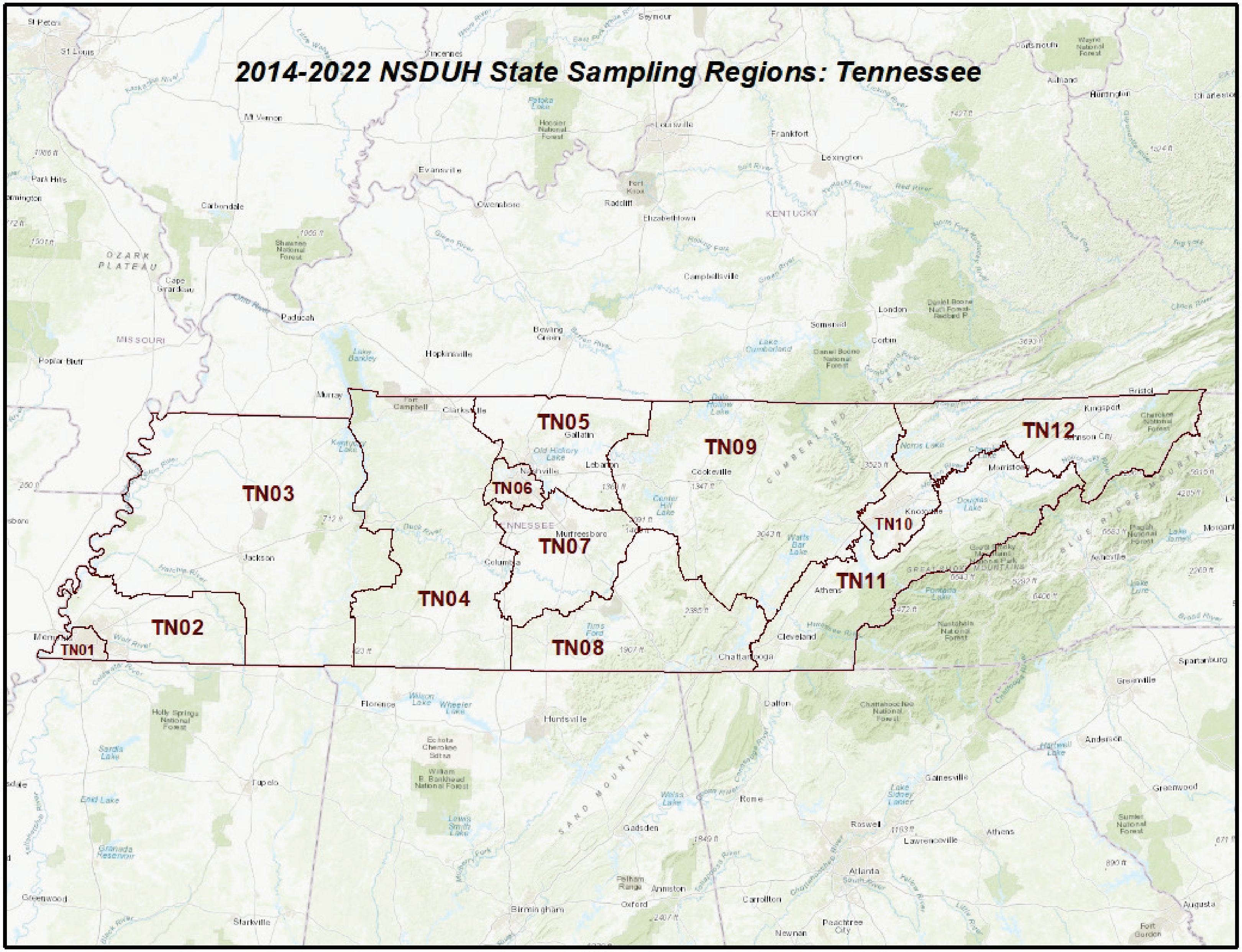

2014-2022 NSDUH State Sampling Regions: Oregon
OR12
OR08
OR11
OR07
OR09
OR10
OR06
OR01
OR05
OR04
OR03
OR02

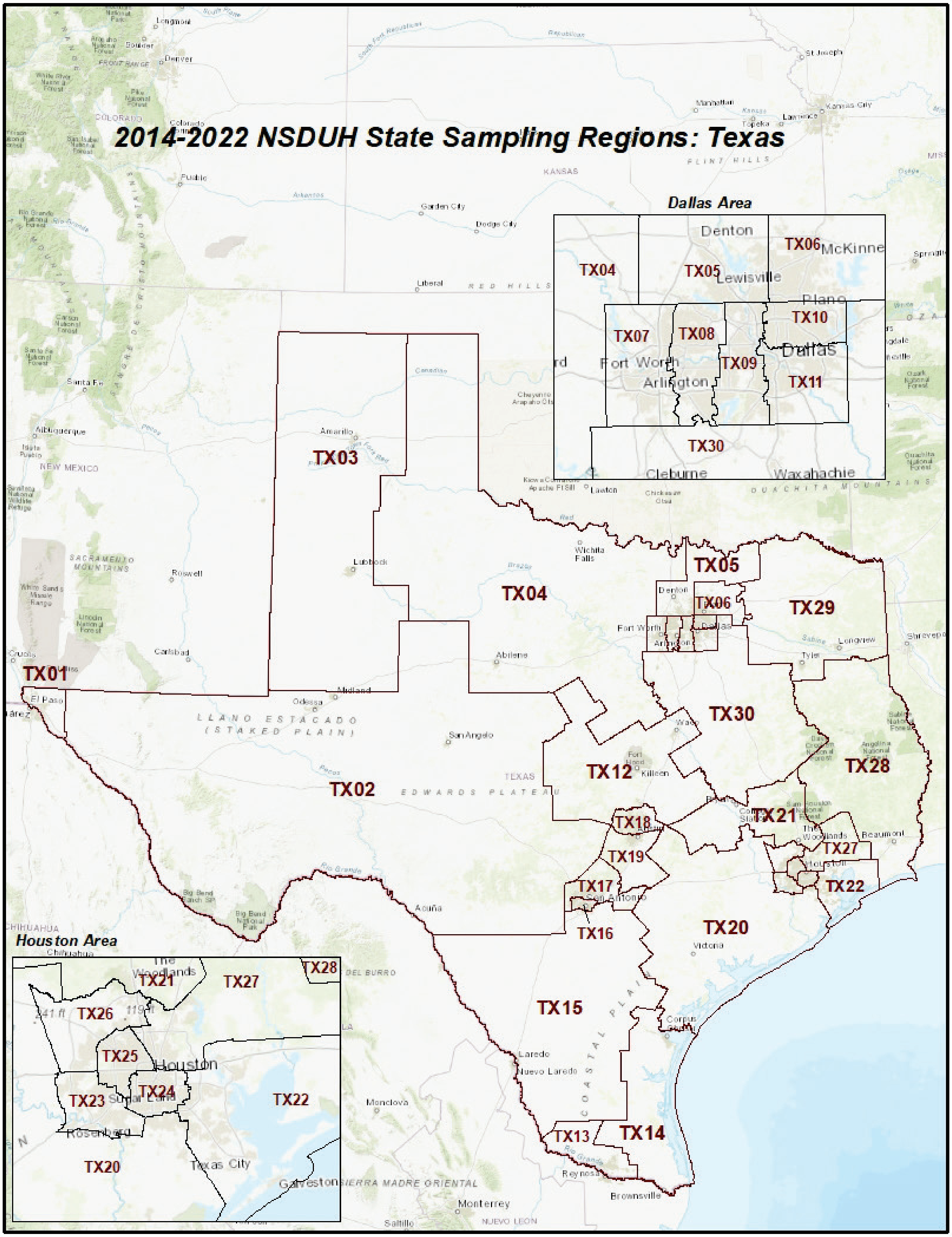

2014-2022 NSDUH State Sampling Regions: Pennsylvania
PA01
PA02
PA03
PA04
PA05
PA06
PA07
PA08
PA09
PA10
PA11
PA12
PA13
PA14
PA15
PA16
PA17
PA18
PA19
PA20
PA21
PA22
PA23
PA24
Buffalo
Erie
Jamestown
Olean
Bradford
Warren
Ithaca
Hornell
Elmira
Binghamton
Oneonta
Scranton
Williamsport
Harrisburg
Lancaster
Reading
Allentown
Hazleton
Wilmington
Philadelphia
Pittsburgh
Altoona
State College
Chambersburg
Cumberland
Hagerstown
Martinsburg
Frederick
Morgantown
Fairmont
Oil City
St Marys
Youngstown
Ashtabula
Allegheny National Forest
Allegany State Park
Sproul State Forest
Bald Eagle State Forest
PENNSYLVANIA
ALLEGHENY MOUNTAINS
CATSKILL
Vineland

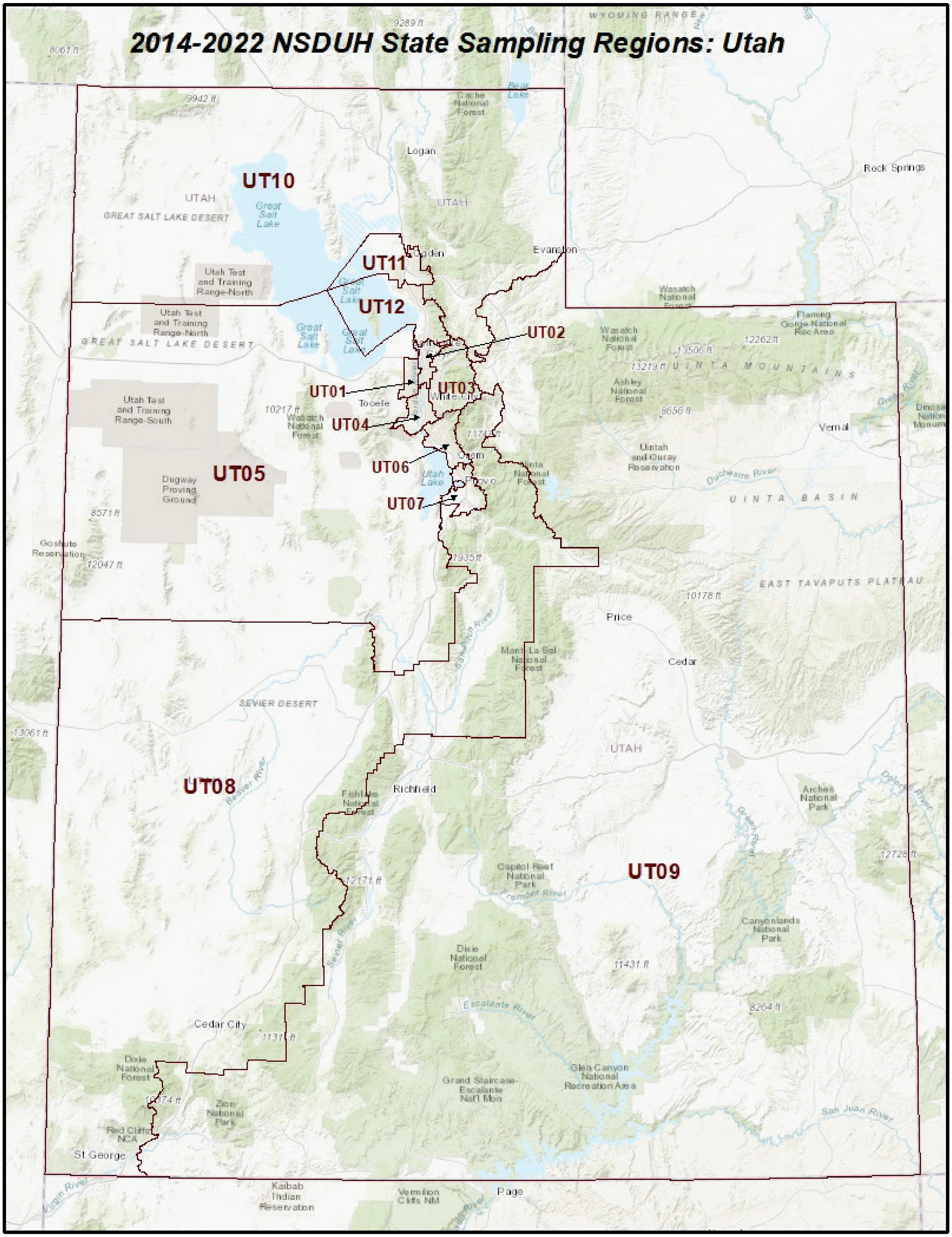

2014-2022 NSDUH State Sampling Regions: Rhode Island
RI01
RI02
RI03
RI04
RI05
RI06
RI07
RI08
RI09
RI10
RI11
RI12
RI11

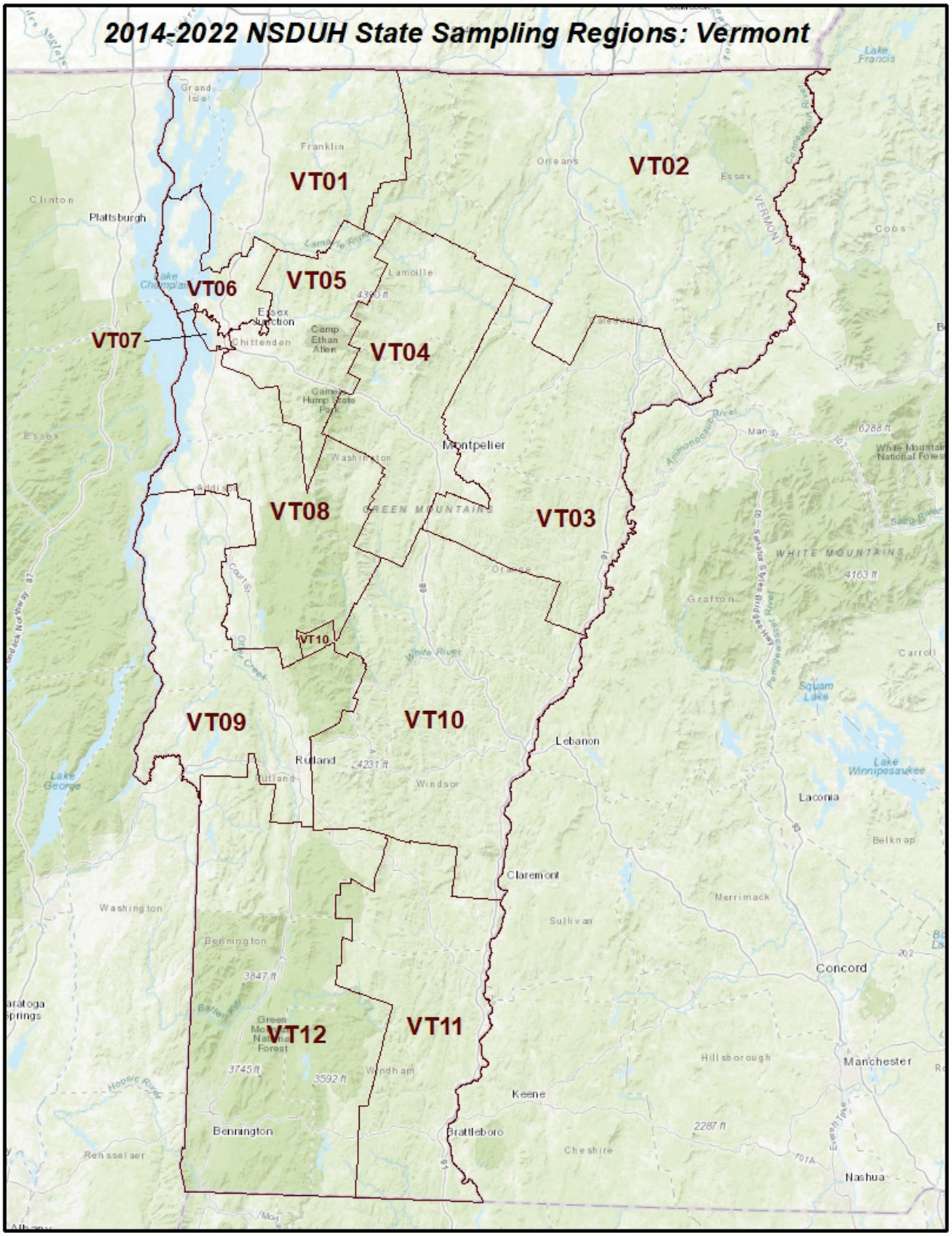

2014-2022 NSDUH State Sampling Regions: South Carolina
SC01
SC02
SC03
SC04
SC05
SC06
SC07
SC08
SC09
SC10
SC11
SC12
SC07
Asheville
Salisbury
Mooresville
Charlotte
Shelby
Spartanburg
Greenville
Anderson
Rockingham
Fayetteville
Lumberton
Florence
Columbia
Greenwood
Gainesville
Athens
Augusta
Fort Gordon
Macon
Warner Robins
Statesboro
Vidalia
Savannah
Fort Stewart
Myrtle Beach
North Charleston
Charleston
Mt Pleasant
Long Bay
NORTH CAROLINA
SOUTH CAROLINA
GEORGIA
Lake Murray
Lake Marion
Lake Moultrie
Francis Marion National Forest
Sumter National Forest
Oconee National Forest
Nantahala National Forest
Uwharrie National Forest
Fort Bragg Military Reservation
Pee Dee Sdtsa
Lumbee Sdtsa
Coharie Sdtsa
Great Smoky Mountains National Park
GREAT SMOKY MOUNTAINS
Broad River
Cape Fear River
Little Pee Dee River
Lynches River
Black River
Santee River
Waccamaw River
Edisto River
North Fork Edisto River
Savannah River
Ogeechee River
Little River
Hartwell Lake
Lake Keowee
Lake Jocassee
Lake Lure
Fontana Lake
6643 ft
6292 ft
6406 ft
1178 ft
671 ft
890 ft
510 ft
379 ft
378 ft
122 ft

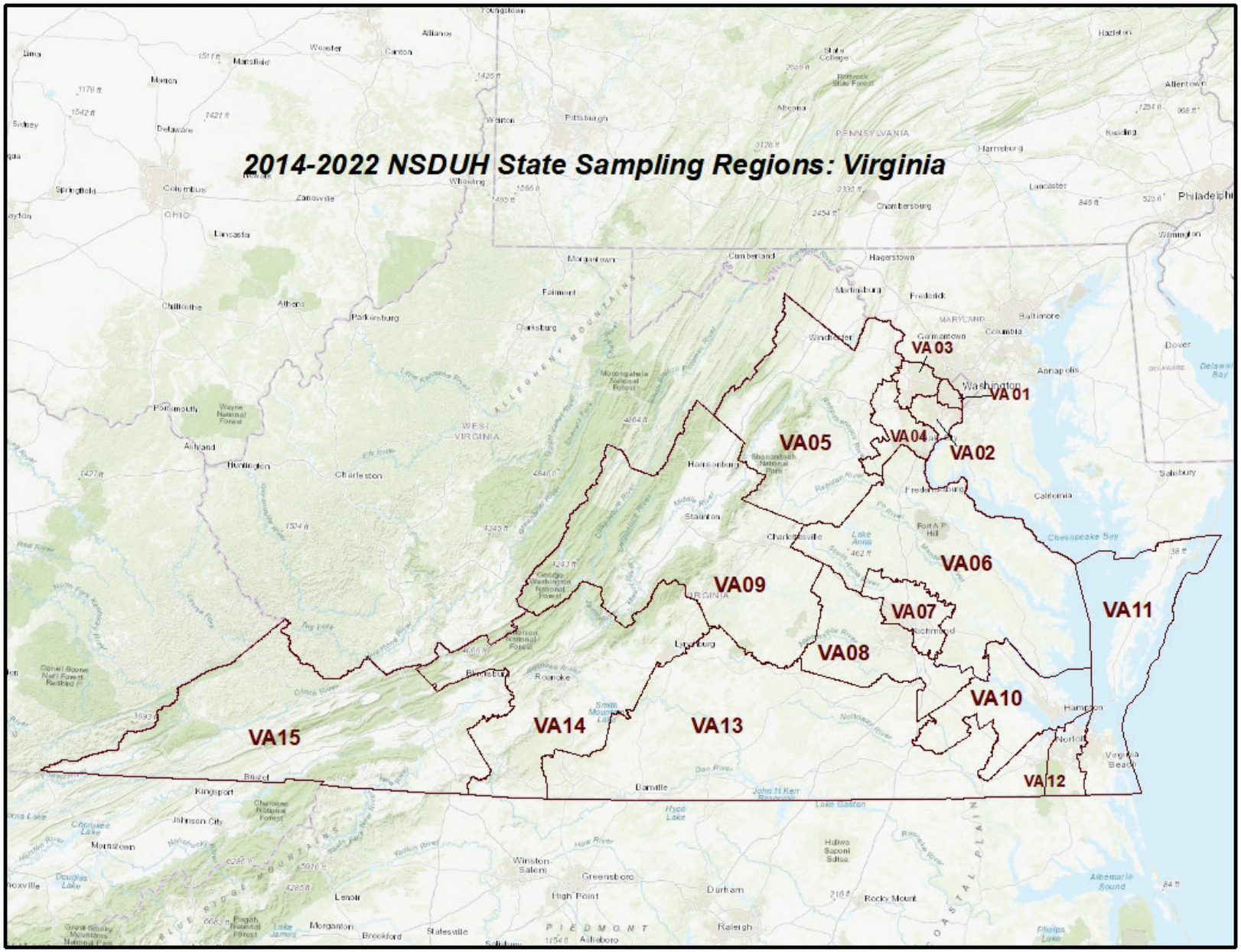

2014-2022 NSDUH State Sampling Regions: South Dakota
SD09
SD01
SD08
SD10
SD02
SD11
SD03
SD12
SD04
SD05
SD07
SD06
Bismarck
Aberdeen
Watertown
Huron
Brookings
Mitchell
Platte
Gregory
Chadron
Valentine
Yankton
Sioux Falls
Rapid City
Standing Rock Reservation
Cheyenne River Reservation
Lower Brule Reservation
Crow Creek Reservation
Pine Ridge Reservation
Rosebud Indian Reservation
Yankton Reservation
SOUTH DAKOTA
BLACK HILLS
THE BADLANDS
PLATEAU DU COTEAU DES PRAIRIES
Grand River
Moreau River
Belle Fourche River
White River
Niobrara River
Missouri River
James River
Big Sioux River
Lake Oahe
Lake Francis Case
Lewis and Clark Lake
Bitter Lake
Long Lake
Custer State Park
Badlands National Park
Black Hills National Forest
Nebraska National Forest
Samuel R. McKelvie National Forest
Valentine Natl Wildlife Refuge
Ainsworth
Winnebago Reservation

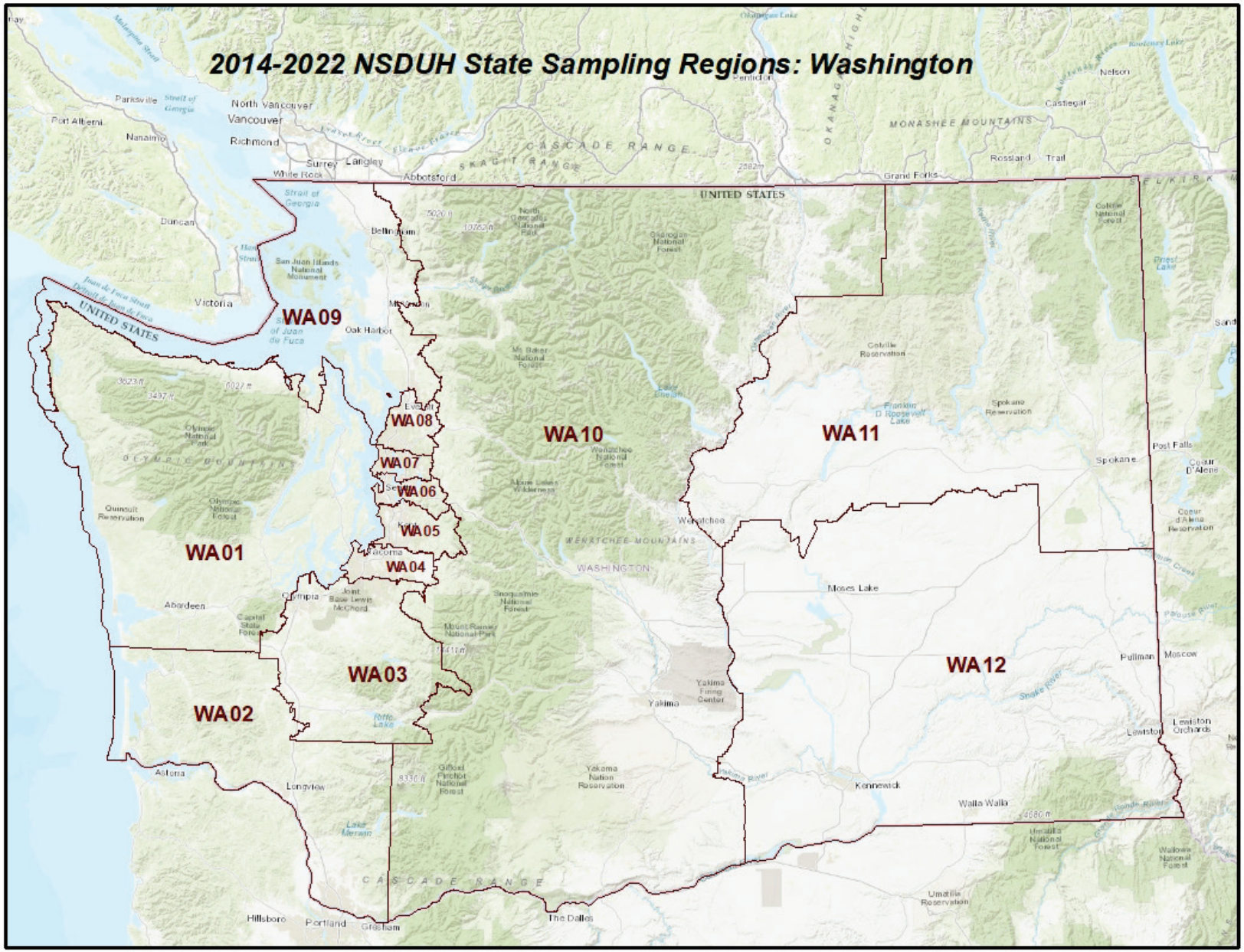

2014-2022 NSDUH State Sampling Regions: Tennessee
TN01
TN02
TN03
TN04
TN05
TN06
TN07
TN08
TN09
TN10
TN11
TN12

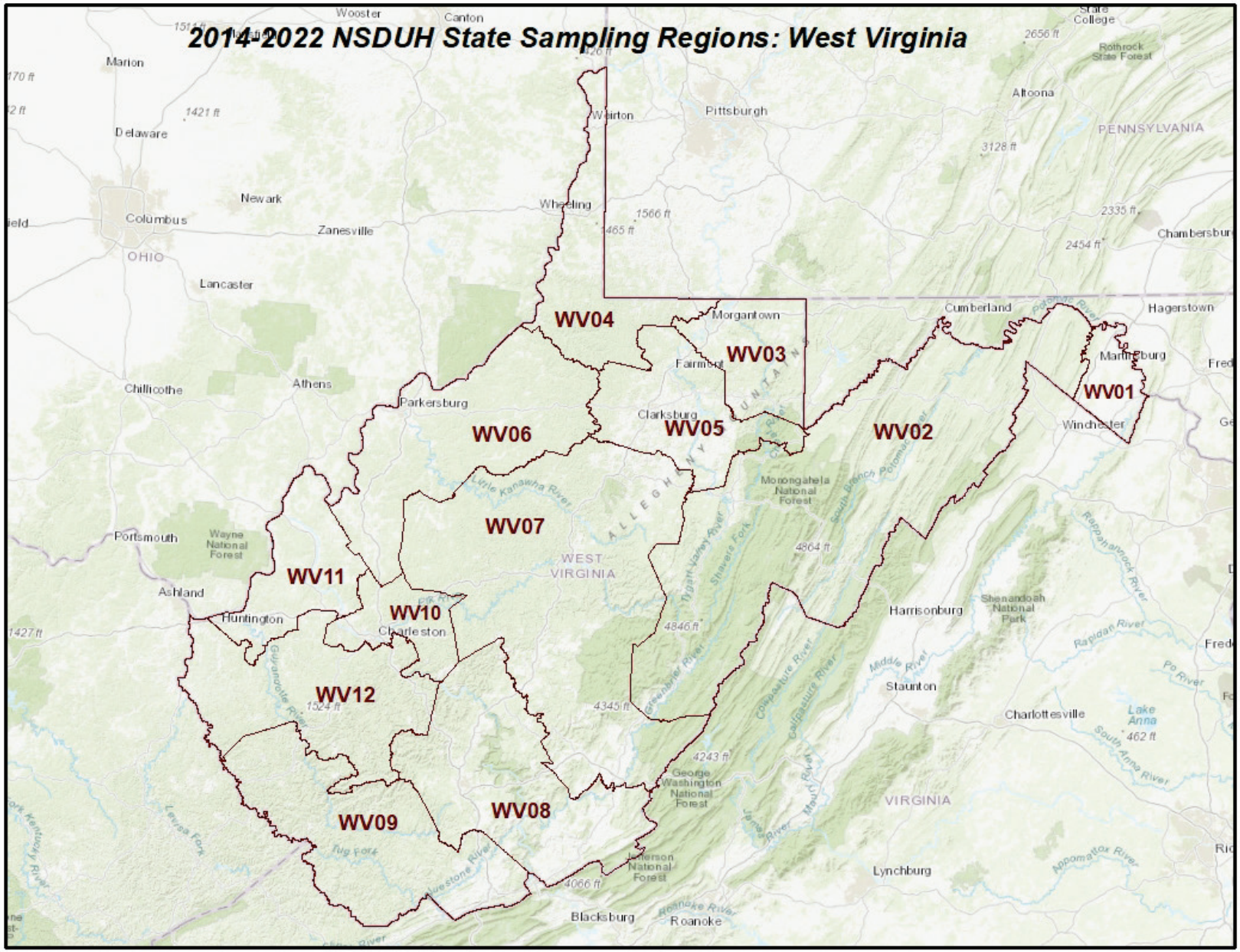

2014-2022 NSDUH State Sampling Regions: Texas
Dallas Area
Houston Area
TX01
TX02
TX03
TX04
TX05
TX06
TX07
TX08
TX09
TX10
TX11
TX12
TX13
TX14
TX15
TX16
TX17
TX18
TX19
TX20
TX21
TX22
TX23
TX24
TX25
TX26
TX27
TX28
TX29
TX30

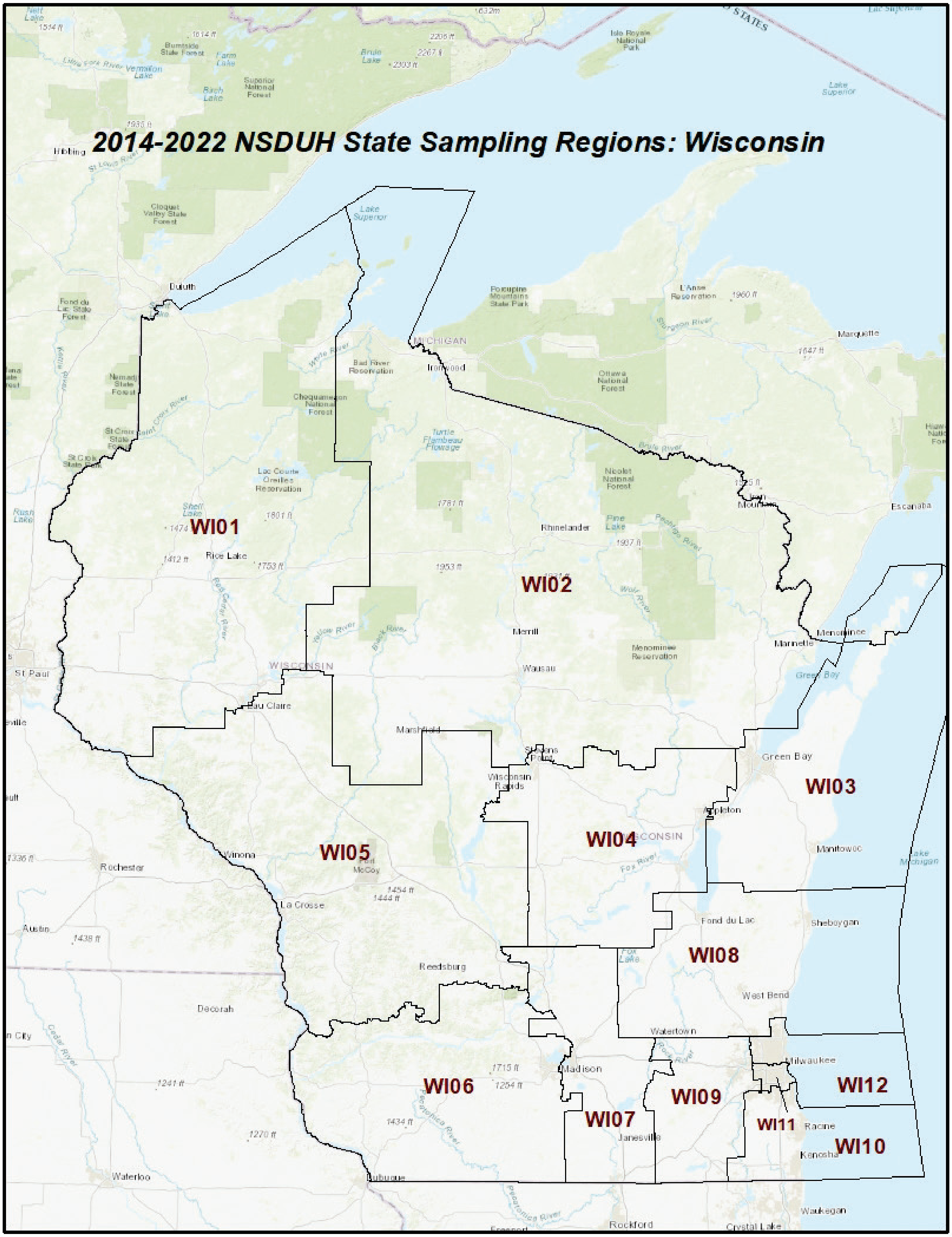

2014-2022 NSDUH State Sampling Regions: Utah
UT01
UT02
UT03
UT04
UT05
UT06
UT07
UT08
UT09
UT10
UT11
UT12
Logan
Ogden
Evanston
Rock Springs
Tooele
Orem
Provo
Vernal
Price
Richfield
Cedar City
St George
Page
Great Salt Lake
Utah Lake
GREAT SALT LAKE DESERT
SEVIER DESERT
UINTA MOUNTAINS
UINTA BASIN
EAST TAVAPUTS PLATEAU
Dugway Proving Ground
Capitol Reef National Park
Canyonlands National Park
Arches National Park
Zion National Park
Glen Canyon National Recreation Area
Grand Staircase-Escalante Nat'l Mon


2014-2022 NSDUH State Sampling Regions: Vermont
VT01
VT02
VT03
VT04
VT05
VT06
VT07
VT08
VT09
VT10
VT11
VT12
Plattsburgh
Montpelier
Rutland
Lebanon
Claremont
Concord
Manchester
Keene
Nashua
Bennington
Brattleboro
Lake Champlain
Lake George
Lake Winnipesaukee
White Mountains
Green Mountains

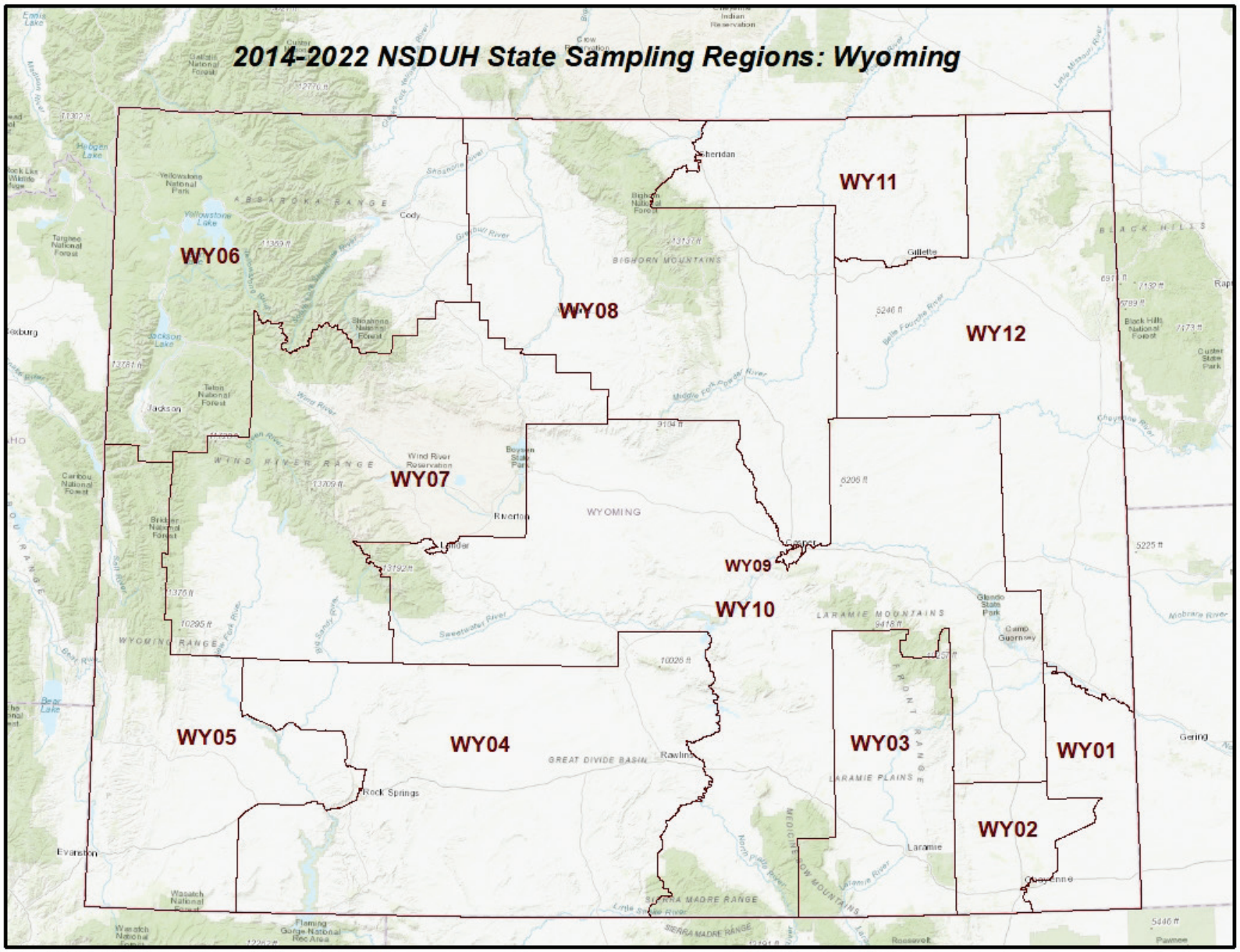

2014-2022 NSDUH State Sampling Regions: Virginia
VA 03
VA 01
VA04
VA05
VA02
VA06
VA09
VA07
VA11
VA08
VA10
VA14
VA13
VA15
VA12


2014-2022 NSDUH State Sampling Regions: Washington
WA01
WA02
WA03
WA04
WA05
WA06
WA07
WA08
WA09
WA10
WA11
WA12
UNITED STATES
Vancouver
Victoria
Seattle
Spokane
Yakima
Portland
Boise
Nelson
Castlegar
Grand Forks
Rossland
Trail
Penticton
Kennewick
Walla Walla
Pullman
Moscow
Lewiston
Moses Lake
Wenatchee
Aberdeen
Olympia
Tacoma
Everett
Bellingham
Oak Harbor
Longview
Astoria
Hillsboro
Gresham
The Dalles
Lewiston Orchards
Coeur d'Alene
Post Falls
CASCADE RANGE
OLYMPIC MOUNTAINS
WENATCHEE MOUNTAINS
MONASHEE MOUNTAINS
OKANAGAN HIGHLAND
WASHINGTON


2014-2022 NSDUH State Sampling Regions: West Virginia
WV01
WV02
WV03
WV04
WV05
WV06
WV07
WV08
WV09
WV10
WV11
WV12
Wooster
Canton
Marion
Delaware
Newark
Columbus
Zanesville
OHIO
Lancaster
Chillicothe
Athens
Portsmouth
Ashland
Huntington
Wayne National Forest
Weirton
Pittsburgh
Wheeling
Morgantown
Parkersburg
Clarksburg
Farmington
WEST VIRGINIA
Charleston
Monongahela National Forest
George Washington National Forest
Jefferson National Forest
Blacksburg
Roanoke
Lynchburg
VIRGINIA
PENNSYLVANIA
State College
Altoona
Cumberland
Hagerstown
Martinsburg
Winchester
Harrisonburg
Shenandoah National Park
Staunton
Charlottesville
Lake Anna
Rothrock State Forest
Chambersburg



2014-2022 NSDUH State Sampling Regions: Wisconsin
WI01
WI02
WI03
WI04
WI05
WI06
WI07
WI08
WI09
WI10
WI11
WI12


2014-2022 NSDUH State Sampling Regions: Wyoming
WY11
WY06
WY08
WY12
WY07
WY09
WY10
WY05
WY04
WY03
WY01
WY02
WYOMING
Sheridan
Cody
Gillette
Jackson
Riverton
Lander
Rawlins
Rock Springs
Evanston
Laramie
Cheyenne
BIGHORN MOUNTAINS
WIND RIVER RANGE
WYOMING RANGE
LARAMIE MOUNTAINS
GREAT DIVIDE BASIN
LARAMIE PLAINS
SIERRA MADRE RANGE
ABSAROKA RANGE
BLACK HILLS

Appendix C:
2020 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.

C.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.

C.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 person). 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.

C.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 2022 design. Simulation analyses based on the 2012 screening data, modified to reflect the required 2014 through 2022 age group sample proportions (but not modified to reflect the new state proportions), were conducted, and $\lambda = 0.25$ was selected (Center for Behavioral Health Statistics and Quality, 2016).

Table C.1 displays the expected pair selection counts (scaled to sum to 67,507) and corresponding pair response rates for $\lambda = 0.25$ based on 2017 screening data.

Table C.1 Expected Pair Selection Counts and Response Rates for λ = 0.25

Age Group Pair	Selected Pairs	Pair Response Rate, %
12+, 12+	28,889	61.4
12 - 17, 12 - 17	3,245	72.6
12 - 17, 18 - 25	2,670	65.6
12 - 17, 26+	8,343	65.9
18 - 25, 18 - 25	4,220	59.2
18 - 25, 26+	4,911	55.8
26+, 26+	5,499	52.7

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2017.

Appendix D:
2020 NSDUH Procedures for Subsegmenting

D.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.

D.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).[49] For the 2020 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.

D.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 originally subsegmented in-house (as described in Section D.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) not enough information was 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

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

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.

D.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 E:
2020 NSDUH Procedures for Adding Missed Dwelling Units

E.1 Introduction

In-person data collection for the 2020 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 2019. Because the listing was done a short time before the 2020 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 2020 NSDUH had in place a procedure for checking for and adding missed DUs on the premises of sampled DUs.[50] 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 through 2018 NSDUHs, the half-open interval (HOI) rule was not implemented in the 2020 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 E.6).

This appendix outlines the 2020 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 E.1 compares the 2014 through 2020 procedures with those used for the 2013 NSDUH.

[50] The check for missed DUs was not implemented for DUs screened via web.

Table E.1 Comparison of 2013 and 2014 with 2020 NSDUH Procedures for Adding Missed Dwelling Units

Missed DU Scenario	2013 Procedure	2014 to 2020 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 using the HOI rule 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.)	FIs were instructed to call their supervisors if they noticed large omissions or changes to the area. These situations were handled using bust procedures; that is, a subsample of the missed DUs was 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; NSDUH = National Survey on Drug Use and Health; 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.

E.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 as the original selection for the segment. To maintain the unequal weighting effect and to control costs associated with adding DUs, a new subsampling procedure was implemented and continued through the 2020 quarter 1 and 4 in-person survey:

Number of Added DUs	Sampling Rate
0	No action
1 to 10	Automatic (all DUs added to the sample)
11 to 25	½
26 to 40	⅓
41 to 50	¼
50 or more	⅕

E.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 E.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 E.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.[51] 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.

E.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 E.6).

E.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.

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

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

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

If more GQ units were 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

[51] 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.

3
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**

E.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 list the segment again 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 E.3.

E.7 Quality Control

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

- Sampling staff ensured 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.

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

The 2014 through 2022 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 9-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 design primary sampling units (PSUs, or 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 design strata (and design PSUs) across quarters to form variance strata (and variance PSUs). A similar approach for the 2014 to 2022 design would yield 750 national variance strata and 12 to 36 variance strata per state. Under the NSDUH design, 1 degree of freedom for variance estimation is associated with each variance 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 is being used for the 2014 through 2022 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 strata across quarters within SSRs, it is equally feasible to form variance strata across SSRs. In addition, if the SSRs that were combined to form a variance stratum come from different states, they provide some additional disclosure protection because an intruder can no longer assume that all respondents in a variance 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 2022 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.[52] 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 variance PSUs within these strata consist of the combined PSU 1 segments and combined PSU 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 $z'_{ki} - \bar{z}'_k = \sum_{h \in k}^{H_k} (z_{hi} - \bar{z}_h)$.

[52] 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 an expected value of zero because sample selection is done independently in each of the original strata. Because this second term is zero 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 G:
2020 NSDUH Census Block Assignment Procedures

G.1 Introduction

A coordinated sample design was developed for the 2014 through 2022 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 fourth 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 2022. Thus, only half of the segments are being listed for the 2015 through 2022 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 2022 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 the U.S. Census Bureau's 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 G.2). Section G.3 describes the quality control (QC) procedures 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.

G.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 5, 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 6, 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 E.6 of Appendix E), 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 5). 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 7, 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 allow users to associate DUs only with blocks that are part of the segment. If a 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 8. 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 appended to the data file include tract (TRACT10), block group (BLKGRP10), and census block (BLOCK10).[53]

G.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.

G.3.1 Built-In Quality Control Checks

- 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 there are no data entry errors and only census blocks contained in the segment are assigned.
- The user is not able to specify a DU location outside of the segment boundaries.
- The user cannot mark a segment as complete until the census blocks for all respondent DUs in the segment have been assigned.

G.3.2 Postprocessing Quality Control Checks

- Confirm that all respondent DUs are assigned to a census block.
- For each DU, confirm that the assigned census block is part of the segment associated with the DU.

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

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

Exhibit 5 Census Block Assignment Application with No Dwelling Units Assigned

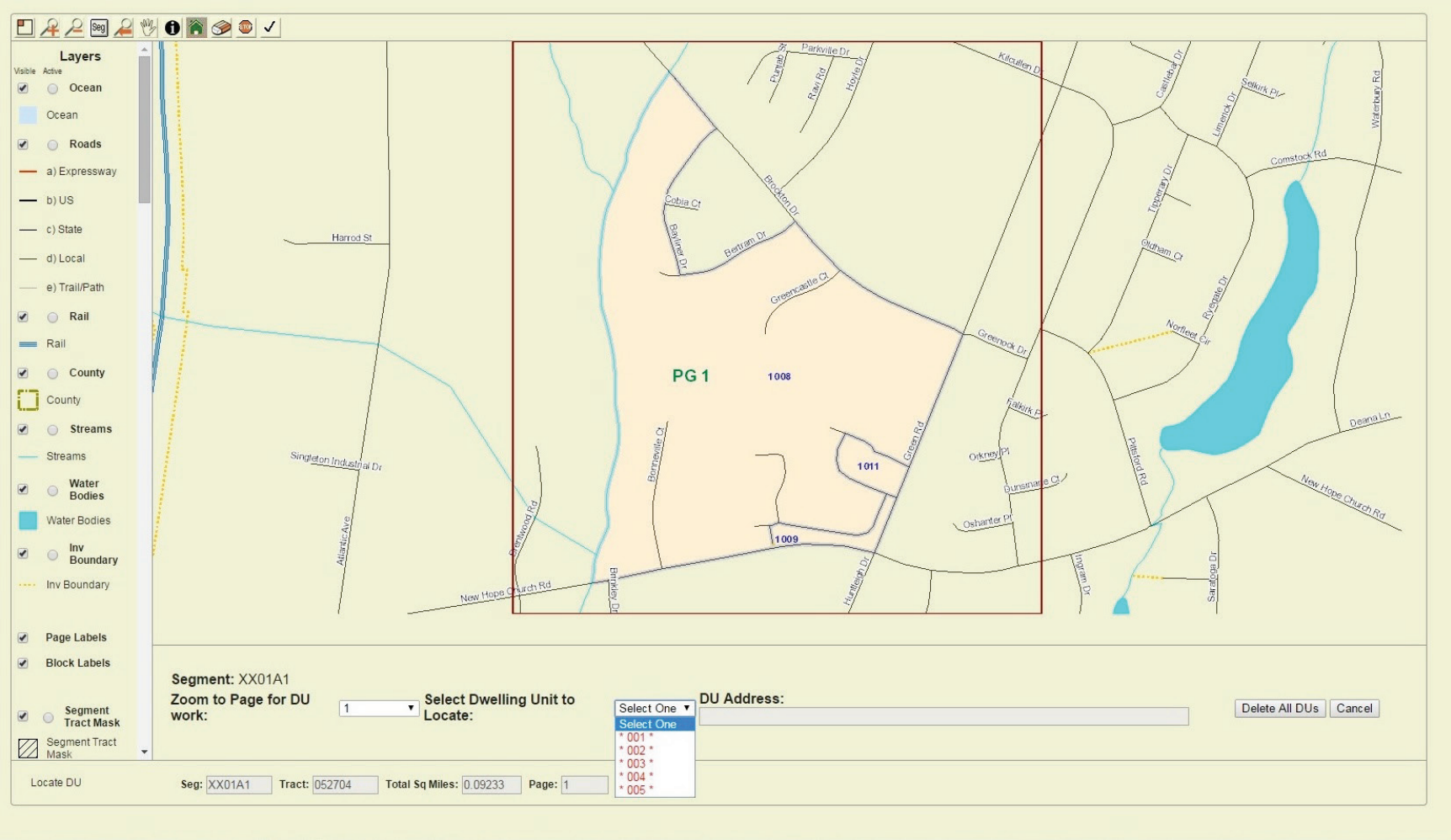


Exhibit 6 Field Listing Map

Exhibit 7 Census Block Assignment Application with Some Dwelling Units Assigned

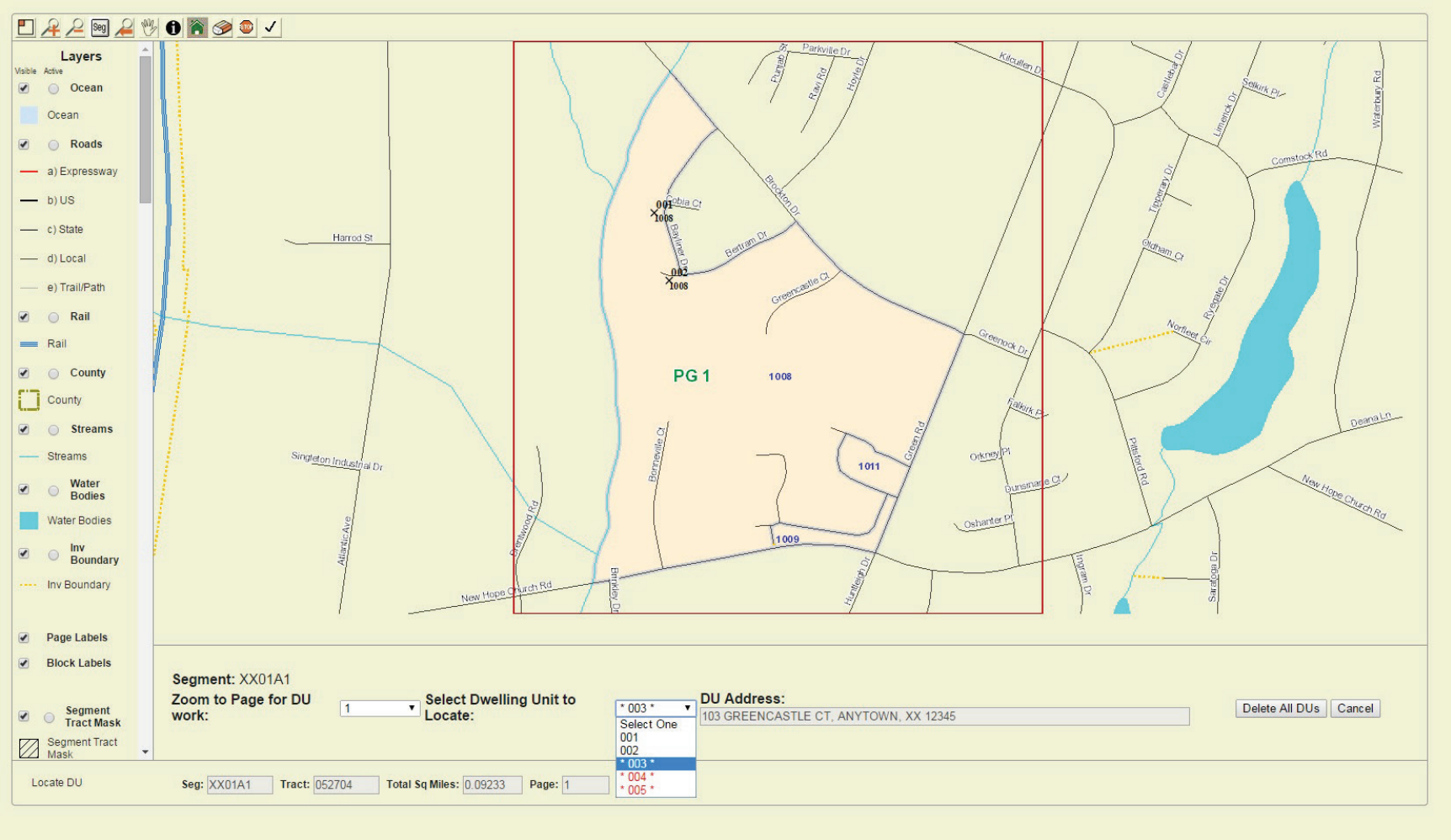
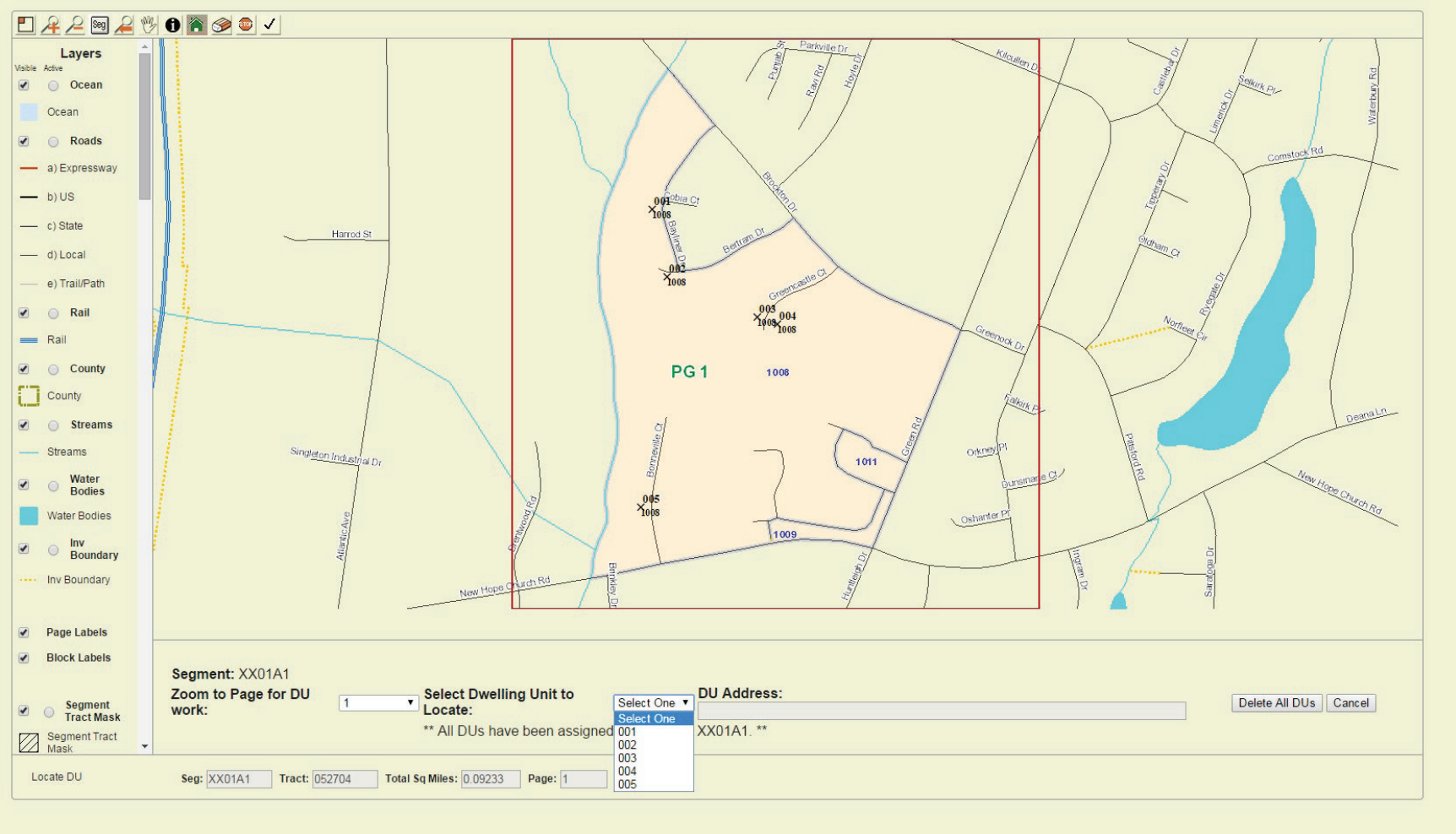


Exhibit 8 Census Block Assignment Application with All Dwelling Units Assigned



This page intentionally left blank

SAMHSA
Substance Abuse and Mental Health
Services Administration

SAMHSA's mission is to reduce the impact of substance abuse and mental illness on America's communities.

1-877-SAMHSA-7 (1-877-726-4727) | 1-800-487-4889 (TDD) | www.samhsa.gov