Table of Contents

Section Page

[List of Tables i-iv](#_Toc500836944)

[R-DAS Introduction i-1](#_Toc500836945)

[1. Overview i-1](#_Toc500836946)

[2. Analytical Options Available on R-DAS i-1](#_Toc500836947)

[3. Structure of the R-DAS Data Files i-2](#_Toc500836948)

[4. Variables on the R-DAS Data Files i-3](#_Toc500836949)

List of Tables

Table Page

[1. Geographic Variables Introduced in the 2002-2011 R-DAS Data File and
Summary Information Derived from the 2002-2011 R-DAS Data File i-9](#_Toc500919630)

[2. Counties with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the
2002-2011 R‑DAS Data File i-9](#_Toc500919631)

[3. CBSAs with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the
2002-2011 R-DAS Data File i-10](#_Toc500919632)

[4. CSAs with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the
2002-2011 R-DAS Data File i-10](#_Toc500919633)

[5. Substate Regions with Coverage Ratios of Less Than 0.8 or More Than
1.2 in the 2002‑2011 R-DAS Data File i-10](#_Toc500919634)

# R-DAS Introduction

## Overview

This codebook provides documentation for the National Survey on Drug Use and Health (NSDUH) Restricted-use Data Analysis System (R-DAS) data files. R-DAS is an online analytic system that allows analysts to produce cross-tabulations using restricted-use NSDUH data files. Restricted-use microdata are not accessible to analysts, but output from the analyses is available as long as the output does not violate any of the disclosure limitation rules that determine what output may be displayed.

This introduction provides details on the structure of NSDUH's R-DAS data files, including the contents, weights, disclosure limitation methods applied to the output, and analysis options. For more information on the survey and data, users should review NSDUH's public use file documentation available at the Substance Abuse and Mental Health Data Archive (SAMHDA), which is accessible online.[[1]](#footnote-1) In addition to R-DAS data files, the Substance Abuse and Mental Health Services Administration (SAMHSA) will continue to disseminate NSDUH's public use files through direct downloads and support an online analysis system for the files.

## Analytical Options Available on R-DAS

The basic R-DAS interface is similar to the online analysis system for NSDUH's public use files; however, in R-DAS, the analyses that can be performed, the data file codebooks, and the available output will differ. Although the current online analysis system for NSDUH's public use files on SAMHDA allows cross-tabulations and logistic regressions for a single year of NSDUH data, NSDUH's R-DAS allows only for the creation of cross-tabulations of variables on the files, excluding the weighting and variance estimation variables. Both systems allow for variance estimation that takes into account NSDUH's complex survey design.

NSDUH's R-DAS does not allow for the creation of single-year estimates, does not permit listing of individual cases, and does not show unweighted frequencies in the R-DAS codebook, nor are users allowed to generate unweighted frequencies. These limitations have been imposed to reduce the potential for disclosure of confidential information via R-DAS. Furthermore, the NSDUH data files used in R-DAS have been subsampled (i.e., they do not include the entire dataset). Revised weights were created because R-DAS produces estimates by combining NSDUH data from 2, 4, 8, 10, 12, 14, or 15 years. The revised weights were constructed in order to be representative of the average annual population across 2, 4, 8, 10, 12, 14, and 15 years.

As mentioned earlier, not all of the output that can be produced using the online data analysis system for NSDUH's public use files is available through NSDUH's R-DAS. R-DAS will not provide unweighted sample sizes, and only standard errors, confidence intervals, and test statistics that incorporate the complex survey design will be available. All weighted totals will be rounded to the nearest thousand, and all prevalence estimates and confidence intervals will be rounded to one decimal point. If any of the cells in a table contain too few unweighted cases, then the entire table will be suppressed.

## Structure of the R-DAS Data Files

From 1971 through 1998, NSDUH employed paper-and-pencil data collection. Since 1999, the NSDUH interview has been carried out using computer-assisted interviewing (CAI). Because of the shift in interviewing method in 1999, estimates from the pre-1999 surveys are not comparable with estimates from the current CAI-based surveys. In addition, although the design of the 2002 through 2016 NSDUHs is similar to the design of the 1999 through 2001 surveys, important methodological differences affect the comparability of the 2002 to 2016 estimates with estimates from prior surveys. Therefore, R-DAS includes data files starting from 2002. Additionally, NSDUH underwent a partial questionnaire redesign in 2015 that resulted in several measures no longer being comparable with their 2014 and earlier counterparts. For data files that span before and after this partial redesign, only comparable variables have been retained for analysis.

Seven separate R-DAS data files for NSDUH data include data from 2002 to 2016. The 2-, 4-, 8-, 10-, 12-, 14-, and 15-year files may all be used to produce national and state-level estimates. The 10‑, 12-, 14-, and 15- year files may also be used to produce substate estimates, such as county-level estimates; for a full list, see Table 1, which is shown at the end of this codebook introduction.

* 2-year NSDUH data file: This file can be used to create combined 2-year estimates. Estimates for the years 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012‑2013, 2014-2015, and 2015-2016 are possible. The weight on this file can be used for 2-year combined estimates only for the pair years specified above. In addition to single pair-year estimates, tables showing pair-year estimates across time can be produced. Multiyear estimates (e.g., estimates based on 4 years of data) should not be produced from the data in this 2‑year file because the totals will not be correct. Moreover, in some cases, it will not be evident which years of data are being used to produce the estimate because it is possible to have data for a certain subgroup of years but not others if there were changes in the questionnaire. For example, the variable ANLKPLMT (ABLE TO KEEP LIMT/USE MORE PN RLVR PST 12 MOS) is available in the 2002‑2003 time period and in the 2004-2005 time period, so estimates may be generated for these 2-year periods. However, ANLKPLMT is not comparable between the two time periods (i.e., 2002-2003 and 2004-2005) because of questionnaire changes between these time periods. Estimates produced using ANLKPLMT across the 4-year time period of 2002-2005 are not appropriate.

Note that a 2-year file is normally created every other year. Thus, the 2014-2015 R-DAS data file was created according to the normal pattern (i.e., 2 years removed from the previous file, which was the 2012-2013 R-DAS data file). However, during this cycle, a 2015-2016 data file was also produced because a number of changes were made in 2015 to the NSDUH questionnaire and data collection procedures. These changes were intended to improve the quality of the data that were collected and to address the changing needs of substance use and mental health policy research.[[2]](#footnote-2) These changes to the questionnaire and data collection procedures resulted in variables not being comparable between 2015 and previous years. Thus, a number of variables were not comparable between 2014 and 2015 and had to dropped from the 2014-2015 R-DAS data file. For those variables to be available to users, therefore, a 2015-2016 R-DAS data file also was created.

* 4-year NSDUH data file: This file allows the creation of combined 4-year estimates. Estimates for the years 2002-2005, 2006-2009, and 2010-2013 are possible. The weight on this file can be used for 4-year combined estimates only for the 4-year combinations of years specified above. Estimates based on more than the specified 4 years of data cannot be produced from the data in this file.
* 8-year NSDUH data file: This file allows the creation of combined 8-year estimates. Only combined estimates for the years 2002-2009 and 2006-2013 are possible with the weight provided. The weight on this file can be used for 8-year combined estimates only for the 8-year combinations of years specified above. Estimates based on more than the specified 8 years of data cannot be produced from the data in this file.
* 10-year NSDUH data file: This file allows the creation of combined 10-year estimates. Only combined estimates for the years 2002-2011 are possible with the weight provided.
* 12-year NSDUH data file: This file allows the creation of combined 12-year estimates. Only combined estimates for the years 2002-2013 are possible with the weight provided.
* 14-year NSDUH data file: This file allows the creation of combined 14-year estimates. Only combined estimates for the years 2002-2015 are possible with the weight provided.

15-year NSDUH data file: This file allows the creation of combined 15-year estimates. Only combined estimates for the years 2002-2016 are possible with the weight provided.

Users will have to decide which R-DAS data file to use based on their analytic needs. Users may not be able to use the 2-year file because of insufficient sample sizes and may need to use a 4-, 8-, 10-, 12-, 14-, or 15-year data file. Users should not combine multiple years of data because the weights have been developed for the specific year combinations on the files. *In all analyses on the 2- and 4-year data files, the variable indicating the years of data should be used either as a variable in the table or as a filter variable used to subset the data file.*

## Variables on the R-DAS Data Files

Not all of the variables on NSDUH's restricted-use data files have been included in the R‑DAS data files. Variables that cannot effectively be used in NSDUH's online analysis system, such as the date of the interview, were not included. Variables indicating geographical locations smaller than a state (e.g., counties and parishes) are not included in the 2-, 4-, and 8-year data files, but they are included in the 10-, 12-, 14-, and 15-year data files. State codes are included in every data file to allow for state-level analyses. Other variables, such as detailed race/ethnicity, country of birth, and initiation of substance use (age at first use), also are included in the R-DAS data files.

NSDUH's seven types of R-DAS data files do not have identical variables. Variables had to be comparable across the different years in consideration. For the 2-year data file, the only variables that were included had consistent and comparable data for each of the two pair years in consideration (e.g., 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013, 2014-2015, and 2015-2016). So, if a new questionnaire item was introduced in 2003 and was consistently collected through 2016, then the 2002-2003 records would not have a valid value for that variable; however, the cases or records from 2004 to 2016 would have valid values for that variable. Similarly, the only variables included in the 2002-2016 data file are ones that were collected in a comparable manner across all 15 years. A crosswalk chart in the documentation provided for NSDUH's R‑DAS data files indicates the variables that are present and comparable across the different years. Users of NSDUH's R-DAS are recommended to carefully look at this crosswalk to make sure that comparisons across time are valid for given variables. Analysts also are encouraged to refer to the instrument specifications for each of the survey years in conjunction with their review of the codebook. The specifications provide detailed information about the logic governing how respondents were routed through the questions in the interviews and any changes to the instrument relative to the survey from the prior year. The 2002 to 2016 specifications are available on SAMHSA's website.[[3]](#footnote-3)

For each variable included in NSDUH's R-DAS data files, the codebook provides the variable name, a description of the variable, the value codes, and their meanings. Unlike NSDUH's public use data file codebooks, NSDUH's R-DAS codebooks do not contain unweighted univariate frequencies because of confidentiality reasons. Most of the variables originated directly as interview items. For a subset of variables created from more than one variable, recoding specifications are provided. In addition, case identification and sampling variables are not included in the files or documented in the codebooks. Although the R-DAS data file variables are listed in the R-DAS codebooks, some variables referenced in codebook comments and appendices may not be available in the R-DAS data files because of concerns about confidentiality or comparability. There may be references in the codebook to appendices that are not available for R-DAS because of similar concerns about confidentiality. The variable documentation shown in the codebooks for the combined-year R‑DAS data files corresponds to the latest year of NSDUH data available in these combined-year data files. For the most part, no adjustments have been made to this documentation. Note that there may be exceptions for the renamed variables where information was added for clarification. The codebook appendix files for a given R-DAS data file correspond to the survey from the last year associated with the data file. For example, the latest NSDUH represented in the 2-year data file is the 2016 survey, so the appendices for the 2-year R-DAS data file come from the 2016 survey. Similarly, the latest NSDUH represented in the 8-year data files is either the 2009 survey or the 2013 survey, so the appendices for the 8‑year R-DAS data files come from the 2009 survey or the 2013 survey.

The overall organization of the R-DAS data files is shown in the table of contents of the R-DAS codebooks. Edited data from core drug modules make up the first portion of the file. Edited data from the noncore (supplemental) self-administered modules and demographic questions are in later sections.[[4]](#footnote-4) Variables from the noncore section contain missing data; for a description of the codes given to different types of missing data, see the section on Standard Code Conventions in the codebook for the public use files.

Imputation-revised drug use variables, as well as selected recoded versions of these variables, are included for core drug use variables. These imputed and recoded drug use variables are in separate sections following the edited core drug use variables. The recoded drug use variables include indicators for lifetime, past year, and past month substance use. The imputation-revised core drug use variables served as the starting point for the recoded core drug use variables. Imputation-revised core and noncore demographic variables also are included toward the end of the codebook. Missing values for all imputation-revised variables from the core drug modules have been replaced with valid values using the statistical imputation procedures described in the public use file codebooks. Imputation indicators are provided for each variable so that users may easily determine whether an observation contains data from the questionnaire or an imputed value. *Where imputed or recoded variables are provided, users are encouraged to use them to produce estimates rather than raw or edited variables from the interview.*

Variables from self-administered noncore modules make up the next major section in the codebook. In addition to edited variables in these noncore modules, recoded variables are in some of these sections. For example, edited variables from the section of the interview pertaining to symptoms of dependence or abuse (Substance Dependence and Abuse section in the codebook) were used to create recoded summary measures of dependence.

For recoded variables, the missing data codes, which are contained in the source variables and defined in the Standard Code Conventions section of the public use codebook, were often recoded to the standard missing code (.). It is recommended that cases containing these missing codes be excluded from an analysis.

In many instances, the codebook itself also indicates in parentheses the question name that corresponds to an edited variable. However, one important feature of the transition in 1999 to CAI from paper-and-pencil interviewing was that respondents could be routed to different versions of a question based on prior information from the interview. For this reason, individual variables do not always exist on NSDUH's R-DAS data files that correspond to every question in the interview.

A few variables specific to the R-DAS data files have names that end with the suffix "\_B." These variable names represent NSDUH variables whose names changed over time but remained analytically comparable. Because very minor changes were made in the documentation for the R-DAS codebooks, the renamed variables will not match the references to the original variable names found in the variable documentation. A general rule to follow is that notes for specific variables are found directly above the variable listing or at the front of that variable codebook section. The variable renaming that was applied for the R-DAS data files is described next.

The renamed variable IRRACE\_B is equivalent to IRRACE in 2002 and to IRRACE2 from 2003 to 2007. The main difference between IRRACE and IRRACE2 is in the handling of multiple race respondents. QD06 did not appear in the 2003 to 2009 NSDUHs. In the 2002 NSDUH, when nonmissing, the QD06 response was used to select a "main race" for multiple race respondents. In the 2003 to 2007 NSDUHs, a single race was imputed from among the races selected by each multiple race respondent, using a random imputation based on actual QD06 responses from 2000 to 2002. Because the purpose of the method used in 2003 to 2007 was to create a variable comparable with IRRACE, the variable names IRRACE and IRRACE2 were considered comparable across 2002 to 2007. In general, IRRACE and IRRACE2 were created by collapsing the categories given in the race questions represented by the raw variables QD051-QD057, QD04RACE, QD05RACE, QD05ASI1-QD05ASI7, and ASIARACE into one of the four listed categories. The detail provided by these questions is given in the variable IRNWRACE. Beginning with the 2008 NSDUH, multiple race respondents were placed in a separate category, and IRRACE2 was discontinued.

The renamed variables HSPRAC\_B and RACE\_B follow the same pattern as IRRACE\_B. These variables are recodes that use IRRACE\_B as one of the source variables. They were also discontinued as of the 2008 NSDUH.

The renamed variable IRENTA\_B is equivalent to IRENTAGE in 2002 and 2003 and to IRENTAG2 from 2004 to 2014. In the 2004 NSDUH, the question QD16 was replaced with three parts, which essentially asked non-U.S. born respondents to provide a more specific time on how long they had lived in the United States. The variable IRENTAG2 gives the age at which an immigrant to the United States entered this country. It was derived from the respondent's age and the variables LIVUS1YR, LIVUSYRS, and LIVUSMOS, which gave the number of years that an immigrant to the United States had lived in this country. Any respondents born in the United States (IRBORNUS=1) were coded as 999. The two variables were deemed similar enough to be comparable over all years from 2002 to 2014. In 2015, the skip logic for the two questions about living in the United States was changed. Thus, IRIMMENTAGE was created started in 2015 and replaced the variable IRENTAG2. Because the questionnaire change in 2015 caused a break in the trend, IRENTA\_B is not included on the 2014-2015, 2015-2016, 2002-2015, or 2002-2016 R-DAS data files. Instead, IRIMMENTAGE is included only on the 2015-2016 R-DAS data file.

Recoded Psychological Distress (2006 and 2007) and Recoded Adult Mental Health (2008 and 2009)―the serious psychological distress (SPD) recodes―were renamed because adjusted variables for the years 2006 and 2007 were similar to variables in 2008 and 2009, although the variable names were different. This recoding applies only to the following analysis periods: 2006-2007, 2008-2009, 2010-2011, 2012-2013, 2006-2009, 2006-2013, and 2010-2013. Note that starting with the 2-year analysis period for 2014-2015, these variables were no longer renamed.

Recoded Adult Depression―the adult major depressive episode (MDE) recodes―was renamed because adjusted variables were available that were similar to later years' variables, even though they had different names. This recoding applies only to the following analysis periods: 2006-2007, 2008-2009, 2010-2011, 2012-2013, 2006-2009, 2006-2013, and 2010-2013. Note that starting with the 2-year analysis period for 2014-2015, these variables were no longer renamed.

Recoded Drug Use―selected initiation variables―was renamed across all years because various changes were made to the initiation variables over time. This recoding applies to all of the analysis periods available with R-DAS except for the following: (1) daily cigarette initiation was not renamed for the following analysis periods: 2014-2015, 2015-2016, 2002-2015, and 2002-2016; and (2) initiation variables deemed comparable across years were not renamed for the following analysis periods: 2014-2015 and 2015-2016.

Recoded Adolescent Depression―several youth MDE variables―was renamed for the 2006-2009 analysis period because in 2006 and 2007 there were variables that contained both adult and youth data, but these variables were broken apart in 2008 and made into age-specific sets of recodes.

Recoded Education―full-time college enrollment variable―was renamed due to a questionnaire change in 2016 to the current school enrollment question to clarify the question for younger respondents. This recoding applies only to the 2015-2016 analysis period.

Recoded Income―poverty variables―was renamed due to a questionnaire change in 2015 that added an additional income response level to the finer income level questions. This recoding applies only to the following analysis periods: 2014-2015, 2002-2015, and 2002-2016.

Starting with the 10-year 2002-2011 R-DAS analysis period, the eight geographic variables listed in Table 1 were added to allow for substate estimation. These eight variables identify counties, core-based statistical areas (CBSAs), combined statistical areas (CSAs), metropolitan divisions, and other substate geographic regions. Some geographic regions represented in the NSDUH samples over the years 2002 through 2011 were combined with others when the number of observations was small (i.e., fewer than 500 cases over the 10-year period) or when a region did not appear in the NSDUH sample in each year from 2002 to 2011.

These geographic variables have also been added to the 12-year 2002-2013 R-DAS data file, the 14-year 2002-2015 R-DAS data file, and the 2002-2016 R-DAS data file. Note that the same unique areas are included for all four files (i.e., if an area was combined with another area on the 2002-2011 R-DAS 10-year data file, it was combined again on subsequent R-DAS data files). Thus, the geographic areas represented on the 10-year data file are the same geographic areas represented on the 12-, 14-, and 15-year data files.

The numeric values of all eight variables, as constructed for the 2002-2011 R-DAS data file, along with their corresponding geographic area descriptions, are listed in Appendix P. Note that the numeric values and geographic area descriptions are the same for the 2002-2013, 2002‑2015, and 2002-2016 R-DAS data files. In addition to these eight geographic variables, all of the R-DAS data files, including the 2002-2013, 2002-2015, and 2002-2016 R-DAS data files, include the state variable to allow for state-level estimates.

Substate region definitions for the STREG10 variable are provided on SAMHSA's website at <https://archive.samhsa.gov/data/>.[[5]](#footnote-5) Substate region definitions for the STREG14 variable are provided in Appendix R of this codebook and on SAMHSA's website at <https://www.samhsa.gov/data/>.[[6]](#footnote-6) CBSAs are defined by the OMB.[[7]](#footnote-7)

Because R-DAS weights are not calibrated to the substate and metropolitan area geographies shown in Table 1 (for convenience, these are all called "substate" areas), it is important to consider the degree to which substate population estimates derived from R‑DAS weights agree with population estimates from other sources. Consequently, the 2002-2011 R‑DAS data file was used to investigate coverage ratios. Coverage ratios, defined as the ratio of R-DAS weighted totals to estimated population counts[[8]](#footnote-8) as of 2008, were created for all substate geographies included in the 2002-2011 R-DAS data file. Substate geographies with coverage ratios of less than 0.8 or more than 1.2 may be more likely to produce biased estimates than other substate geographies. Coverage ratios for all 2002-2011 R-DAS substate geographies are listed in Appendix P of this codebook. Additionally, updated coverage ratios using the 2002-2016 R‑DAS data file are included in Appendix Q. Note that only the 2002-2011 R-DAS coverage ratios are discussed in this introduction.

The assessment of the 2002-2011 R-DAS data file revealed a total of 20 state-county geographies, 14 CBSAs, 6 CSAs, and 14 substate regions with coverage ratios of less than 80 percent or more than 120 percent. These geographies and their associated coverage ratios are shown in Tables 2, 3, 4, and 5, respectively.

Table 1. Geographic Variables Introduced in the 2002-2011 R-DAS Data File and Summary Information Derived from the 2002-2011 R-DAS Data File

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Description | Unique Areas in R-DAS | Total | Unique Areas with Coverage Ratio of Less Than 0.8 or More Than 1.2 |
| CBSA09\_N | CORE-BASED STATSTICAL AREA CODE | 208 | 955 | 14 |
| CBSTA109 | 2009 CBSA 1 STATUS | 3 | 3 | N/A |
| CBSTA209 | 2009 CBSA 2 STATUS | 4 | 4 | N/A |
| CSA09\_N | COMBINED STATISTICAL AREA CODE | 97 | 128 | 6 |
| CYSTAT09 | CBSA COUNTY STATUS | 3 | 3 | N/A |
| METDV09N | METROPOLITAN DIVISION CODE | 29 | 29 | 0 |
| STREG101 | 2008-2010 SUBSTATE REGION CODE | 383 | 383 | 14 |
| STREG141 | 2012-2014 SUBSTATE REGION CODE | 384 | 384 | -- |
| STCTYCOD | STATE-COUNTY CODE | 291 | 3,142 | 20 |

-- coverage ratios were not calculated; CBSA = core-based statistical area; CSA = combined statistical area; N/A = not applicable; R-DAS = Restricted-use Data Analysis System.

NOTE: With the exception of the STREG10 and STREG14 variables (see footnote 1), the same variables contained on the 2002-2011 R-DAS data file are also contained on the 2002-2013, 2002-2015, and 2002-2016 R-DAS data files.

1 The 2002-2011 and 2002-2013 R-DAS data files contain 2008-2010 substate region definitions (STREG10), whereas the 2002-2015 and 2002-2016 R-DAS data files contain 2012-2014 substate region definitions (STREG14).

Table 2. Counties with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 2002-2011 R‑DAS Data File

|  |  |  |
| --- | --- | --- |
| STCYCODE | Description | Coverage Ratio |
| 7 | 02: AK, 170: Matanuska-Susitna Borough | 0.762 |
| 27 | 06: CA, 077: San Joaquin County | 0.757 |
| 30 | 06: CA, 107: Tulare County | 1.405 |
| 39 | 08: CO, 069: Larimer County | 1.221 |
| 66 | 12: FL, 117: Seminole County | 1.362 |
| 71 | 13: GA, 121: Fulton County | 0.694 |
| 84 | 17: IL, 019: Champaign County | 1.393 |
| 95 | 17: IL, 179: Tazewell County | 1.364 |
| 96 | 17: IL, 197: Will County | 0.785 |
| 113 | 22: LA, 033: East Baton Rouge Parish | 1.226 |
| 149 | 26: MI, 145: Saginaw County | 1.210 |
| 192 | 34: NJ, 027: Morris County | 1.249 |
| 198 | 35: NM, 045: San Juan County | 1.333 |
| 201 | 36: NY, 005: Bronx County | 0.777 |
| 202 | 36: NY, 007: Broome County | 1.283 |
| 206 | 36: NY, 055: Monroe County | 1.205 |
| 231 | 39: OH, 085: Lake County | 1.403 |
| 232 | 39: OH, 089: Licking County | 1.293 |
| 280 | 45: SC, 079: Richland County | 1.269 |
| 302 | 48: TX, 491: Williamson County | 1.303 |

R-DAS = Restricted-use Data Analysis System; STCYCODE = state-county code.

Table 3. CBSAs with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 2002-2011 R-DAS Data File

|  |  |  |
| --- | --- | --- |
| CBSA09\_N | Description | Coverage Ratio |
| 35 | 16580: Champaign-Urbana, IL | 1.359 |
| 39 | 16860: Chattanooga, TN-GA | 1.256 |
| 58 | 20500: Durham-Chapel Hill, NC | 1.351 |
| 64 | 22140: Farmington, NM | 1.333 |
| 67 | 22660: Fort Collins-Loveland, CO | 1.221 |
| 80 | 25420: Harrisburg-Carlisle, PA | 0.717 |
| 86 | 26380: Houma-Bayou Cane-Thibodaux, LA | 1.286 |
| 102 | 29340: Lake Charles, LA | 1.244 |
| 130 | 33860: Montgomery, AL | 0.793 |
| 131 | 34060: Morgantown, WV | 0.768 |
| 173 | 40980: Saginaw-Saginaw Township North, MI | 1.210 |
| 184 | 43340: Shreveport-Bossier City, LA | 0.769 |
| 192 | 44700: Stockton, CA | 0.757 |
| 202 | 47300: Visalia-Porterville, CA | 1.405 |

CBSA = core-based statistical area; R-DAS = Restricted-use Data Analysis System.

Table 4. CSAs with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 2002-2011 R-DAS Data File

|  |  |  |
| --- | --- | --- |
| CSA | Description | Coverage Ratio |
| 7 | 138: Beckley-Oak Hill, WV | 0.794 |
| 13 | 174: Chattanooga-Cleveland-Athens, TN-GA | 1.219 |
| 36 | 276: Harrisburg-Carlisle-Lebanon, PA | 0.727 |
| 52 | 338: Lima-Van Wert-Wapakoneta, OH | 1.223 |
| 87 | 508: Shreveport-Bossier City-Minden, LA | 0.789 |
| 90 | 526: Sunbury-Lewisburg-Selinsgrove, PA | 1.292 |

CSA = combined statistical area; R-DAS = Restricted-use Data Analysis System.

Table 5. Substate Regions with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 2002‑2011 R-DAS Data File

|  |  |  |
| --- | --- | --- |
| STREG10 | Description | Coverage Ratio |
| 12 | 4: AZ, 4: Rural South | 0.609 |
| 29 | 6:CA, 9: Region 12R | 0.773 |
| 30 | 6: CA, 10: Region 13 (Riverside) | 0.792 |
| 36 | 6: CA, 16: Region 19R (Imperial) | 1.517 |
| 174 | 26: MI, 12: Saginaw | 1.219 |
| 193 | 29: MO, 3: Eastern (excluding St. Louis) | 0.781 |
| 226 | 36: NY, 1: Region 1 | 0.776 |
| 227 | 36: NY, Region 10 | 1.236 |
| 228 | 36: NY, Region 11 | 1.22 |
| 243 | 37: NC, 3: ECCS | 0.743 |
| 272 | 39: OH, 15: Boards 28, 43, and 67 | 1.228 |
| 280 | 40: OK, 2: East Central | 0.786 |
| 302 | 42: PA, 11: Regions 5, 18, 23, 24, and 46 | 0.793 |
| 383 | 56: WY, 9: Judicial District 9 | 0.789 |

R-DAS = Restricted-use Data Analysis System STREG10 = 2008-2010 substate region code included on the 2002-2011 and 2002-2013 R-DAS data files.

1. See the following web page: <https://datafiles.samhsa.gov/>. [↑](#footnote-ref-1)
2. The exact changes are documented in the 2015 NSDUH's Office of Management and Budget (OMB) clearance package and in a summary report on NSDUH's 2014 and 2015 redesign changes. The summary report and the 2015 questionnaire are available on SAMHSA's website at <https://www.samhsa.gov/data/>. To access these documents, click on the "Population Data / NSDUH" tab, go to the "Methodology & Questionnaires" tab, then look in the 2015 folder. The 2015 questionnaire is a file in the 2015 NSDUH Methodology and Resource Book. [↑](#footnote-ref-2)
3. For the 2002 to 2009 NSDUHs, see <https://archive.samhsa.gov/data/>, click on "Methodology Reports," then download the specific "NSDUH Methodological Resource Book" for the desired dataset year's questionnaire specifications. For the 2010 to 2016 NSDUHs, see <https://www.samhsa.gov/data/population-data-nsduh/reports?tab=38>, then click on the specific "NSDUH Methodological Resource Book" in the Methodology & Questionnaires tab for the desired dataset year's questionnaire specifications (e.g., 2011's specifications are in the file named "2k11Q.pdf"). [↑](#footnote-ref-3)
4. A core set of questions critical for basic trend measurement of prevalence estimates remains in the survey every year and comprises the first part of the NSDUH interview. Noncore questions or modules (which can be revised, dropped, or added from year to year) make up the remainder of the interview. The core consists of initial demographic items (which are interviewer-administered) and self-administered questions pertaining to the use of tobacco, alcohol, marijuana, cocaine, crack cocaine, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives. [↑](#footnote-ref-4)
5. Click on the "Behavioral Health Prevalence (NSDUH)" tab, then under the 2012 set of files, click on "2008-2010 NSDUH Substate Estimates of Substance Use and Mental Disorders." The substate region definitions for the STREG10 variable are in the file on the "2008-2010 National Survey on Drug Use and Health Substate Region Definitions." [↑](#footnote-ref-5)
6. Click on the "Population Data / NSDUH" tab, then go to the 2014 files in the "Substate" folder to access the file on the "2012-2014 NSDUH Substate Region Definitions." [↑](#footnote-ref-6)
7. See the following reference: Office of Management and Budget. (2009, December 1). *OMB Bulletin No. 10-02: Update of statistical area definitions and guidance on their uses*. Washington, DC: The White House. Retrieved from <https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/bulletins/2010/b10-02.pdf> [↑](#footnote-ref-7)
8. Estimated population counts were provided by Claritas, Inc. Claritas is a market research firm headquartered in Ithaca, New York (see <https://www.claritas.com/>). Claritas used to be affiliated with Nielsen Holdings, from which they became independent in January 2017. [↑](#footnote-ref-8)