

Data Analysis System, Restricted-use Files (DAS-R)

Introduction to Codebooks

Substance Abuse and Mental Health Services Administration
Center for Behavioral Health Statistics and Quality
Rockville, Maryland 20857

February 2024

Table of Contents

Section	Page
<u>DAS-R Introduction.....</u>	<u>i-1</u>
<u>1. Overview.....</u>	<u>i-1</u>
<u>2. Analytical Options Available on DAS.....</u>	<u>i-1</u>
<u>3. History and Structure of DAS-R.....</u>	<u>i-2</u>
<u>4. Variables in DAS-R.....</u>	<u>i-4</u>
<u>5. Summary of the DAS-R Weights.....</u>	<u>i-10</u>
<u>5.1 Poststratification Adjustment.....</u>	<u>i-10</u>
<u>5.2 Final DAS-R Weights.....</u>	<u>i-11</u>

List of Tables

Table	Page
<u>1. Geographic Variables Introduced in the 2002-2011 DAS-R and Summary Information Derived from the 10-Year 2002-2011 DAS-R</u>	<u>i-12</u>
<u>2. Counties with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2002-2011 DAS-R</u>	<u>i-13</u>
<u>3. Counties with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2010-2019 DAS-R</u>	<u>i-14</u>
<u>4. CBSAs with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2002-2011 DAS-R</u>	<u>i-14</u>
<u>5. CBSAs with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2010-2019 DAS-R</u>	<u>i-15</u>
<u>6. CSAs with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2002-2011 DAS-R</u>	<u>i-15</u>
<u>7. CSAs with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2010-2019 DAS-R</u>	<u>i-15</u>
<u>8. Substate Regions with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2002-2011 DAS-R</u>	<u>i-16</u>
<u>9. Substate Regions with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 15-Year 2002-2016 DAS-R</u>	<u>i-16</u>
<u>10. Substate Regions with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 16-Year 2002-2017 DAS-R</u>	<u>i-17</u>
<u>11. Demographic Variables Used in DAS-R Weight Calibration</u>	<u>i-18</u>
<u>12. Outcome Variables Used in DAS-R Weight Calibration</u>	<u>i-19</u>

DAS-R Introduction

1. Overview

This introduction provides details on the structure of the National Survey on Drug Use and Health (NSDUH) Data Analysis System, restricted-use data files (DAS-R), including the contents, weights, disclosure limitation methods applied to the output, and analysis options. The Data Analysis System (DAS) is an online analytic system that allows analysts to produce cross-tabulations using both public use and 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 can be displayed.

For more information on the survey and data, refer to the NSDUH public use file documentation available on the Substance Abuse and Mental Health Data Archive (SAMHDA) website at <https://datafiles.samhsa.gov/>. In addition to DAS-R, the Substance Abuse and Mental Health Services Administration (SAMHSA) continues to disseminate NSDUH public use data through direct downloads.

2. Analytical Options Available on DAS

The basic DAS interface is similar for the NSDUH restricted-use and public use files; however, with DAS-R, the data file codebooks and the available output differ. Although DAS allows cross-tabulations for a single year of public use NSDUH data, it allows only for the creation of pooled data cross-tabulations of variables for DAS-R. Whatever the file type, DAS estimates the variances that take into account the complex survey design of NSDUH.

In DAS, using the restricted files means users cannot do the following:

- create single-year estimates,
- list individual cases, and
- generate unweighted frequencies, which are also not shown in the restricted-use file codebook.

These limitations have been imposed to reduce the potential for disclosure of confidential information via DAS-R. Furthermore, the NSDUH data files available in DAS have been subsampled (i.e., they do not include the entire dataset). Revised weights were also created because DAS-R produces estimates by combining NSDUH data from 2, 4, 8, 10, 12, 14, 15, or 16 years. The revised weights were constructed in order to be representative of the average annual population across 2, 4, 8, 10, 12, 14, 15, and 16 years. See Section 5 for more details about the weights.

Weighted estimates, standard errors, confidence intervals, and test statistics that incorporate the complex survey design are all available through DAS-R. All weighted totals are rounded to the nearest thousand, and all prevalence estimates and confidence intervals are rounded to one decimal point. If any of the cells in a table contain too few unweighted cases, then the entire table is suppressed (i.e., users will get a message that the table cannot be produced because of suppression).

3. History and Structure of DAS-R

NSDUH has incorporated two types of changes over the years: (1) changes in data collection procedures and (2) changes in survey methods.

Data collection procedures have changed multiple times since the survey started. From 1971 through 1998, NSDUH employed paper-and-pencil data collection in a respondent's place of residence. From 1999 to 2019, the data were collected via face-to-face (in-person) interviews at a respondent's place of residence using a combination of computer-assisted personal interviewing conducted by an interviewer and audio computer-assisted self-interviewing. Because of the coronavirus disease 2019 (COVID-19) pandemic, the 2020 data collection was paused for a time and an additional web data collection mode was introduced. Multimode data collection procedures continued to be used in the 2021 and 2022 surveys. Because of the shift in the interviewing method in 1999, estimates from the pre-1999 pencil-and-paper surveys are not comparable with estimates from the current computer-assisted interviewing surveys. Due to the impact of COVID-19 on data collection procedures, the 2020 data are not comparable with any other years of NSDUH data. Additionally, estimates based on multimode data collection used in 2021 and 2022 are not comparable with estimates from 2020 or prior years. Thus, the 2021-2022 DAS-R estimates should not be compared with the DAS-R estimates from prior years.

In addition to data collection procedures, other changes in survey methods have occurred across the years. Although the design of the 2002-2022 NSDUHs is similar to the design of the 1999-2001 surveys, important methodological¹ differences affect the comparability of the 2002-2022 estimates with estimates from prior surveys. Therefore, DAS only includes data files starting from 2002. NSDUH underwent a partial questionnaire redesign in 2015 that resulted in several measures no longer being comparable with their 2014 and earlier counterparts. For all DAS-R files that stretch multiple years, only comparable variables have been retained on the data files for analysis. Files that include years both before and after the redesign are especially likely to exclude variables for this reason.

Eight separate types of NSDUH DAS-R files include data from 2002 to 2022.² The 2-, 4-, 8-, 10-, 12-, 14-, 15-, and 16-year files can all be used to produce national and state-level estimates of annual averages pooled across the years in the file. The 10-, 12-, 14-, 15-, and 16-year files can also be used to produce substate estimates, such as county-level estimates for select counties. For a full list of substate variables, see [Table 1](#), which appears at the end of this codebook introduction.

Multiyear pooled estimates should not be produced from combining the data in different files because the totals would not be correct. For example, estimates based on 4 years of data

¹ Changes starting with the 2002 NSDUH included paying each respondent who completed the interview \$30. Also, the name of the survey was changed in 2002 from the National Household Survey on Drug Abuse (NHSDA) to the current name, and updated population data from the 2000 decennial census were incorporated into the sample weights starting with the 2002 estimates.

² The 2019-2020 DAS-R was produced, but it has since been removed from the SAMHSA website. Methodological investigations found that the unusual societal circumstances in 2020 and the resulting methodological revisions to NSDUH data collection have affected the comparability of 2020 estimates with estimates from 2019 and earlier and 2021 and beyond. Consequently, estimates and datasets that involve combining data from 2020 with data from prior years have been removed from the SAMHSA website.

should not be produced from combining data from two different 2-year files. Moreover, in some cases, it would not be clear which years of data are being used to produce the estimates because it is possible to have data for some years but not for others when the questionnaire changes. For example, the variable ANLKPLMT (ABLE TO KEEP LIMIT/USE MORE PN RLVR PST 12 MOS) is available in the 2002-2003 time period and in the 2004-2005 time period, so it is possible to generate estimates for both of 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. Estimates produced using ANLKPLMT across the 4-year time period of 2002-2005 are not appropriate. A crosswalk chart in the documentation provided for the NSDUH DAS-R indicates the variables that are present across the different years.

The following list summarizes the eight types of DAS-R files and provides relevant updated information about each type:

- **2-Year NSDUH Data Files:** These files allow for the creation of 2-year estimates. Estimates for the years 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013, 2014-2015, 2015-2016, 2016-2017, 2017-2018, 2018-2019, and 2021-2022 are possible. The weight on these files can be used for 2-year combined estimates only for the pair years specified above. Starting with 2014-2015, a 2-year file has been created every year except 2019-2020 and 2020-2021.³ Prior to that, the 2-year file was created every other year (e.g., the 2010-2011 and 2012-2013 DAS-R files that were published 2 years apart).
- **4-Year NSDUH Data Files:** These files allow for the creation of combined 4-year estimates. Estimates for the years 2002-2005, 2006-2009, 2010-2013, 2014-2017, and 2015-2018 are possible. The 2014-2017 file excluded variables that were not comparable between 2014 and 2015 because of the changes in 2015 to the NSDUH questionnaire and data collection procedures that resulted in some variables not being comparable. 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.⁴ Thus, a 2015-2018 file was created to have a 4-year file that contained those variables and other new variables that were the result of the 2015 questionnaire changes. The weight on this file can be used for 4-year combined estimates only for the 4-year combinations of years specified above.
- **8-Year NSDUH Data Files:** These files allow for the creation of combined 8-year estimates. Estimates for the years 2002-2009 and 2006-2013 are possible with the weight provided.

³ The 2019-2020 DAS-R was produced, but it has since been removed from the SAMHSA website. Methodological investigations found that the unusual societal circumstances in 2020 and the resulting methodological revisions to NSDUH data collection have affected the comparability of 2020 estimates with estimates from 2019 and earlier and 2021 and beyond. Consequently, estimates and datasets that involve combining data from 2020 with data from prior years have been removed from the SAMHSA website.

⁴ The exact changes are documented in the 2015 NSDUH Office of Management and Budget (OMB) clearance package and in a summary report on the NSDUH 2014 and 2015 redesign changes. The summary report and the 2015 questionnaire are available at <https://www.samhsa.gov/data/report/summary-effects-2015-nsduh-questionnaire-redesign-implications-data-users> and <https://www.samhsa.gov/data/report/2015-nsduh-questionnaire>.

- **10-Year NSDUH Data Files:** These files allow for the creation of combined 10-year estimates. Estimates for the years 2002-2011 and 2010-2019 are possible with the weight provided.
- **12-Year NSDUH Data File:** This file allows for 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 for 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 for the creation of combined 15-year estimates. Only combined estimates for the years 2002-2016 are possible with the weight provided.
- **16-Year NSDUH Data File:** This file allows for the creation of combined 16-year estimates. Only combined estimates for the years 2002-2017 are possible with the weight provided.

Users can decide which DAS-R file to use based on their analytic needs. Users may not be able to use the 2-year file because of insufficient sample sizes for some analyses and may need to use a 4-, 8-, 10-, 12-, 14-, 15-, or 16-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.

4. Variables in DAS-R

Not all of the variables on the NSDUH restricted-use data files have been included in DAS-R. The following types of variables have been excluded from DAS-R:

- variables that cannot effectively be used in the NSDUH online analysis system, such as the date of the interview;
- variables for analyzing interactions between members of the same household (i.e., pair-level variables);
- variables related to survey administration and processing because they are irrelevant to most users, which, starting in 2021-2022, includes variables about which unit a respondent reported a measurement in (e.g., HTANSWER [preferred way to report height] is not available);
- starting in 2021-2022, the variable for the quarter of the year that the respondent took the survey is not available.

Unlike in the 1-year and concatenated public use DAS, state codes are included in every DAS-R file to allow for state-level analyses. Variables indicating geographical locations smaller than a state (e.g., selected counties and parishes) are not included in the 2-, 4-, and 8-year data files, but they are included in the 10-, 12-, 14-, 15-, and 16-year data files. These other variables, such as detailed race/ethnicity, country of birth, and initiation of substance use (age at first use), also are included in DAS-R.

The eight types of DAS-R files include 2-, 4-, 8-, 10-, 12-, 14-, 15-, and 16-year files. These different types of DAS-R files do not have identical variables. Variables had to be consistently defined across the specific years in consideration to be included in a particular

DAS-R file. This is different from the DAS public use concatenated files, which include all variables that are comparable in at least two of the included file years. For instance, in a 2-year data file, the only variables that were included had consistent data for both pooled years (e.g., 2018-2019 or 2021-2022). So, if a new questionnaire item was introduced in 2003 and was consistently collected through 2022, then the 2002-2003 restricted-use file would not include that variable; however, the cases or records from 2004 to 2022 would have valid values for that variable and would be included on those files. Similarly, the only variables included in the 2002-2017 data file are ones that were collected in a comparable manner across all 16 years.

A crosswalk chart⁵ in the documentation provided for the NSDUH DAS-R indicates the variables that are present across the different years. 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 2002-2022 specifications are available on the SAMHDA website.⁶

For each variable included in the NSDUH DAS-R, the codebook provides the variable name, a description of the variable, the value codes, and their meanings. Unlike the NSDUH public use data file codebooks, the DAS-R codebooks do not contain unweighted frequencies for confidentiality reasons. Most of the variables originated directly as interview items. For those variables created from more than one interview item, recoding specifications are provided as well.

The variable documentation shown in the codebooks for the combined-year DAS-R 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, although there may be exceptions for the renamed variables where information was added for clarification (see information about renamed variables later in this section). The codebook appendix files for a given DAS-R 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 2022 survey, so the appendices for the 2-year DAS-R come from the 2022 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 DAS-R come from the 2009 survey or the 2013 survey.

Case identification and sampling variables are not included in the files or documented in the codebooks. Some variables referenced in the codebook comments and appendices may also not be available in DAS-R because of concerns about confidentiality or comparability. In addition, there may be references in the codebook to appendices that are not available for DAS-R for similar reasons.

⁵ For the crosswalk chart, go to <https://www.datafiles.samhsa.gov/>, click on the “Data Sources” tab, then click on “Population Data (NSDUH).” From there, scroll down and click “RUFs Variable Crosswalk Chart: Multi-Year (xlsx).”

⁶ For the questionnaire specifications, go to <https://www.datafiles.samhsa.gov/>, click on the “Data Sources” tab, then click on “Population Data (NSDUH).” From there, pick any year from the menu on the right, then click on one of the questionnaire specifications links below the “Dataset Documentation” heading.

Both recoded and imputation-revised variables are included in DAS-R. Imputation-revised drug use and demographic variables, as well as selected recoded versions of these variables, are included. Variables that did not undergo statistical imputation (or those that are not based on imputed variables) 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 that accompanies each public use file. Missing values for all imputation-revised variables from the drug sections 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 can easily determine whether an observation contains data from the questionnaire or an imputed value.

Users are encouraged to use imputed or recoded variables to produce estimates when they are provided, rather than raw or edited variables from the interview. Users should identify recoded variables of their topic of interest first, if available. If recoded versions are not available, they should determine whether any imputed versions are available before using raw or edited variables. Recoded and imputed variables are found in sections of the codebook that are labeled as such. Starting with the 2015 NSDUH, all new recoded variables have an “RC” in the front portion of the variable label to help users identify them. In DAS, all recoded variables are labeled “ – recoded” and imputation-revised variables are labeled “- imputation revised.” 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 file codebook, were often recoded to the standard missing code (.). It is recommended that cases containing these missing codes be excluded from an analysis. Note that this could lead to potential estimation bias due to the missingness.⁷

The overall organization of DAS-R is shown in the table of contents of the DAS-R codebooks. In 2021 and prior years, edited data from the questionnaire’s drug sections make up the first portion of the file. Edited data from the second set of self-administered sections and demographic questions are in later portions of the file.⁸ Starting in 2022, the first portion of the codebook consists of documentation for the demographic variables, followed by edited substance use variables.

Imputed and recoded drug use variables are in separate sections following the edited drug use variables. The recoded drug use variables include indicators for lifetime, past year, and past month substance use. The imputation-revised drug use variables served as the starting point for

⁷ See Section 4 in the following reference: Center for Behavioral Health Statistics and Quality. (2023). *2021 National Survey on Drug Use and Health (NSDUH) methodological resource book, Section 13: Statistical inference report*. <https://www.samhsa.gov/data/report/nsduh-2021-statistical-inference-report>

⁸ A main 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. Edited data from the second set of self-administered sections and demographic questions (which can be revised, dropped, or added from year to year) are in later sections and make up the remainder of the interview. The main set of questions consists of the 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. Starting with the 2015 NSDUH, questions pertaining to the use of methamphetamine were moved into their own section.

the recoded drug use variables. Imputation-revised demographic variables are also included (until 2022, toward the end of the codebook; after that, at the beginning).

Variables from physical and mental health sections make up the next major portion of the codebook. In addition to edited variables in these self-administered sections, recoded variables are in some of these sections. For example, edited variables from the section of the interview pertaining to the symptoms of depression for adults (the Adult Depression section in the codebook) were used to create a recoded measure classifying a respondent as having a past year major depressive episode (MDE).

For many edited variables, the codebook also indicates in parentheses the name for the question that the variable was based on. However, respondents can be routed to different versions of a question based on prior information from the interview. For this reason, individual variables do not always exist in the NSDUH DAS-R that correspond to every question in the interview.

A few variables specific to the DAS-R have names that end with the suffix “_B.” These represent NSDUH variables whose names changed over time but remained comparable for analysis. Because only very minor changes were made in the documentation for the DAS-R codebooks, the renamed variables do 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 DAS-R 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. In the 2002 NSDUH, the QD06 response was used to select a “main race” for multiple race respondents, but question QD06 was not included in subsequent years. In order to make IRRACE2 comparable with IRRACE, a single race was imputed from among the races selected by each multiple race respondent from 2003 to 2007, using a random imputation based on actual QD06 responses from 2000 to 2002. Because the purpose of this method was to create a variable comparable with IRRACE, the variables IRRACE and IRRACE2 were considered comparable across 2002-2007. 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, 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 in 2015 and replaced the variable IRENTAG2. Because the questionnaire change in 2015 caused a break in the trend, IRENTA_B is not included in the 2002-2015, 2002-2016, or 2002-2017 DAS-R. Instead, IRIMMENTAGE is included in the 2015-2016, 2016-2017, 2017-2018, 2015-2018, 2018-2019, and 2021-2022 DAS-R. Note that IRENTA_B is included only in the following DAS-R: 2002-2005, 2002-2009, 2002-2011, and 2002-2013.

Recoded Psychological Distress (2006 and 2007) and Recoded Adult Mental Health (2008 and 2009)—the serious psychological distress recodes (K6MXAJ_B and SPDYAJ_B)—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 and the 4-year analysis period for 2014-2017, these variables were no longer renamed.

Recoded Adult Depression—the adult 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, 2010-2013, and 2006-2013. Note that starting with the 2-year analysis period for 2014-2015 and the 4-year analysis period for 2014-2017, 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 DAS-R except for the following: (1) daily cigarette initiation was not renamed for the following analysis periods: 2014-2015, 2015-2016, 2016-2017, 2017-2018, 2018-2019, 2019-2020, 2014-2017, 2015-2018, 2002-2015, 2002-2016, 2002-2017, and 2010-2019; and (2) initiation variables deemed comparable across years were not renamed for the following analysis periods: 2014-2015, 2015-2016, 2016-2017, 2017-2018, 2018-2019, 2021-2022, 2014-2017, 2015-2018, and 2010-2019.

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 (COLLENRFT_B)—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 and 2015-2018 analysis periods.

Recoded Income—poverty variables (POVERTY_B)—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, 2014-2017, 2002-2015, 2002-2016, 2002-2017, and 2010-2019.

Starting with the 10-year 2002-2011 DAS-R 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 2002-2011 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 DAS-R, the 14-year 2002-2015 DAS-R, the 15-year 2002-2016 DAS-R, and the 16-year 2002-2017 DAS-R. Note that the same unique areas are included for all five files (i.e., if an area was combined with another area in the 2002-2011 DAS-R 10-year file, it was combined again in subsequent DAS-R files). Thus, the geographic areas represented in the 10-year 2002-2011 data file are the same geographic areas represented in the 12-, 14-, 15-, and 16-year data files.

For the 2010-2019 DAS-R, updates were made to the way some geographical areas were combined with others when the number of observations was small (i.e., fewer than 500 cases over the new 10-year period) or when a region did not appear in the NSDUH sample in each year from 2010 to 2019. Because the variables were created differently, they have different names in the 2010-2019 DAS-R (see [Table 1](#) for details).

The numeric values of all eight geographic variables, as constructed for the 2002-2011 DAS-R, along with their corresponding geographic area descriptions, are listed in Appendix R. Note that the numeric values and geographic area descriptions are the same for the 2002-2013, 2002-2015, 2002-2016, and 2002-2017 DAS-R. The numeric values of all eight variables, as constructed for the 2010-2019 DAS-R, along with their corresponding geographic area descriptions, are listed in Appendix T. In addition to these eight geographic variables, all of the DAS-R files, including the 2002-2013, 2002-2015, 2002-2016, 2002-2017, and 2010-2019 DAS-R, include the state variable to allow for state-level estimates.

Substate region definitions are available at <https://www.samhsa.gov/data/report/2008-2010-nsduh-substate-region-definitions> for the STREG10 variable and at <https://www.samhsa.gov/data/report/2012-2014-nsduh-substate-region-definitions> for the STREG14 variable. Substate region definitions for the STREG16 variable are available at <https://www.samhsa.gov/data/report/2014-2016-nsduh-substate-region-definitions>. CBSAs are defined by the Office of Management and Budget (OMB).²

DAS-R weights are not calibrated to the substate and metropolitan area geographies shown in [Table 1](#) (for convenience, these are all called “substate” areas below). It is important

² Office of Management and Budget. (2009, December 1). *OMB Bulletin No. 10-02: Update of statistical area definitions and guidance on their uses*. https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/bulletins/2010/b10-02.pdf

for users to consider the degree to which substate population estimates derived from DAS-R weights agree with population estimates from other sources, so the 2002-2011 DAS-R was used to investigate coverage ratios. Coverage ratios, defined as the ratio of DAS-R weighted totals to estimated population counts¹⁰ as of 2008, were created for all substate geographies included in the 2002-2011 DAS-R. 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 DAS-R substate geographies are listed in Appendix P of the 2002-2011 DAS-R codebook. Additionally, updated coverage ratios using the 2002-2016 DAS-R are included in Appendix Q, coverage ratios using the 2002-2017 DAS-R are included in Appendix R, and coverage ratios using the 2010-2019 DAS-R are included in Appendix T.


The assessment of the 2002-2011 DAS-R revealed a total of 20 state-county geographies, 14 CBSAs, 6 CSAs, and 13 substate regions with coverage ratios of less than 0.8 or more than 1.2. These geographies and their associated coverage ratios are shown in [Tables 2 to 10](#) at the end of this codebook introduction.


5. Summary of the DAS-R Weights

Analysis weights for DAS-R have been developed for each DAS-R file. In each NSDUH cycle, the process of developing DAS-R weights involved two steps: (1) subsampling the full NSDUH sample and (2) poststratifying the NSDUH analysis weights for the DAS-R subsample to the national population estimates for the civilian, noninstitutionalized population aged 12 years or older residing in the 50 states and the District of Columbia. The NSDUH data files used in DAS-R were subsampled from the full NSDUH sample (i.e., they did not include the entire dataset) by demographic group. The single-year DAS-R weights were combined and adjusted for multiple-year DAS-R files (i.e., 2 years, 4 years, 8 years, 10 years, 12 years, 14 years, 15 years, and 16 years).

5.1 Poststratification Adjustment

The NSDUH analysis weights of the DAS-R subsample cases were poststratified to the national population estimates for demographic domains. In addition, the weighted prevalence rates of the DAS-R subsample for three outcome variables in some demographic domains were forced to match the weighted prevalence rates from the full NSDUH sample. The generalized exponential model (GEM) developed by Folsom and Singh in 2000 was used for the poststratification.¹¹ Similar to the process of developing NSDUH analysis weights, the poststratification adjustment for DAS-R weights was done in nine model groups corresponding to the nine census divisions.

¹⁰ Estimated population counts were provided by Claritas, Inc., a market research firm headquartered in Cincinnati, Ohio (see <https://www.claritas.com/> .

¹¹ 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). <http://www.asasrms.org/Proceedings/index.html> .

The input weight for poststratification was ANALWT¹² for the DAS-R subsample cases. The control totals used for poststratification were from two sources. For the demographic domains, the control totals were annual population estimates obtained from the U.S. Census Bureau. For the outcome variables, the control totals were the final weighted totals from the full NSDUH sample.¹³ In the 2021-2022 DAS-R weighting process, education, interactions of education with demographic variables, and interactions of education with state were added to the poststratification adjustment model. The education population estimates were obtained from the American Community Survey.

The same demographic variables used in the poststratification adjustment in developing the standard analysis weight for the full sample (i.e., ANAWLT, or ANALWT2 for 2021-2022) were the initial variables for the poststratification adjustment for DAS-R weighting (see [Table 11](#)). Three outcome variables—past month alcohol use (ALCMON), past month cigarette use (CIGMON), and past month marijuana use (MRJMON)—and their interactions with demographic variables and state were also used (see [Table 12](#)). Some variables listed in [Table 11](#) had to be dropped or collapsed because of model convergence issues in GEM. The variables kept in the final models varied across model groups and years.

For each record in DAS-R, the final DAS-R analysis weights were the product of ANALWT (or ANALWT2 for 2021-2022) and the poststratification adjustment factor from GEM. The sum of the final DAS-R weights matched the population estimates for demographic variables controlled in the poststratification. The weighted prevalence rates for three outcome variables and their interactions with demographic variables were the same as from the full NSDUH sample.

The distributions of the weights before and after calibration were compared to uncover any unusual impact of the weight adjustments on the initial weights. In addition to the weight distributions, the ratios of the maximum weight to the mean weight and the unequal weighting effects were compared across various domains before and after the poststratification adjustment. Coverage bias analysis based on the slippage rates (the distance between the total subsample's weighted count and the target population count) also was conducted to check the impact of poststratification on various noncontrolled domains (i.e., those factors that were dropped or collapsed in the models).

5.2 Final DAS-R Weights

The DAS-R weights for a single year were pooled with those of other year(s) and adjusted into multiple-year weights for the multiple-year combined DAS-R files. That is, the adjusted DAS-R weights for multiple NSDUH years were created as the pooled set of single-year DAS-R analysis weights divided by the number of years of combined data. For example,

¹² In 2021 and 2022, the analysis weight called ANALWT2 was developed to reduce the bias caused by interview mode and was used as the input weight for the 2021-2022 DAS-R. In addition to the same set of demographic and geographic variables used in the usual poststratification adjustment, interview mode was added as a main effect, in which the target for web mode was 30 percent of the population estimate and the target for the in-person mode was 70 percent of the population estimate. The 30/70 web and in-person split matches the expected mode proportions for the 2022 and future NSDUHs.

¹³ The three outcome variables were (1) past month alcohol use, (2) past month cigarette use, and (3) past month marijuana use.

combined 2021 and 2022 DAS-R weights were calculated as the single-year DAS-R weights divided by 2. Four-year combined 2002-2005 DAS-R weights were calculated as the DAS-R weights divided by 4. Fifteen-year combined 2002-2016 DAS-R weights were calculated as the DAS-R weights divided by 15.

Table 1. Geographic Variables Introduced in the 2002-2011 DAS-R and Summary Information Derived from the 10-Year 2002-2011 DAS-R

Variable	Description	Unique Areas in DAS-R	Total	Unique Areas with Coverage Ratio of Less Than 0.8 or More Than 1.2
CBSA09 N ¹	Core-Based Statistical Area Code	208	942	14
CBSA09 N2 ¹	Core-Based Statistical Area Code	216	942	14 ²
CBSTA109	2009 CBSA 1 Status	3	3	N/A
CBSTA209	2009 CBSA 2 Status	4	4	N/A
CSA09 N ¹	Combined Statistical Area Code	97	125	6
CSA09 N2 ¹	Combined Statistical Area Code	94	125	3 ²
CYSTAT09	CBSA County Status	3	3	N/A
METDV09N	Metropolitan Division Code	29	29	0
STREG10 ³	2008-2010 Substate Region Code	383	383	13
STREG14 ³	2012-2014 Substate Region Code	384	384	11 ²
STREG16 ³	2014-2016 Substate Region Code	406	406	18 ²
STCTYCOD ¹	State-County Code	291	3,142	20
STCTYCOD2 ¹	State-County Code	309	3,143	28 ²

CBSA = core-based statistical area; CSA = combined statistical area; DAS-R = Data Analysis System, restricted-use data files; N/A = not applicable.

NOTE: With the exception of the STREG10, STREG14, STREG16, CBSA09_N, CBSA09_N2, CSA09_N, CSA09_N2, STCTYCOD, and STCTYCOD2 variables (see table footnotes 1 and 2), the same variables contained in the 2002-2011 DAS-R are also contained in the 2002-2013, 2002-2015, 2002-2016, 2002-2017, and 2010-2019 DAS-R.

¹ CBSA09_N, CSA_09_N, and STCTYCOD were created for the 2002-2011 DAS-R (i.e., the unique areas in DAS-R have a sample size of 500 or more, and the areas appear in all 10 years). These same variables were included in the 2002-2013, 2002-2015, 2002-2016, and 2002-2017 DAS-R. The unique areas were re-evaluated for the 2010-2019 DAS-R (using the same criteria that were used for the 2002-2011 DAS-R). Thus, the variables included in the 2010-2019 DAS-R are CBSA09_N2, CSA09_N2, and STCTYCOD2. The unique areas can differ between the two sets of variables.

² The number of areas listed here for CBSA09_N2, CSA09_N2, and STCTYCOD2 are based on coverage ratios calculated based on the 2010-2019 DAS-R, the number of areas listed here for STREG14 are based on coverage ratios calculated based on the 2002-2016 DAS-R, and the number of areas for STREG16 are based on coverage ratios calculated based on the 2002-2017 DAS-R.

³ The 2002-2011 and 2002-2013 DAS-R contain 2008-2010 substate region definitions (STREG10), the 2002-2015 and 2002-2016 DAS-R contain 2012-2014 substate region definitions (STREG14), and the 2002-2017 and 2010-2019 DAS-R contain 2014-2016 substate region definitions (STREG16).

Table 2. Counties with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2002-2011 DAS-R

STCTYCOD	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

DAS-R = Data Analysis System, restricted-use data files; STCTYCOD = state-county code.

Table 3. Counties with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2010-2019 DAS-R

STCTYCOD2	Description	Coverage Ratio
29	06: CA, 075: San Francisco County	0.767
31	06: CA, 081: San Mateo County	1.225
36	06: CA, 111: Ventura County	0.788
60	12: FL, 033: Escambia County	0.783
62	12: FL, 069: Lake County	1.295
64	12: FL, 073: Leon County	1.218
68	12: FL, 097: Osceola County	1.331
73	12: FL, 115: Sarasota County	1.201
84	15: HI, 007: Kauai County	1.242
90	16: ID, 083: Twin Falls County	1.362
97	17: IL, 111: McHenry County	1.213
99	17: IL, 163: St. Clair County	1.235
103	18: IN, 003: Allen County	1.229
125	23: ME, 003: Aroostook County	1.214
143	25: MA, 023: Plymouth County	1.288
184	31: NE, 153: Sarpy County	0.781
211	35: NM, 045: San Juan County	1.224
219	36: NY, 055: Monroe County	1.214
224	36: NY, 071: Orange County	1.201
227	36: NY, 103: Suffolk County	1.219
237	38: ND, 101: Ward County	0.753
245	39: OH, 089: Licking County	1.261
247	39: OH, 095: Lucas County	0.775
253	39: OH, 173: Wood County	1.273
308	48: TX, 121: Denton County	1.330
313	48: TX, 439: Tarrant County	1.302
317	49: UT, 005: Cache County	1.292
345	54: WV, 061: Monongalia County	0.791

DAS-R = Data Analysis System, restricted-use data files; STCTYCOD2 = state-county code.

Table 4. CBSAs with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2002-2011 DAS-R

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; DAS-R = Data Analysis System, restricted-use data files.

Table 5. CBSAs with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2010-2019 DAS-R

CBSA09_N2	Description	Coverage Ratio
7	11540: Appleton, WI	1.347
20	13780: Binghamton, NY	1.223
36	16580: Champaign-Urbana, IL	1.225
65	22140: Farmington, NM	1.224
86	25860: Hickory-Lenoir-Morganton, NC	1.283
103	28180: Kapaa, HI	1.242
117	30860: Logan, UT-ID	1.237
127	33500: Minot, ND	0.799
133	34060: Morgantown, WV	0.784
148	37100: Oxnard-Thousand Oaks-Ventura, CA	0.788
191	43780: South Bend-Mishawaka, IN-MI	1.215
196	44600: Steubenville-Weirton, OH-WV	1.252
202	45820: Topeka, KS	0.787
206	46300: Twin Falls, ID	1.375

CBSA = core-based statistical area; DAS-R = Data Analysis System, restricted-use data files.

Table 6. CSAs with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2002-2011 DAS-R

CSA_N	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; DAS-R = Data Analysis System, restricted-use data files.

Table 7. CSAs with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2010-2019 DAS-R

CSA_N2	Description	Coverage Ratio
1	102: Albany-Corvallis-Lebanon, OR	1.250
27	222: Dothan-Enterprise-Ozark, AL	1.260
75	466: Rockford-Freeport-Rochelle, IL	1.238

CSA = combined statistical area; DAS-R = Data Analysis System, restricted-use data files.

Table 8. Substate Regions with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 10-Year 2002-2011 DAS-R

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
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

DAS-R = Data Analysis System, restricted-use data files; STREG10 = 2008-2010 substate region code included in the 2002-2011 and 2002-2013 DAS-R.

Table 9. Substate Regions with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 15-Year 2002-2016 DAS-R

STREG14	Description	Coverage Ratio
12	4: AZ, 4: Rural South	0.639
29	6: CA, 9: Region 12R	0.737
41	6: CA, 21: Region 5R (San Francisco)	0.780
91	13: GA, 5: Region 5	0.797
135	22: LA, 1: Region 1	0.760
251	37: NC, 14: Trillium Health Resources 2	0.676
271	39: OH, 13: Boards 23 and 45	1.222
273	39: OH, 15: Boards 28, 43, and 67	1.223
342	49: UT, 2: Central, Four Corners, San Juan, and Southwest	0.790
359	53: WA, 4: Peninsula	0.726
384	56: WY, 9: Judicial District 9	0.750

DAS-R = Data Analysis System, restricted-use data files; STREG14 = 2012-2014 substate region code included in the 2002-2015 and 2002-2016 DAS-R.

Table 10. Substate Regions with Coverage Ratios of Less Than 0.8 or More Than 1.2 in the 16-Year 2002-2017 DAS-R

STREG16	Description	Coverage Ratio
12	4: AZ, 4: South B	0.631
29	6: CA, 9: Region 12R	0.716
41	6: CA, 21: Region 5R (San Francisco)	0.773
48	8: CO, 2: Region 2	0.724
63	11: DC, 1: Ward 1	0.790
93	13: GA, 5: Region 5	0.797
108	17: IL, 3: Region 1.3 (North Central Side)	0.569
117	17: IL, 12: Region 3a (Champaign)	1.228
148	22: LA, 1: Region 1	0.744
250	36: NY, 11: Region 5: Mohawk Valley	0.770
254	36: NY, 15: Region 9: Southern Tier	1.229
259	37: NC, 5: Cardinal Innovations Healthcare Solutions 3	0.796
268	37: NC, 14: Trillium Health Resources 2	0.675
271	38: ND, 3: North Central	0.762
326	44: RI, 3: Region 3: Providence	0.776
366	49: UT, 5: Southwest	0.726
383	53: WA, 6: Salish	0.740
406	56: WY, 9: Judicial District 9	0.769

DAS-R = Data Analysis System, restricted-use data files; STREG16 = 2014-2016 substate region code included in the 2002-2017 DAS-R.

Table 11. Demographic Variables Used in DAS-R Weight Calibration

Label	Number of Levels	Levels
State	Varied	Depended on model groups corresponding to census divisions
Quarter	4	Four calendar quarters
Age Group	6	Six age groups: 12-17, 18-25, 26-34, 35-49, 50-64, 65 or older
Race (5)	5	Five race categories: White, Black/African American, American Indian/Alaska Native, Asian/Pacific Islander, multiple races
Gender	2	Male, female
Hispanicity	2	Hispanic, non-Hispanic
Education	4	Less than high school, high school graduate, some college, college graduate
Age Group × Race (3)	18	Six age groups, three race categories: White, Black/African American, others
Age Group × Hispanicity	12	Six age groups, two Hispanicity categories
Age Group × Gender	12	Six age groups, two gender categories
Race (3) × Hispanicity	6	Three race categories, two Hispanicity categories
Race (3) × Gender	6	Three race categories, two gender categories
Age Group (5) × Education	20	Five age groups, four education categories
Race (3) × Education	12	Three race categories, four education categories
Gender × Education	8	Two gender categories, four education categories
Hispanicity × Education	8	Two Hispanicity categories, four education categories
Hispanicity × Gender	4	Two Hispanicity categories, two gender categories
State × Quarter	Varied	Depended on model groups corresponding to census divisions
State × Age Group	Varied	Depended on model groups corresponding to census divisions
State × Race (5)	Varied	Depended on model groups corresponding to census divisions
State × Hispanicity	Varied	Depended on model groups corresponding to census divisions
State × Gender	Varied	Depended on model groups corresponding to census divisions
State × Education	Varied	Depended on model groups corresponding to census divisions
Age Group × Race (3) × Hispanicity	36	Six age groups, three race categories, two Hispanicity categories
Age Group × Race (3) × Gender	36	Six age groups, three race categories, two gender categories
Age Group × Hispanicity × Gender	24	Six age groups, two Hispanicity categories, two gender categories
Race (3) × Hispanicity × Gender	12	Three race categories, two Hispanicity categories, two gender categories
State × Age Group × Race (3)	Varied	Depended on model groups corresponding to census divisions
State × Age Group × Hispanicity	Varied	Depended on model groups corresponding to census divisions
State × Age Group × Gender	Varied	Depended on model groups corresponding to census divisions
State × Race (3) × Hispanicity	Varied	Depended on model groups corresponding to census divisions
State × Race (3) × Gender	Varied	Depended on model groups corresponding to census divisions
State × Hispanicity × Gender	Varied	Depended on model groups corresponding to census divisions

DAS-R = Data Analysis System, restricted-use data files.

NOTE: The education domain was added to the weight calibration in 2020 and beyond. Prior to 2020, education and its interactions were not used.

NOTE: In 2021 and 2022, survey mode was added as a main effect in the DAS-R weight calibration, the same as in developing the main analysis weights.

Table 12. Outcome Variables Used in DAS-R Weight Calibration

Label	Number of Levels	Levels
ALCMON	1	Past month alcohol use
CIGMON	1	Past month cigarette use
MRJMON	1	Past month marijuana use
Age Group × ALCMON	6	Past month alcohol use by six age groups
Age Group × CIGMON	6	Past month cigarette use by six age groups
Age Group × MRJMON	6	Past month marijuana use by six age groups
Race (3) × ALCMON	3	Past month alcohol use by three race categories
Race (3) × CIGMON	3	Past month cigarette use by three race categories
Race (3) × MRJMON	3	Past month marijuana use by three race categories
Hispanicity × ALCMON	2	Past month alcohol use by two Hispanicity categories
Hispanicity × CIGMON	2	Past month cigarette use by two Hispanicity categories
Hispanicity × MRJMON	2	Past month marijuana use by two Hispanicity categories
Gender × ALCMON	2	Past month alcohol use by two gender categories
Gender × CIGMON	2	Past month cigarette use by two gender categories
Gender × MRJMON	2	Past month marijuana use by two gender categories
State × ALCMON	Varied	Depended on model groups corresponding to census divisions
State × CIGMON	Varied	Depended on model groups corresponding to census divisions
State × MRJMON	Varied	Depended on model groups corresponding to census divisions

ALCMON = past month alcohol use; CIGMON = past month cigarette use; DAS-R = Data Analysis System, restricted-use data files; MRJMON = past month marijuana use.