**Addendum to the 2002-2019**

**Combined Public Use Data File**

## Introduction

This codebook provides documentation for the 1,005,421 records on the combined 2002‑2019 National Survey on Drug Use and Health (NSDUH) public use data file. It contains 4,741 variables, the majority of which are available across two or more of the individual-year public use files (PUFs). This addendum provides information specific to the combined file. Detailed information specific to each of the survey years is available in the documentation associated with the individual PUFs and can be accessed online through the Substance Abuse and Mental Health Data Archive (SAMHDA) at <https://datafiles.samhsa.gov/>.

## Organization of the Data File

The records on this data file are combined from each individual PUF for 2002 through 2019. Because the identifier for each record on the individual PUFs, QUESTID2, may be repeated across years, a new variable, YEAR, was added to the combined data file. The combination of YEAR and QUESTID2 will result in unique observations. The variable YEAR may also be used to subset the data to specific study years for analysis purposes.

The majority of the variables found in each of the single-year PUFs are included on the 2002-2019 combined PUF. For the most part, as long as a variable was available on more than 1 year of the data files, it was included on this combined file. Retaining or dropping variables from the combined PUF was based on the analytic utility for multiple-year data analysis. Some variables were retained on the combined PUF, even if they were only available in a single year, if they were considered analytically useful. A few variables that had low analytic utility or were available for only 1 or 2 years were not included on this combined PUF. These variables can be obtained from the individual PUFs. Some variables that were not included on the 2002-2018 combined PUF are now retained on the 2002‑2019 PUF because they are available for 2 years (2018 and 2019) of data. Variables that either were new in 2015 or were not considered comparable with their counterparts from 2002 to 2014 due to the partial questionnaire redesign in the 2015 NSDUH (e.g., the cancer diagnostic variables and those dealing with the use of prescription psychotherapeutics) continue to be retained on the 2002-2019 combined PUF. Similar to the 2002-2018 combined PUF, some demographic variables, such as education and employment status, and several substance use outcome variables, such as "recoded any illicit drug use in past month" and "recoded binge alcohol use in past 30 days," which were excluded for 2015 from the 2002-2015 combined PUF, remain available on the 2002-2019 combined PUF. The marital status variable IRMARIT was recreated for the 2016-2019 PUFs (similar to 2002-2014) and reflects the move of the marital status questions from self-administration in 2015 back to interviewer administration in 2016 to 2019. The 2015 marital status variable IRMARITSTAT, the imputation indicator variable IIMARITSTAT, and the recoded variable MARIEDNUM2 (number of times married) were also retained on the 2002-2019 combined PUF. Analytic goals should be considered prior to pooling or comparing marital status data from 2015 with data from other years. For details on how the two marital status variables differ, see the 2016 PUF codebook.[[1]](#footnote-1) In addition, variables with the same names from the Marijuana Purchases section were added back in 2018 (and retained in 2019) after last being used in 2014 (they are not available in the years from 2015 to 2017). Because of this break in years, data users are cautioned about comparing 2018 and 2019 estimates for these variables with estimates from 2014 or earlier years. Data users also are advised not to pool data for these variables from 2018 and 2019 with corresponding data from 2014 or earlier years.

It is worth noting that several variables that were not comparable across time were retained on the file. Reasons for variables not being comparable across study years may include questionnaire changes, skip logic (i.e., routing) changes, or changes in how recoded variables were created. Additionally, from 2015 onward, COUTYP4 (COUNTY METRO/NONMETRO STATUS [2013 3-LEVEL]), which was created based on the 2013 Rural/Urban Continuum Codes (RUCC13), was included to replace COUTYP2, which was created based on the 2003 Rural/Urban Continuum Codes (RUCC03). COUTYP2 was still retained for the 2002-2014 data. Also, the poverty variable, POVERTY3 (RC-POVERTY LEVEL-NEW INC [% OF US CENSUS POVERTY THRESHOLD]) created and used from 2015 onward is comparable with POVERTY2 in previous years. For details regarding this variable, see the 2015 PUF codebook.[[2]](#footnote-2) A crosswalk (referred to as the "combined PUF compatibility spreadsheet") in the documentation provided for the combined 2002-2019 PUF indicates the variables that are present and comparable across the different years. Users are encouraged to look carefully at this crosswalk to ensure that comparisons across time are valid for given variables.

Analysts are encouraged to refer to the questionnaires (i.e., the computer-assisted interviewing [CAI] specifications) for each of the survey years in conjunction with their review of the 2002-2019 codebook. The questionnaires provide detailed information about how respondents were routed through the questions in the interviews and changes to the instrument relative to the survey from the prior year.

The 2002-2019 questionnaires, as well as other resource materials and earlier years' documentation, can be found on the Substance Abuse and Mental Health Services Administration (SAMHSA) website and on SAMHDA's webpages:

* Start at SAMHSA's main webpage at <https://www.samhsa.gov/>, choose "Data We Collect" from the drop-down list under the "Data" tab, and then choose "National Survey on Drug Use and Health" in the center of the page. To find a specific year's questionnaire, scroll down to "Latest Survey Resources" and choose a year (e.g., 2019) from the drop-down list beside "Year Data Collected" on the right-hand side of the page. Scroll through the search results (displayed on multiple web pages) until you find "NSDUH 2019 Questionnaire."

Alternatively, start at SAMHDA's main webpage at <https://datafiles.samhsa.gov/>, click on the "Download Data" heading, then click on "National Survey on Drug Use and Health (NSDUH)," and then choose a dataset year (e.g., NSDUH-2019 National Survey on Drug Use and Health) or a set of aggregated years (NSDUH-2002-2018 National Survey on Drug Use and Health) from the list. To access the dataset's documentation (e.g., Codebook.pdf, Questionnaire-Specs.pdf, Questionniare-Showcards.pdf), click on "NSDUH-2019-DS0001" under "Datasets in this Study" on the right-hand side of the page.

Because of changes in how variables may have been created or recoded across years, the number of levels or level contents may vary from year to year for some variables. For example, for the industry and occupation variables (WRKIDST2, WRKIDSY2, WRKOCUP2, and WRKOCUY2), the category "Armed Forces" did not exist from 2008 to 2014. For certain other variables, the same values in the categories may have different meanings. For example, for the WHAT YEAR LAST WORKED variable, WRKLSTY2, the year 1968 stands for the exact year last worked in the 2007 and prior data. However, in 2008, it stands for 1968 and earlier. Similar differences also apply to the counterpart variable WRKLASTYR2 in years 2015 to 2019. For the four weight gain or weight loss variables (ADWRGNL2, ADWRLSL2, YOWRGNL2, and YOWRLSL2), levels 21 and higher mean different weight categories for years before and after 2010. Caution should be taken, therefore, when combining years with such data, and further collapsing may be needed to make the levels comparable. Additional details on these variables are provided in Tables A.1 to A.4 later in this document.

## Weights and Design Variables for the Combined PUF

When analyzing any single year of data or when conducting yearly trend analysis, the variable ANALWC1 should be used. This variable is the same as the variable ANALWT\_C that is found on the single-year PUFs. However, with a combined file, analysts have the option of using pooled data from 2 or more years. Therefore, in addition to the analysis weights for a single year of data (ANALWC1), additional weight variables, ANALWC2 to ANALWC18, were created to allow for multiple-year data analysis. These additional weight variables were created by adjusting the single-year weights by a scalar factor (i.e., the number of years of data used) so that the estimated numbers of individuals reported is representative of the national population. For example, ANALWC2, which can be used for producing estimates of annual averages using any combination of 2 years of data (e.g., pooled 2002-2003 data, pooled 2018-2019 data, or even 2008 and 2010 combined data), was obtained by dividing the single-year weight ANALWC1 by 2. Similarly, ANALWC18 was obtained by dividing ANALWC1 by 18 and can be used for producing estimates of annual averages using all 18 years of NSDUH data (i.e., combined data from 2002 to 2019).

Selecting which weights to use depends on the years of data being analyzed. This can depend on the years that certain variables are available because some of the analytic variables are not available or not comparable for all 18 years. (For details, see the combined 2002-2019 compatibility spreadsheet.) For example, a user may be interested in obtaining estimates from 18 years of data, but for variables like OTHINS (OTHER HEALTH INSURANCE), which is available for only 17 of those 18 years (i.e., no other health insurance data for 2002 were collected), ANALWC17 should be used instead of ANALWC18 during analysis for that particular variable. Also, for variables, such as YUFCSOR (STAYED IN FOSTER CARE FOR SOME OTHER REASON), if data from 2002 to 2004 (3 years) are used, ANALWC3 should be used, and if data from 2005 to 2019 (15 years) are used, ANALWC15 should be used. Note that, for YUFCSOR, data collected prior to 2004 is not comparable with data collected after 2005. Therefore, caution should be taken when choosing the weights for multiple-year data analysis. Examples of SAS®, Stata, and SUDAAN® code focusing on how to use the weights appear toward the end of this document.

For years when a split-sample design was implemented (i.e., in 2004 and 2008), a special weight was developed for the split samples. When an analysis with pooled data from multiple years is conducted, analysts should check the codebook introduction and other documents for that specific year on how to use the analysis weights in order to use a proper analysis weight for the pooled data analysis.

For example, in the 2004 NSDUH, the sample of adult respondents aged 18 or older was split approximately evenly, where respondents in sample A were administered the full adult mental health questionnaire section as it had been administered in 2002 and 2003, and respondents in sample B were administered a short version of the adult mental health questionnaire section in addition to the adult depression questionnaire section as it was administered in 2005, 2006, and 2007. Thus, analyses using 2004 data from either the adult mental health section (renamed the psychological distress section in 2005, 2006, and 2007)or the adult depression section need to be conducted using a different weight variable. For 2004, analyses that include the adult mental health variables for major depressive episode and/or unadjusted serious psychological distress (SPD) should involve the following:

* Select either sample A or sample B by using the ADLTSAMP variable (ADLTSAMP = 1 corresponds to sample A, and ADLTSAMP = 2 corresponds to sample B) to restrict the analysis to the appropriate half sample.

When combining 2004 adult half-sample data with data for survey years that have comparably defined adult mental health variables, the weight variable (the SPD analysis weight for 2004 and the person-level analysis weight ANALWC1 for the other years in the analysis) should be divided by the total number of years in the analysis.

For variance estimation, no adjustment needs to be made to the sample design variables VESTR (variance estimation [pseudo] stratum) and VEREP (variance estimation [pseudo] replicate within stratum). Note that there are 60 pseudo strata (resulting in 60 degrees of freedom for variance estimation) per year for the 2002-2013 PUF data and 50 pseudo strata (resulting in 50 degrees of freedom) per year for the 2014 and subsequent data. This change is due to the sample redesign implemented in the 2014 NSDUH.[[3]](#footnote-3) When combining any pair of years of data (e.g., 2015 and 2016, 2018 and 2019), the degrees of freedom remain the same as if the pair of years was a single year (e.g., 50 for national estimates) when these years are part of the same sample design. When combining years with different degrees of freedom (e.g., 2013 and 2014), the specific number of denominator degrees of freedom (DDF) can be computed by counting the unique values of VESTR. For example, when combining data for 2015 and 2016, DDF = 50 can be used because the sample design remained the same across those 2 years. When producing estimates by combining data for 2013 and 2014, DDF = 110 can be used because the sample design changed in 2014. When comparing estimates in two domains with different degrees of freedom, researchers should be conservative and use the smaller degrees of freedom. For example, when comparing 2013 estimates with 2014 estimates, DDF = 50 should be used. For details about degrees of freedom, see Chapter 6 in the 2017 statistical inference report.[[4]](#footnote-4) As with the single-year PUFs, users of multiyear PUFs should first sort the combined data by the sample design variables VESTR and VEREP, then specify them in a statistical software package, such as SUDAAN,[[5]](#footnote-5) to estimate variances and standard errors (SEs).

## Variables with Differences in Levels in the Combined PUF

This section describes the variables that have levels that are not comparable across the 2002-2019 PUFs, where variables in items 1 and 2 pertain to the 2003-2014 data, variables in item 3 pertain to the 2002-2019 data, and variables in item 4 pertain to the 2004-2019 data.

### 1. Industry Variables

**WRKIDST2:** TYPE OF BUSINESS OR INDUSTRY

**WRKIDSY2:** TYPE OF BUSINESS OR INDUSTRY

Table A.1 provides descriptions of the industry variable levels. Note that, starting in 2008, the "Armed Forces" category (17) became unavailable. It was collapsed with "Public Administration" (15). These two variables are not available from 2015 onward.

Table A.1 Descriptions of the Industry Variable Levels

|  |  |  |
| --- | --- | --- |
| Level | Description (2003-2007 NSDUHs) | Description (2008-2014 NSDUHs) |
| 1 | Agriculture, Forestry, Fishing, & Hunting | Agriculture, Forestry, Fishing, & Hunting |
| 2 | Mining | Mining |
| 3 | Construction | Construction |
| 4 | Manufacturing, Nondurable Goods | Manufacturing, Nondurable Goods |
| 5 | Manufacturing, Durable Goods | Manufacturing, Durable Goods |
| 6 | Transportation & Utilities | Transportation & Utilities |
| 7 | Information & Communications | Information & Communications |
| 8 | Wholesale Trade, Durable Goods | Wholesale Trade, Durable Goods |
| 9 | Wholesale Trade, Nondurable Goods | Wholesale Trade, Nondurable Goods |
| 10 | Retail Trade | Retail Trade |
| 11 | Finance, Insur, Real Estate, Rental & Leasing | Finance, Insur, Real Estate, Rental & Leasing |
| 12 | Professional/Scientific/Mgmt/Admin/Waste Mgmt | Professional/Scientific/Mgmt/Admin/Waste Mgmt |
| 13 | Education, Health & Social Services | Education, Health & Social Services |
| 14 | Arts/Entertain/Recreation/Accommodation/Food Svcs | Arts/Entertain/Recreation/Accommodation/Food Svcs |
| 15 | Public Administration | Public Administration |
| 16 | Other Services (Except Public Admin) | Other Services (Except Public Admin) |
| 17 | Armed Forces | — |

### 2. Occupation Variables

**WRKOCUP2:** KIND OF WORK/JOB TITLE PRIMARY JOB

**WRKOCUY2:** KIND OF WORK/JOB TITLE PRIMARY JOB

Table A.2 provides descriptions of the occupation variable levels. Note that, starting in 2008, the "Armed Forces" category (15) became unavailable. It was collapsed with "Protective Service Occupations" (8). These two variables are not available from 2015 onward.

Table A.2 Descriptions of Occupation Variable Levels

|  |  |  |
| --- | --- | --- |
| Level | Description (2003-2007 NSDUHs) | Description (2008-2014 NSDUHs) |
| 1 | Executive/Administrative/Managerial/Financial | Executive/Administrative/Managerial/Financial |
| 2 | Professional (not Education/Entertainment/Media) | Professional (not Education/Entertainment/Media) |
| 3 | Education & Related Occupations | Education & Related Occupations |
| 4 | Entertainers, Sports, Media, & Communications | Entertainers, Sports, Media, & Communications |
| 5 | Technicians & Related Support Occupations | Technicians & Related Support Occupations |
| 6 | Sales Occupations | Sales Occupations |
| 7 | Office & Administrative Support Workers | Office & Administrative Support Workers |
| 8 | Protective Service Occupations | Protective Service Occupations |
| 9 | Service Occupations, Except Protective | Service Occupations, Except Protective |
| 10 | Farming, Fishing, & Forestry Occupations | Farming, Fishing, & Forestry Occupations |
| 11 | Installation, Maintenance & Repair Workers | Installation, Maintenance & Repair Workers |
| 12 | Construction Trades & Extraction Workers | Construction Trades & Extraction Workers |
| 13 | Production, Machinery Setters/Operators/Tenders | Production, Machinery Setters/Operators/Tenders |
| 14 | Transportation & Material Moving Workers | Transportation & Material Moving Workers |
| 15 | Armed Forces | — |

### 3. "What Year Last Worked" Variable

**WRKLSTY2:** WHAT YEAR LAST WORKED for 2002 to 2014

**WRKLASTYR2:** WHAT YEAR LAST WORKED for 2015 to 2019

Before 2008, the first category for this variable was coded to "1960 or earlier." For 2008 and 2009, it was coded to (SURVEY YEAR - 40). For 2010 and later years, it was coded to (SURVEY YEAR - 39). The cut point lumps all of the prior data together. Therefore, values on or before the cut points have different meanings for the combined PUF. Table A.3 provides a breakdown of the years for which the first category was coded.

Table A.3 Coding of the First Category for the "What Year Last Worked" Variable

|  |  |
| --- | --- |
| NSDUH Survey Year | First Category Coded As |
| 2007 and earlier | 1960 or earlier |
| 2008 | 1968 or earlier |
| 2009 | 1969 or earlier |
| 2010 | 1971 or earlier |
| 2011 | 1972 or earlier |
| 2012 | 1973 or earlier |
| 2013 | 1974 or earlier |
| 2014 | 1975 or earlier |
| 2015 | 1976 or earlier |
| 2016 | 1977 or earlier |
| 2017 | 1978 or earlier |
| 2018 | 1979 or earlier |
| 2019 | 1980 or earlier |

### 4. Weight Gain/Loss Variables

Four recoded weight gain and weight loss variables were derived from the adult and youth depression questionnaire sections. ADWRGNL2 and ADWRLSL2 are the recoded adult depression section variables, and YOWRGNL2 and YOWRLSL2 are the recoded adolescent depression section variables. If respondents reported gaining or losing weight and these gains or losses could not be attributed to factors other than depression (e.g., growth, pregnancy, dieting), respondents were asked to report the number of pounds they gained or lost.

**ADWRGNL2:** WHEN PRBLMS WORST # LBS GAINED WITHOUT TRYING

**ADWRLSL2:** WHEN PRBLMS WORST # LBS LOST WITHOUT TRYING

**YOWRGNL2:** WHEN PRBLMS WORST # LBS GAINED WITHOUT TRYING

**YOWRLSL2:** WHEN PRBLMS WORST # LBS LOST WITHOUT TRYING

The levels were consistent from 2004 to 2009 and from 2010 to 2019. Table A.4 provides descriptions of the weight gain and weight loss variable levels.

Table A.4 Descriptions of the Weight Gain and Weight Loss Variable Levels

|  |  |  |
| --- | --- | --- |
| Level | Description (2004-2009 NSDUHs) | Description (2010 and Later NSDUHs) |
| 0-20 | Single level for 0-20 pounds (i.e., 0, 1, 2, …, 20 pounds) | Single level for 0-20 pounds (i.e., 0, 1, 2, …, 20 pounds) |
| 21 | 21 pounds | 21-25 pounds |
| 22 | 22 pounds | 26-30 pounds |
| 23 | 23 pounds | 31-35 pounds |
| 24 | 24 pounds | 36-40 pounds |
| 25 | 25 pounds | 41-45 pounds |
| 26 | 26 pounds | 46-50 pounds |
| 27 | 27 pounds | 51 or more pounds |
| 28-50 | Single level for 28-50 pounds (i.e., 28, 29, 30, …, 50 pounds) | — |
| 51 | 51 or more pounds | — |

## Examples of SUDAAN, Stata, and SAS Code

This section provides some examples of SUDAAN, SAS, and Stata code for generating estimates (means along with SEs) (see Exhibits A.1 to A.5) from the combined NSDUH PUF. Additionally, three tables showing estimates for tobacco product use, any mental illness (AMI), and use of prescription psychotherapeutics are included, using the combined 2002-2019 PUF (Tables A.5 to A.7). PUF users can ensure that PUF data are downloaded correctly, or that code is implemented correctly, by replicating estimates in these tables.

Exhibit A.1 Using SUDAAN DESCRIPT Procedure to Produce Mean Estimate and Standard Error of Past Month Alcohol Use (ALCMON), by Gender (IRSEX), Using 2002-2019 Combined Public Use File Data

|  |
| --- |
| PROC SORT DATA=DATANAME; */\*SAS code to sort output dataset by Nesting Variables\*/*BY VESTR VEREP;RUN;  PROC DESCRIPT DATA=DATANAME DDF=170 DESIGN=WR FILETYPE=SAS DEFT4;*/\*The DOF here is based on 2002-2019 combined data. It may change depending on what years the combined data contain and whether or not the combined years cross survey designs. The specific number of degrees of freedom can be computed by counting the unique values of VESTR\*/*  NEST VESTR VEREP;WEIGHT ANALWC18; */\*Use analwc18 because 18 years of NSDUH data are being used.\*/*VAR ALCMON; */\*Past month alcohol analysis variable\*/*SUBGROUP IRSEX; */\*Gender variable, where male=1 & female=2\*/*  LEVELS 2; TABLES IRSEX; */\*by gender\*/*PRINT WSUM NSUM MEAN SEMEAN TOTAL SETOTAL / REPLACE STYLE=NCHS;OUTPUT WSUM MEAN SEMEAN TOTAL SETOTAL NSUM DEFFMEAN /REPLACE  NSUMFMT=F8.0 WSUMFMT=F12.0 MEANFMT=F15.10 SEMEANFMT=F15.10 DEFFMEANFMT=F15.10 TOTALFMT=F12.0 SETOTALFMT=F12.0 FILENAME="OUT.SUDFILE";  TITLE "ESTIMATES OF PAST MONTH ALCOHOL BY GENDER";RUN;  Note: The following CLASS statement could be used in place of SUBGROUP and LEVELS statements in the above example:  *CLASS IRSEX;* |

Exhibit A.2 Using SUDAAN DESCRIPT Procedure to Produce Mean Estimate and Standard Error of Any Mental Illness (AMIYR\_U), by Age Group (CATAGMH), Using 2008‑2019 Combined Public Use File Data

|  |
| --- |
| DATA DATANAME;  SET DATANAME;  YR=YEAR+0; */\*YEAR is a character variable\*/*  IF YR>=2008; */\*get 2008-2019 combined data\*/*  RUN;  PROC SORT DATA=DATANAME; */\*SAS code to sort output dataset by Nesting Variables\*/*BY VESTR VEREP;RUN;  PROC DESCRIPT DATA=DATANAME DDF=110 DESIGN=WR FILETYPE=SAS DEFT4;*/\*The DOF here is based on 2008-2019 combined data. It may change depending on what years the combined data contain and whether or not the combined years cross survey designs. The specific number of degrees of freedom can be computed by counting the unique values of VESTR \*/*  NEST VESTR VEREP;WEIGHT ANALWC12; */\*Use analwc12 because 12 years of NSDUH data are being used.\*/*VAR AMIYR\_U; */\*Any mental illness variable\*/*SUBGROUP CATAGEMH; */\*Adult age group variable that is recoded from AGE2, where 18-25=1, 26-49=2, 50+=3, and 12-17=4. Youths 12-17 are not asked this question, so level 4 is missing; only levels 1, 2, and 3 are of interest\*/*  LEVELS 3;  SUBPOPN CATAGMH <4;TABLES CATAGEMH; */\*by age group\*/*PRINT WSUM NSUM MEAN SEMEAN TOTAL SETOTAL / REPLACE STYLE=NCHS;OUTPUT WSUM MEAN SEMEAN TOTAL SETOTAL NSUM DEFFMEAN /REPLACE  NSUMFMT=F8.0 WSUMFMT=F12.0 MEANFMT=F15.10 SEMEANFMT=F15.10 DEFFMEANFMT=F15.10 TOTALFMT=F12.0 SETOTALFMT=F12.0 FILENAME="OUT.SUDFILE";  TITLE " ESTIMATES OF ANY MENTAL ILLNESS IN THE PAST YEAR BY AGE GROUP";RUN;  Note: The following CLASS statement could be used in place of SUBGROUP and LEVELS statements in the above example:  *CLASS CATAGEMH;* |

Exhibit A.3 Using SUDAAN DESCRIPT Procedure to Produce Mean Estimate and Standard Error of Breast Cancer (CABREAST) among Women, by Race (NEWRACE2), Using 2015-2019 Combined Public Use File Data

|  |
| --- |
| DATA DATANAME;  SET DATANAME;  YR=YEAR+0; */\*YEAR is a character variable\*/*  IF YR>=2015; */\*get 2015-2019 combined data\*/*  RUN;  PROC SORT DATA=DATANAME; */\*SAS code to sort output dataset by Nesting Variables\*/*BY VESTR VEREP;RUN;  PROC DESCRIPT DATA=DATANAME DDF=50 DESIGN=WR FILETYPE=SAS DEFT4;*/\*The DOF here is based on 2015-2019 combined data. It may change depending on what years the combined data contain and whether or not the combined years cross survey designs. The specific number of degrees of freedom can be computed by counting the unique values of VESTR \*/*  NEST VESTR VEREP;WEIGHT ANALWC5; */\*Use analwc5 because 5 years of NSDUH data are being used.\*/*VAR CABREAST\_r; */\*Breast Cancer: recode variable CABREAST to combine the yes and logically assigned yes as YES, and all other levels as NO\*/*SUBGROUP NEWRACE2; */\*Race categories\*/*  LEVELS 7; SUBPOPN IRSEX=2;  TABLES NEWRACE2; */\*by race category\*/*PRINT WSUM NSUM MEAN SEMEAN TOTAL SETOTAL / REPLACE STYLE=NCHS;OUTPUT WSUM MEAN SEMEAN TOTAL SETOTAL NSUM DEFFMEAN /REPLACE  NSUMFMT=F8.0 WSUMFMT=F12.0 MEANFMT=F15.10 SEMEANFMT=F15.10 DEFFMEANFMT=F15.10 TOTALFMT=F12.0 SETOTALFMT=F12.0 FILENAME="OUT.SUDFILE";  TITLE " ESTIMATES OF BREAST CANCER AMONG WOMEN BY RACE";RUN;  Note: The following CLASS statement could be used in place of SUBGROUP and LEVELS statements in the above example:  *CLASS NEWRACE2;* |

Exhibit A.4 Using Stata Commands svy: mean and svy: total to Produce Mean Estimate and Standard Error of Past Month Alcohol Use (ALCMON), by Gender (IRSEX), Using 2002-2019 Combined Public Use File Data

|  |
| --- |
| use using ".\\dataname.dta", clear  */\*Ensure all variables are lower case\*/*  rename \*, lower  */\* ID Nesting variables (VESTR and VEREP) and weight variable (ANALWC18 – person analysis weight for 18 years of combined NSDUH data). The DOF here is based on 2002-2019 combined data. It may change depending on what years the combined data contain and whether or not the combined years cross survey designs. The specific number of degrees of freedom can be computed by counting the unique values of VESTR\*/*  svyset verep [pw=analwc18], strata(vestr) dof(170)  gen total\_out=.  gen setotal=.  gen mean\_out=.  gen semean=.  gen nsum=.  gen wsum=.  gen deffmean=.  */\*Estimated means of past month alcohol use by gender\*/*  */\*Gender variable, where male=1 & female=2\*/*  svy: mean alcmon, over(irsex)  matrix M=e(b) */\*Store mean estimates in matrix M\*/*  matrix S=e(V) */\*Store variances in matrix S\*/*  matrix N=e(\_N) */\*Store sample size in matrix N\*/*  matrix W=e(\_N\_subp) */\*Store weighted sample size in matrix W\*/*  estat effects, deff srssubpop*/\*Obtain design effect\*/*  matrix D=e(deff) */\*Store design effect in matrix D\*/*  */\*Extract values stored in the M, S, N, W, and D matrices defined above to the mean\_out, semean, nsum, wsum, and deffmean variables. The loop ensures that the appropriate values are extracted for each value of gender.\*/*  forvalues j=1/2 { */\* number of gender categories\*/*  replace mean\_out=(M[1,`j’]) if irsex==`j'  replace semean=(sqrt(S[`j’,`j’])) if irsex==`j’  replace nsum=(N[1,`j’]) if irsex==`j’  replace wsum=(W[1,`j’]) if irsex==`j’  replace deffmean=(D[1,`j’]) if irsex==`j’  } |

Exhibit A.4 Using Stata Commands svy: mean and svy: total to Produce Mean Estimate and Standard Error of Past Month Alcohol Use (ALCMON), by Gender (IRSEX), Using 2002-2019 Combined Public Use File Data (continued)

|  |
| --- |
| */\*Estimated Totals\*/*  svy: total alcmon, over(irsex)    matrix M=e(b) */\*Store total estimates in matrix M\*/*  matrix S=e(V) */\*Store variances in matrix S\*/*  */\*Extract values stored in the M and S matrices defined above to the total\_out and setotal variables. The loop ensures that the appropriate values are extracted for value of gender.\*/*  forvalues j=1/2 { */\* number of gender categories\*/*  replace total\_out=(M[1,`j’]) if irsex==`j’  replace setotal=(sqrt(S[`j’,`j’])) if irsex==`j’  }  keep wsum mean\_out semean total\_out setotal nsum deffmean irsex  duplicates drop irsex, force */\*keep one record per subpopulation*  *of interest\*/*  */\*Format wsum, mean\_out, semean, total\_out, setotal, nsum, and deffmean variables to control appearance in output.\*/*  format wsum %-12.0fc  format mean\_out %-15.10f  format semean %-15.10f  format total\_out %-12.0fc  format setotal %-12.0fc  format nsum %-8.0fc  format deffmean %-15.10f  */\*Estimates of past month alcohol by year and gender\*/*  list irsex wsum nsum mean\_out semean total\_out setotal |

Exhibit A.5 Using SAS SURVEYMEANS Procedure to Produce Mean Estimate and Standard Error of Past Month Alcohol Use (ALCMON), by Gender (IRSEX), Using 2002-2019 Combined Public Use File Data

|  |
| --- |
| PROC SURVEYMEANS DATA=DATANAME SUMWGT NOBS MEAN SUM;  STRATA VESTR; */\*Nesting variable - strata\*/*  CLUSTER VEREP; */\*Nesting variable - PSU\*/*  WEIGHT ANALWC18; */\*Use analwc18 since we are using 18 years of NSDUH data.\*/*VAR ALCMON; */\*Past month alcohol analysis variable\*/*  DOMAIN IRSEX; */\*Gender variable, where male=1 & female=2\*/*  ODS OUTPUT DOMAIN=OUT.SASFILE;  RUN; |

191112

Table A.5 Tobacco Product Use in the Lifetime, Past Year, and Past Month among Individuals Aged 12 or Older, by Demographic Characteristics: Percentages and Standard Errors, 2002-2019 NSDUH Combined Public Use File Estimates

| Demographic Characteristic | LIFETIME | | PAST YEAR | | PAST MONTH | |
| --- | --- | --- | --- | --- | --- | --- |
| Estimate | SE | Estimate | SE | Estimate | SE |
| **TOTAL** | 67.4 | 0.08 | 31.8 | 0.08 | 26.4 | 0.08 |
| **GENDER** |  |  |  |  |  |  |
| Male | 74.9 | 0.11 | 39.1 | 0.12 | 32.5 | 0.12 |
| Female | 60.2 | 0.11 | 24.9 | 0.11 | 20.7 | 0.10 |
| **HISPANIC ORIGIN AND RACE** |  |  |  |  |  |  |
| Not Hispanic or Latino | 69.9 | 0.09 | 32.8 | 0.10 | 27.6 | 0.09 |
| White | 74.5 | 0.09 | 34.2 | 0.11 | 28.7 | 0.11 |
| Black or African American | 56.8 | 0.26 | 30.7 | 0.22 | 26.6 | 0.21 |
| American Indian or Alaska Native | 75.2 | 0.79 | 46.9 | 0.96 | 40.5 | 0.94 |
| Native Hawaiian or Other Pacific Islander | 59.8 | 1.47 | 32.2 | 1.45 | 26.7 | 1.31 |
| Asian | 38.4 | 0.45 | 15.7 | 0.30 | 11.9 | 0.26 |
| Two or More Races | 69.9 | 0.51 | 39.0 | 0.60 | 33.1 | 0.55 |
| Hispanic or Latino | 53.1 | 0.21 | 25.8 | 0.20 | 19.9 | 0.19 |

SE = standard error that is associated with the public use file (PUF) estimate.

NOTE: The combined NSDUH PUFs are available for download at <https://datafiles.samhsa.gov/>.

NOTE: The combined PUF weight variable ANALWT18 is used to produce these estimates based on combined 2002-2019 data (18 years).

NOTE: Tobacco products include cigarettes, smokeless tobacco (i.e., snuff, dip, chewing tobacco, or snus), cigars, or pipe tobacco. Tobacco product use in the past year excludes past year pipe tobacco use but includes past month pipe tobacco use.

NOTE: No precision-based suppression rules have been applied to this table. Users are encouraged to apply appropriate suppression rules as needed.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2002-2019.

191112

Table A.6 Any Mental Illness in the Past Year among Adults Aged 18 or Older, by Age Group and Demographic Characteristics: Percentages and Standard Errors, 2008-2019 NSDUH Combined Public Use File Estimates

| Demographic Characteristic | TOTAL | | AGED 18-25 | | AGED 26-49 | | AGED 50+ | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Estimate | SE | Estimate | SE | Estimate | SE | Estimate | SE |
| **TOTAL** | 18.6 | 0.10 | 21.6 | 0.13 | 21.7 | 0.13 | 14.8 | 0.15 |
| **GENDER** |  |  |  |  |  |  |  |  |
| Male | 14.8 | 0.13 | 16.8 | 0.18 | 17.3 | 0.17 | 11.6 | 0.22 |
| Female | 22.2 | 0.13 | 26.4 | 0.17 | 26.0 | 0.17 | 17.5 | 0.22 |
| **HISPANIC ORIGIN AND RACE** |  |  |  |  |  |  |  |  |
| Not Hispanic or Latino | 19.1 | 0.11 | 22.3 | 0.15 | 23.1 | 0.14 | 14.8 | 0.17 |
| White | 19.8 | 0.12 | 23.7 | 0.18 | 24.8 | 0.16 | 15.1 | 0.18 |
| Black or African American | 16.4 | 0.24 | 16.6 | 0.26 | 18.6 | 0.35 | 13.9 | 0.46 |
| American Indian or Alaska Native | 21.8 | 0.98 | 22.1 | 1.18 | 25.3 | 1.49 | 18.1 | 1.69 |
| Native Hawaiian or Other Pacific Islander | 20.3 | 1.86 | 17.9 | 1.70 | 22.3 | 2.11 | 18.5 | 3.85 |
| Asian | 13.4 | 0.35 | 20.3 | 0.52 | 14.1 | 0.43 | 9.2 | 0.80 |
| Two or More Races | 27.2 | 0.70 | 29.2 | 0.80 | 30.7 | 0.91 | 22.9 | 1.43 |
| Hispanic or Latino | 15.9 | 0.22 | 18.9 | 0.26 | 15.8 | 0.28 | 13.9 | 0.54 |
| **HEALTH INSURANCE1** |  |  |  |  |  |  |  |  |
| Private | 16.3 | 0.12 | 21.8 | 0.17 | 19.0 | 0.14 | 12.3 | 0.20 |
| Medicaid/CHIP2 | 29.1 | 0.28 | 22.7 | 0.26 | 32.5 | 0.34 | 28.7 | 0.62 |
| Other3 | 17.2 | 0.18 | 22.5 | 0.44 | 31.7 | 0.50 | 14.9 | 0.19 |
| No Coverage | 21.2 | 0.28 | 20.1 | 0.27 | 22.5 | 0.35 | 18.6 | 0.72 |

SE = standard error that is associated with the public use file (PUF) estimate.

NOTE: The combined NSDUH PUFs are available for download at <https://datafiles.samhsa.gov/>.

NOTE: The combined PUF weight variable ANALWT12 is used to produce these estimates based on combined 2008-2019 data (12 years).

NOTE: Any mental illness (AMI) is defined as having a diagnosable mental, behavioral, or emotional disorder, other than a developmental or substance use disorder, assessed by the Mental Health Surveillance Study (MHSS) *Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition—Research Version—Axis I Disorders* (MHSS-SCID), which is based on the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). These mental illness estimates are based on a predictive model and are not direct measures of diagnostic status. For details on the methodology, see Section 3.4.7 in the *2019 National Survey on Drug Use and Health: Methodological Summary and Definitions* at <https://www.samhsa.gov/data/>.

NOTE: No precision-based suppression rules have been applied to this table. Users are encouraged to apply appropriate suppression rules as needed.

1 Respondents could indicate multiple types of health insurance. Thus, these response categories are not mutually exclusive.

2 CHIP is the Children's Health Insurance Program. Individuals aged 19 or younger are eligible for this plan.

3 Other Health Insurance is defined as having Medicare, CHAMPUS, TRICARE, CHAMPVA, the VA, military health care, or any other type of health insurance.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2019.

191112

Table A.7 Any Use of Prescription Psychotherapeutics in the Past Year among Individuals Aged 12 or Older, by Age Group and Demographic Characteristics: Percentages and Standard Errors, 2015-2019 NSDUH Combined Public Use File Estimates

| Demographic Characteristic | AGED 12-17 | | AGED 18-25 | | AGED 26+ | |
| --- | --- | --- | --- | --- | --- | --- |
| Estimate | SE | Estimate | SE | Estimate | SE |
| **TOTAL** | 24.6 | 0.21 | 40.0 | 0.25 | 44.2 | 0.17 |
| **GENDER** |  |  |  |  |  |  |
| Male | 24.4 | 0.26 | 36.7 | 0.37 | 40.5 | 0.25 |
| Female | 24.8 | 0.33 | 43.4 | 0.35 | 47.7 | 0.24 |
| **HISPANIC ORIGIN AND RACE** |  |  |  |  |  |  |
| Not Hispanic or Latino | 25.1 | 0.22 | 41.6 | 0.30 | 45.8 | 0.20 |
| White | 26.2 | 0.27 | 45.1 | 0.39 | 48.1 | 0.21 |
| Black or African American | 23.8 | 0.63 | 35.3 | 0.62 | 41.9 | 0.43 |
| American Indian or Alaska Native | 20.5 | 1.87 | 39.6 | 1.86 | 48.8 | 1.87 |
| Native Hawaiian or Other Pacific Islander | 20.1 | 3.58 | 39.9 | 3.92 | 42.5 | 2.19 |
| Asian | 16.7 | 0.79 | 23.9 | 0.90 | 25.3 | 0.73 |
| Two or More Races | 27.8 | 1.10 | 46.3 | 1.48 | 52.4 | 1.54 |
| Hispanic or Latino | 22.9 | 0.48 | 34.5 | 0.51 | 35.3 | 0.48 |
| **EDUCATION** |  |  |  |  |  |  |
| < High School | da | da | 36.9 | 0.63 | 39.3 | 0.49 |
| High School Graduate | da | da | 37.2 | 0.47 | 43.8 | 0.38 |
| Some College/Associate's Degree | da | da | 43.3 | 0.38 | 49.7 | 0.28 |
| College Graduate | da | da | 39.9 | 0.70 | 41.7 | 0.30 |
| **CURRENT EMPLOYMENT** |  |  |  |  |  |  |
| Full-Time | da | da | 40.8 | 0.37 | 41.0 | 0.22 |
| Part-Time | da | da | 40.2 | 0.51 | 45.2 | 0.47 |
| Unemployed | da | da | 42.2 | 0.77 | 44.4 | 0.88 |
| Other1 | da | da | 37.6 | 0.49 | 48.7 | 0.38 |

da = does not apply; SE = standard error that is associated with the public use file (PUF) estimate.

NOTE: The combined NSDUH PUFs are available for download at <https://datafiles.samhsa.gov/>.

NOTE: The combined PUF weight variable ANALWT5 is used to produce these estimates based on combined 2015-2019 data (5 years).

NOTE: Prescription psychotherapeutics include pain relievers, tranquilizers, stimulants, or sedatives and do not include over-the-counter drugs.

NOTE: Any use of prescription psychotherapeutics is defined as (a) the use of one's own prescription medication as directed by a doctor or (b) misuse of prescription psychotherapeutics. Misuse of prescription psychotherapeutics is defined as use in any way not directed by a doctor, including use without a prescription of one's own; use in greater amounts, more often, or longer than told; or use in any other way not directed by a doctor. Prescription psychotherapeutics do not include over-the-counter drugs.

NOTE: Prescription psychotherapeutic subtypes were revised in 2016. One effect was the comparability of codeine products between 2015 and 2016.

NOTE: No precision-based suppression rules have been applied to this table. Users are encouraged to apply appropriate suppression rules as needed.

1 The Other Employment category includes students, persons keeping house or caring for children full time, retired or disabled persons, or other persons not in the labor force.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2015-2019.

1. Center for Behavioral Health Statistics and Quality. (2017). *National Survey on Drug Use and Health: 2016 public use file and codebook*. Retrieved from <https://datafiles.samhsa.gov/> [↑](#footnote-ref-1)
2. Center for Behavioral Health Statistics and Quality. (2016). *National Survey on Drug Use and Health: 2015 public use file and codebook*. Retrieved from <https://datafiles.samhsa.gov/> [↑](#footnote-ref-2)
3. Center for Behavioral Health Statistics and Quality. (2015). *2014 National Survey on Drug Use and Health: Methodological resource book (Section 2, Sample design report)*. Retrieved from <https://www.samhsa.gov/data/> [↑](#footnote-ref-3)
4. Center for Behavioral Health Statistics and Quality. (2019). *2017 National Survey on Drug Use and Health: Methodological resource book (Section 13, Statistical inference report).* Retrieved from <https://www.samhsa.gov/data/> [↑](#footnote-ref-4)
5. RTI International. (2013). *SUDAAN® language manual, Release 11.0.1*. Research Triangle Park, NC: Author. [↑](#footnote-ref-5)