SAMHSA
Substance Abuse and Mental Health Services Administration

2018 National Survey on Drug Use and Health (NSDUH): Methodological Resource Book

Section 13: Statistical Inference Report

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

March 2020

2018 National Survey on Drug Use and Health (NSDUH): Methodological Resource Book

Acknowledgments

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

Public Domain Notice

All material appearing in this report is in the public domain and may be reproduced or copied without permission from SAMHSA. Citation of the source is appreciated. However, this publication may *not* be reproduced or distributed for a fee without the specific, written authorization of the Office of Communications, SAMHSA, HHS.

Electronic Access

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

Recommended Citation

Recommended Citation: Center for Behavioral Health Statistics and Quality. (2020). *2018 National Survey on Drug Use and Health (NSDUH): Methodological Resource Book, Section 13, Statistical Inference Report*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Originating Office

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

Nondiscrimination Notice

SAMHSA complies with applicable federal civil rights laws and does not discriminate on the basis of race, color, national origin, age, disability, or sex. SAMHSA cumple con las leyes federales de derechos civiles aplicables y no discrimina por motivos de raza, color, nacionalidad, edad, discapacidad o sexo.

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

March 2020

Table of Contents

Chapter **Page**

1. Introduction ... 1
2. Background ... 3
   - 2.1 Sample Design ... 3
   - 2.2 Changes to Questionnaire Content and Survey Methodology ... 6
     - 2.2.1 2017 Questionnaire Changes ... 7
     - 2.2.2 2018 Questionnaire Changes ... 7
3. Prevalence Estimates ... 9
   - 3.1 Adult Major Depressive Episode ... 10
   - 3.2 Serious Psychological Distress ... 10
   - 3.3 Mental Illness ... 11
   - 3.4 Substance Use Disorders ... 16
   - 3.5 Substance Use Treatment ... 17
   - 3.6 Perceptions of Risk and Availability ... 18
   - 3.7 Prescription Drug Subtypes ... 18
   - 3.8 Adult Mental Health Outpatient Treatment ... 19
   - 3.9 Youth Reasons for Receiving Mental Health Services ... 20
   - 3.10 Measures of Benzodiazepine Use and Misuse ... 20
   - 3.11 Measures of Tranquilizer or Sedative Use and Misuse ... 22
   - 3.12 Decennial Census Effects on NSDUH Substance Use and Mental Health Estimates ... 22
   - 3.13 Using Revised Estimates for 2006 to 2010 ... 23
4. Missingness ... 25
   - 4.1 Potential Estimation Bias Due to Missingness ... 25
   - 4.2 Variance Estimation in the Presence of Missingness ... 28
5. Sampling Error ... 29
6. Degrees of Freedom ... 35
   - 6.1 Background ... 35
   - 6.2 Degrees of Freedom Used in Key NSDUH Analyses ... 38
7. Statistical Significance of Differences ... 41
   - 7.1 Impact of the 2014 Sample Redesign on Significance Testing between Years ... 41
   - 7.2 Impact of 2017 Data Quality Improvements on Significance Testing between Years ... 42
   - 7.3 Impact of 2018-2022 Sample Design Continuation on Significance Testing between Years ... 43
   - 7.4 Comparing Prevalence Estimates between Years ... 43
   - 7.5 Example of Comparing Prevalence Estimates between Years ... 44
   - 7.6 Example of Comparing Prevalence Estimates between Years in Excel ... 45
   - 7.7 Comparing Prevalence Estimates in Categorical Subgroups ... 47
   - 7.8 Comparing Prevalence Estimates to Identify Linear Trends ... 47
   - 7.9 Impact of Rounding in Interpreting Testing Results ... 48

Table of Contents (continued)

Chapter **Page**

8. Confidence Intervals 51
   - 8.1 Example of Calculating Confidence Intervals Using Published Prevalence Estimates and Standard Errors 53
   - 8.2 Example of Calculating Confidence Intervals in Excel Using Published Prevalence Estimates and Standard Errors 54
   - 8.3 Example of Calculating Standard Errors Using Published Confidence Intervals 55
   - 8.4 Example of Calculating Standard Errors in Excel Using Published Confidence Intervals 56

9. Initiation Estimates 59
   - 9.1 Initiation of Misuse of Prescription Psychotherapeutic Drugs 60
   - 9.2 Initiation of Use of Substances Other Than Prescription Psychotherapeutic Drugs 61

10. Suppression of Estimates with Low Precision 65

References 71

List of Contributors 77

Appendix

A Documentation for Conducting Various Statistical Procedures: SUDAAN®, Stata®, SAS®, and R Examples 79

List of Tables

Table		Page
2.1	NSDUH Sample Selection, 2014-2022	4
3.1	Final SMI Prediction Models in the 2008-2012 MHSS	14
3.2	Cut Point Probabilities for SMI, AMI, and SMMI, by 2012 Model	14
5.1	Demographic and Geographic Domains Shown in the First Findings Reports and Detailed Tables That Use the Alternative Standard Error Estimation Method for Calculating Standard Errors of the Estimated Number of Persons (Totals), 2018	31
6.1	Ninety-Five Percent Confidence Intervals for the Percentage of Past Month Users of Alcohol, Using Different Degrees of Freedom, 2015	36
6.2	Degrees of Freedom for Specific States per the NSDUH Sample Design Based on the Restricted-Use Dataset	37
6.3	Key NSDUH Analyses and Degrees of Freedom for the Restricted-Use Data File and the Public Use Data File, by Sample Design Years, 2002-2022	39
10.1	Summary of 2018 NSDUH Suppression Rules	67
A.1	Summary of SUDAAN, Stata, SAS, and R Exhibits	79
A.2	SUDAAN Matrix Shell	102
A.3	Stata Matrix Shell	103
A.4	Contrast Statements for Exhibits A.55 and A.56	127

This page intentionally left blank

List of Exhibits

Exhibit		Page
6.1	97.5th Percentiles of *t*-Distributions for Varying Degrees of Freedom	36
10.1	Required Effective Sample in the 2018 NSDUH as a Function of the Proportion Estimated	66
A.1	SUDAAN DESCRIPT Procedure (Estimate Generation: Single Year and Pooled Years of Data)	83
A.2	Stata COMMANDS svy: mean and svy: total (Estimate Generation: Single Year and Pooled Years of Data)	83
A.3	SAS SURVEYMEANS Procedure (Estimate Generation: Single Year and Pooled Years of Data)	86
A.4	R Code: svytotal and svymean (Estimate Generation: Single Year and Pooled Years of Data)	86
A.5	SAS Code Based on SUDAAN Output (Calculation of Standard Error of Totals for Fixed Domains)	89
A.6	Stata Code (Calculation of Standard Error of Totals for Fixed Domains)	89
A.7	SAS Code Based on SAS Output (Calculation of Standard Error of Totals for Fixed Domains)	89
A.8	R Code (Calculation of Standard Error of Totals for Fixed Domains)	90
A.9	SAS Code Based on SUDAAN Output (Implementation of Suppression Rule)	91
A.10	Stata Code (Implementation of Suppression Rule)	91
A.11	SAS Code Based on SAS Output (Implementation of Suppression Rule)	92
A.12	R Code (Implementation of Suppression Rule)	93
A.13	SUDAAN DESCRIPT Procedure (Tests of Differences)	95
A.14	Stata COMMANDS svy: mean and svy: total (Tests of Differences)	96
A.15	SAS Code (Tests of Differences)	99
A.16	R Code (Tests of Differences)	99
A.17	SAS Code Based on SUDAAN Output (Calculation of the *P* Value for the Test of Differences between Totals for Nonfixed Domains)	100
A.18	Stata Code (Calculation of the *P* Value for the Test of Differences between Totals for Nonfixed Domains)	100

List of Exhibits (continued)

Exhibit		Page
A.19	SAS Code (Calculation of the *P* Value for the Test of Differences between Totals for Nonfixed Domains)	100
A.20	R Code (Calculation of the *P* Value for the Test of Differences between Estimated Number Totals for Nonfixed Domains)	100
A.21	SUDAAN DESCRIPT Procedure (Covariance Matrix)	101
A.22	Stata COMMAND svy: mean (Covariance Matrix)	102
A.23	SAS Code Based on SUDAAN Output (Identification of Covariance Components)	103
A.24	Stata Code (Identification of Covariance Components)	104
A.25	SAS Code Based on SUDAAN Output (Calculation of Variances)	104
A.26	Stata Code (Calculation of Variances)	105
A.27	SAS Code Based on SUDAAN Output (Calculation of the *P* Value for the Test of Differences between Totals for Fixed Domains)	105
A.28	Stata Code (Calculation of the *P* Value for the Test of Differences between Totals for Fixed Domains)	106
A.29	SAS Code (Covariance Matrix, Calculations of Variances, and Calculation of the *P* Value for the Test of Differences between Totals for Fixed Domains)	106
A.30	R Code (Covariance Matrix, Calculations of Variances, and Calculation of the *P* Value for the Test of Differences between Totals for Fixed Domains)	107
A.31	SAS Code (Recoding a Variable) and SUDAAN DESCRIPT Procedure (Estimate Generation with (1) Missing Values and (2) Using Subpopulation)	109
A.32	Stata Code (Recoding a Variable, Estimate Generation with (1) Missing Values and (2) Using Subpopulation)	110
A.33	SAS Code (Recoding a Variable, Estimate Generation with (1) Missing Values and (2) Using Subpopulation)	111
A.34	R Code (Estimate Generation with (1) Missing Values and (2) Using Subpopulation)	112
A.35	SAS Code Based on SUDAAN Output (Calculating a 95 Percent Confidence Interval)	113

List of Exhibits (continued)

Exhibit		Page
A.36	Stata Code (Calculating a 95 Percent Confidence Interval for a Mean)	113
A.37	SAS Code (Calculates a CI for Alcohol Drinker Prevalence and Estimated Totals Produced in Exhibit A.3)	114
A.38	R Code (Calculates a CI for Alcohol Drinker Prevalence and Estimated Totals Produced in Exhibit A.4)	114
A.39	SUDAAN DESCRIPT Procedure (Frequency of Use; i.e., Number of Days Used Substance in the Past Month among Past Month Users)	115
A.40	Stata Code (Frequency of Use; i.e., Number of Days Used Substance in the Past Month among Past Month Users)	116
A.41	SAS (Frequency of Use; i.e., Number of Days Used Substance in the Past Month among Past Month Users)	116
A.42	R (Frequency of Use; i.e., Number of Days Used Substance in the Past Month among Past Month Users)	116
A.43	SAS Code (Stacking a Dataset) and SUDAAN DESCRIPT Procedure (Test of Difference when Two Groups Overlap Using Stacked Data)	117
A.44	Stata Code (Test of Difference when Two Groups Overlap Using Stacked Data)	118
A.45	SAS Code (Statistical Tests of Differences between Two Groups when the Two Groups Overlap)	119
A.46	R Code (Statistical Tests of Differences between Two Groups when the Two Groups Overlap)	120
A.47	SUDAAN CROSSTAB Procedure (Test for Independence Based on a Log-Linear Model)	121
A.48	Stata Code (Test for Independence Based on a Log-Linear Model)	121
A.49	SAS Code (Tests of the Independence of the Prevalence Variable and Subgroup Variable)	121
A.50	R Code (Tests of the Independence of the Prevalence Variable and Subgroup Variable)	122
A.51	SUDAAN DESCRIPT Procedure (Pairwise Testing)	123
A.52	Stata Code (Pairwise Testing)	123
A.53	SAS Code (Pairwise Testing)	124
A.54	R Code (Pairwise Testing)	125
A.55	SUDAAN DESCRIPT Procedure (Test of Linear Trends with DESCRIPT)	127

List of Exhibits (continued)

Exhibit		Page
A.56	Stata Code (Test of Linear Trends with TEST Statements)	128
A.57	SAS Code (Linear Trend Test of Significance across Years Using Test Statements)	130
A.58	R Code (Linear Trend Test of Significance across Years Using Test Statements)	130
A.59	SUDAAN RLOGIST Procedure (Modeling Test of Linear Trends)	132
A.60	Stata Code (Modeling Test of Linear Trends)	133
A.61	SAS Code (Linear Trend Test of Significance across Years Using Modeling)	133
A.62	R Code (Linear Trend Test of Significance across Years Using Modeling)	134

1. Introduction

Statistical inference occurs whenever data obtained from sample observations belonging to and considered representative of a larger target population are used to make generalizations concerning the larger population. The target population for the 2018 National Survey on Drug Use and Health (NSDUH)[1] was the U.S. civilian, noninstitutionalized population aged 12 or older (at the time of their interview) in 2018. Measurements for this target population were the responses to the survey questions provided by people participating in the 2018 survey. Examples of conducting statistical inference include using the weighted estimate and the corresponding standard error of the number of users of illicit drugs[2] based on a sample to make a statement about the number of users in the U.S. civilian, noninstitutionalized population. Another example is conducting a significance test to determine whether the percentage of adults with serious mental illness increased over time.

Statistical inferences concerning characteristics of interest for this population and various subpopulations are presented in the form of estimates (number of people and associated prevalence estimates) derived from the sample data collected. Examples of the inferences made from the 2018 NSDUH data are presented in the 2018 detailed tables[3] (Center for Behavioral Health Statistics and Quality [CBHSQ], 2019c) and the 2018 national-level first findings report (FFR) that focuses on key substance use and mental health indicators in the United States (CBHSQ, 2019b). A glossary of key definitions and other supporting information that are relevant to estimates of substance use and mental health issues from the 2018 NSDUH accompanies the 2018 methodological summary and definitions (CBHSQ, 2019a) and the detailed tables (CBHSQ, 2019c).

The focus of this report is to describe the statistical inference procedures used to produce design-based estimates as presented in the 2018 detailed tables and the 2018 FFR, which are based on *restricted-use data*. Users of NSDUH's *public use data* (CBHSQ/Substance Abuse and Mental Health Services Administration, 2019), therefore, may find inconsistencies in the variable names referenced in this report's Appendix A, the information presented in Table 5.1 in Chapter 5, and other specific numbers presented in this report (i.e., degrees of freedom). For examples of statistical analyses using NSDUH public use data and tables presenting estimates for selected measures based on the public use data, see Appendix H in the 2018 NSDUH public use codebook.[4]

The examples in this report are based on various years of NSDUH data, but all the examples are relevant to the 2018 detailed tables and the 2018 FFR. The specific estimates shown in examples within the body of this report are based on the 2015 restricted-use dataset that was used to create the 2015 detailed tables (CBHSQ, 2016c) and the 2015 FFRs (CBHSQ, 2016b; Hughes et al., 2016; Lipari, Forsyth, Bose, Kroutil, & Lane, 2016; Lipari, Williams,

[1] Before 2002, the survey was called the National Household Survey on Drug Abuse.

[2] NSDUH obtains information on the following 10 categories of drugs: marijuana, cocaine (including crack), heroin, hallucinogens, inhalants, and methamphetamine, as well as the misuse of prescription pain relievers, tranquilizers, stimulants, and sedatives. Estimates of "illicit drug use" reported from NSDUH reflect the use of drugs in any of these 10 categories.

[3] Starting with the 2015 NSDUH, the mental health detailed tables were combined with the detailed tables on substance use and other measures.

[4] NSDUH public use files going back to 1979 are available on the Substance Abuse and Mental Health Data Archive, which can be accessed at https://datafiles.samhsa.gov/.

Copello, & Pemberton, 2016; Medley et al., 2016; Park-Lee, Lipari, Hedden, Copello, & Kroutil, 2016; Piscopo, Lipari, Cooney, & Glasheen, 2016). Appendix A's examples are based on the NSDUH restricted-use datasets from 2002 through 2018 showing statistical procedures implemented in the detailed tables and various NSDUH reports. Prior to this report, Appendix A included example code in two and sometimes three programming languages to illustrate how to apply each of the statistical procedures using the NSDUH data. Starting with this report, Appendix A was expanded to include code in four programming languages for each of the statistical procedures. The statistical procedures and information found in this report can also be generally applied to analyses based on the public use file. To emphasize key points for analyzing NSDUH data, certain sentences throughout this report have been italicized.

This report is organized as follows: Chapter 2 provides background information concerning the survey design, including redesign and questionnaire changes; Chapter 3 discusses the prevalence estimates and how they were calculated, including specifics on topics such as major depressive episode, serious psychological distress, mental illness, substance use disorder, substance use treatment, perceptions of risk and availability, prescription drug subtypes, adult mental health outpatient treatment, youth reasons for receiving mental health services, benzodiazepine use and misuse, and tranquilizer or sedative use and misuse; Chapter 4 briefly discusses how missing item responses of variables that are not imputed may lead to biased estimates; Chapter 5 discusses sampling errors and how they were calculated; Chapter 6 describes degrees of freedom and how they were used when comparing estimates; and Chapter 7 discusses how the statistical significance of differences between estimates was determined. Chapter 8 discusses confidence interval estimation, and Chapter 9 describes how past year initiation of drug use was computed. Finally, Chapter 10 discusses the conditions under which estimates with low precision were suppressed. Appendix A contains examples that demonstrate how to conduct various statistical procedures documented within this report. Examples include using SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2013), Stata® (StataCorp LP, 2017), SAS® (SAS Institute Inc., 2017), and R (R Core Team, 2018).

2. Background

The respondent universe for the National Survey on Drug Use and Health (NSDUH) is the civilian, noninstitutionalized population aged 12 or older residing within the 50 states and the District of Columbia. The survey covers residents of households (e.g., individuals living in houses or townhouses, apartments, and condominiums; civilians living in housing on military bases) and individuals living in noninstitutional group quarters (e.g., shelters, rooming/boarding houses, college dormitories, migratory workers' camps, halfway houses). Excluded from the survey are individuals with no fixed household address (e.g., homeless and/or transient people not in shelters), active-duty military personnel, and residents of institutional group quarters, such as correctional facilities, nursing homes, mental institutions, and long-term hospitals.

The final respondent sample of 67,791 people for the 2018 NSDUH provides a sufficient sample to create domain estimates for a broad range of ages, other demographic characteristics, geographic characteristics, and socioeconomic categories. Individual observations are weighted so that the weighted sample represents the civilian, noninstitutionalized population aged 12 or older for the nation as a whole and for each state. *The person-level weights in NSDUH are calibrated by adjusting for nonresponse and poststratifying to known population estimates (or control totals) obtained from the U.S. Census Bureau.* For more information on the person-level sampling weight calibration in the *2018 NSDUH Methodological Resource Book*, see Center for Behavioral Health Statistics and Quality (CBHSQ) (2020b).

2.1 Sample Design

A coordinated sample design was developed for the 2014 through 2017 NSDUHs. A large reserve sample was selected at the time the 2014 through 2017 NSDUH sample was selected. This reserve sample was (or will be) used to field the 2018 through 2022 NSDUHs. Thus, the 2018 through 2022 NSDUH designs simply continue the coordinated design. The 2014 through 2022 coordinated sample design is state based, with an independent, multistage area probability sample within each state and the District of Columbia (see Table 2.1). As a result, states are viewed as the first level of stratification and as a variable for reporting estimates. Each state was further stratified into approximately equally populated state sampling regions (SSRs). The number of SSRs varied by state and was related to the state's sample size. SSRs were contiguous geographic areas designed to yield approximately the same number of interviews within a given state.[5] There was a total of 750 SSRs for 2018. Creation of the multistage area probability sample then involved selecting census tracts within each SSR, census block groups within census tracts, and area segments (i.e., a collection of census blocks) within census block groups. Finally, dwelling units (DUs) were selected within segments, and (within each selected DU) up to two residents who were at least 12 years old were selected for the interview. If two eligible residents within the same DU were selected, they formed a within-DU pair. For more detailed information on the sample selection process, see the 2018 sample design report (CBHSQ, 2019d).

[5] Sampling areas were defined using 2010 census geography. Counts of dwelling units and population totals were obtained from the 2010 decennial census data supplemented with revised population projections from Claritas, a market research firm (see https://www.claritas.com/).

Table 2.1 NSDUH Sample Selection, 2014-2022

Stage of Selection	Unit	Details
Stratification	State	50 States plus the District of Columbia.
Stratification	State sampling region (SSR)	750 SSRs.
First	Census tract	Select 48 systematically per SSR with probability proportional to size (PPS).
Second[1]	Block group	Select 1 per selected census tract with PPS.
Third	Segment	Select 1 per selected block group with PPS. Eight segments per SSR are used each year. Fifty percent of the segments overlap from one year to the next.
Fourth	Dwelling unit (DU)	Systematic selection of DUs within segments.
Fifth	Person	Select 0, 1, or 2 people per household using predetermined State and age group sampling rates.

[1] The 2005-2013 sample selection process excluded this stage of selection.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014-2022.

The coordinated sample design for 2014 through 2022 includes a 50 percent overlap in third-stage units (area segments) within each successive 2-year period from 2014 through 2022. In addition to reducing costs, this designed sample overlap may slightly increase the precision of estimates of year-to-year trends when there is a small but positive correlation in successive survey years due to an overlapped segment being somewhat homogeneous. DUs that are not sampled the first year are eligible for selection the following year. There is no planned overlap of sampled residents; however, individuals may be selected in consecutive years if they move and their new residence is selected the year after their original DU was sampled.

The 2014 through 2022 sample design allocates more interviews to the largest 12 states (compared with the 1999 to 2013 design).[6] For the 2018 NSDUH, the target sample size for the largest 12 states was between 1,500 and 4,560 completed interviews and approximately 960 interviews in each of the remaining 37 states and the District of Columbia. *This design change moved the sample from two state sample size groups (large and small) to essentially five state sample size groups, making the sample sizes more proportional to the state population sizes and improving the precision of NSDUH estimates.* This change also allows for a more cost-efficient sample allocation to the largest states while slightly increasing the sample sizes in smaller states to improve the precision of state estimates by either direct methods (by pooling multiple years of data) or using small area estimation (SAE).[7] Population projections based on the 2010 census and data from the 2006 to 2010 American Community Surveys were used to construct the

[6] In the 1999 to 2013 design, the eight largest states each had a target sample size of 3,600. The remaining states and the District of Columbia each had a sample size of 900. In 2014, the sample design was modified so that the sample size per state was relatively more proportional to the state population. For a full list of target sample size per state in 2013 and from 2014 to 2022, see Table 2.1 in Chapter 2 of the 2018 methodological summary and definitions (CBHSQ, 2019a).

[7] SAE is a hierarchical Bayes modeling technique used to make state-level estimates for 30 measures related to substance use and mental health. For more details, see "2017-2018 National Survey on Drug Use and Health: Guide to State Tables and Summary of Small Area Estimation Methodology" at https://www.samhsa.gov/data/.

sampling frame for the 2014 to 2022 NSDUHs. In contrast, projections based on the 2000 census were used in constructing the sampling frame for the 2005 to 2013 NSDUHs.

Similar to the 2005 through 2013 NSDUHs, the first stage of selection for the 2014 through 2022 NSDUHs was census tracts.[8] Within each SSR, 48 census tracts[9] were selected with probability proportional to a composite measure of size.[10] This stage has been included since 2005 to contain sampled areas within a single census tract to the extent possible in order to facilitate merging to external data sources. Within sampled census tracts, adjacent census block groups were combined as necessary to meet the minimum DU size requirements.[11] One census block group or second-stage sampling unit then was selected within each sampled census tract with probability proportional to population size. Compared with the selection process used for the 2005 through 2013 NSDUHs, the selection of census block groups is an additional stage of selection that was included to facilitate possible transitioning to an address-based sampling design in a future survey year (see Table 2.1). For the third stage of selection, adjacent blocks were combined within each sampled census block group to form area segments. One area segment was selected within each sampled census block group with probability proportionate to a composite measure of size.

Although only 20 segments per SSR were needed to support the coordinated 4-year sample for the 2014 through 2017 NSDUHs, an additional 28 segments per SSR were selected to serve as replacements when segment DUs are depleted and/or to support any supplemental studies that the Substance Abuse and Mental Health Services Administration may choose to field.[12] These additional segments are being used to support the 2018 through 2022 NSDUHs. Of the eight sampled segments per SSR, four were used in the previous year and were selected for reuse in the current year, and four more were selected for use for the first time in the current year and will be used again the following year.

These sampled segments were allocated equally into four separate samples, one for each 3-month period (calendar quarter) during the year. That is, a sample of addresses was selected from two segments per SSR in each calendar quarter so that field data collection occurred essentially year round. *The primary objective of the fourth stage of sample selection (listing units) was to select the minimum number of DUs needed in each segment to meet the targeted sample sizes for all age groups.* After DU selections were made, an interviewer visited each selected DU to obtain a roster of all people residing in the DU. Using the roster information

[8] Census tracts are relatively permanent statistical subdivisions of counties and parishes and provide a stable set of geographic units across decennial census periods.

[9] Some census tracts had to be aggregated in order to meet the minimum DU requirement. In California, Florida, Georgia, Illinois, Michigan, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Texas, and Virginia, this minimum size requirement was 250 DUs in urban areas and 200 DUs in rural areas. In the remaining states and the District of Columbia, the minimum requirement was 150 DUs in urban areas and 100 DUs in rural areas.

[10] The composite measure of size is a weighted population size where the weights are the sampling rates defined for specified age groups.

[11] The minimum DU size requirements for census tracts also were applied to census block groups. The purpose of the minimum DU size is to ensure that each sampled area has a sufficient number of DUs to field two NSDUH samples and one field test.

[12] Eight segments per SSR were needed to field the 2014 through 2017 NSDUHs (8 segments × 4 years = 32 segments per SSR). For the 2015 through 2017 NSDUHs, half of the segments were carried over from the prior year (4 segments × 3 years = 12 segments per SSR). Thus, 20 unique segments per SSR are needed to field the 4-year sample (32 – 12 = 20).

obtained from an eligible member of the selected DU, zero, one, or two people were selected for the survey. Compared with selecting one eligible person from each selected DU, the selection of zero, one, or two eligible people allows better control of the age group distribution in order to meet targeted sample sizes. Further, the selection algorithm (a modification of the Brewer [1963, 1975] method for selecting samples of size two [Chromy & Penne, 2002]) provides a mechanism for controlling the number of survey-eligible pairs that are selected. Sampling rates were preset by age group and state. Roster information was entered directly into the electronic screening instrument, which automatically implemented the fifth stage of selection based on the state and age group sampling parameters.

Although the overall sample design remained similar, various design elements did change starting with the 2014 NSDUH. Beginning with the 2014 NSDUH redesign, a change was implemented in the allocation of sample by age group. In the 2005 through 2013 NSDUHs, the sample was allocated equally among three age groups: 12 to 17, 18 to 25, and 26 or older. Starting in 2014, the allocation of the NSDUH sample is 25 percent for adolescents aged 12 to 17, 25 percent for adults aged 18 to 25, and 50 percent for adults aged 26 or older. The sample of adults aged 26 or older is further divided into three subgroups: aged 26 to 34 (15 percent), aged 35 to 49 (20 percent), and aged 50 or older (15 percent). These age allocation changes were designed to reflect more closely the actual population distributions by state and age group, so the precision of estimates overall and for older age groups could be improved. *The sample redesign is not expected to affect the prevalence estimates of outcome variables, but the nature of the design changes is expected to affect the precision of those estimates.* Additionally, changes in the sample design with respect to age group and state necessitated a review of the pair sampling strategy (Chromy & Penne, 2002); therefore, the number of pairs (i.e., two eligible residents within the same DU were selected for the interview) selected for the 2014 through 2018 surveys would be reduced from what was selected in surveys before 2014.

2.2 Changes to Questionnaire Content and Survey Methodology

NSDUH's primary purpose is to measure the prevalence and correlates of substance use and mental health issues in the United States. A strength of NSDUH is the stability of its sample and survey design, which allows for trend analysis and for multiple years of data to be combined to examine specific subgroups. Over time, changes have been made to the questionnaire and data collection procedures to help improve data quality; however, these changes could have also affected the ability to trend NSDUH estimates or combine years of NSDUH data. The next two subsections describe the 2017 and 2018 questionnaire changes and evaluations done to determine whether there were any implications on comparability of variables due to the changes. In general, depending on the topic, NSDUH estimates can be compared back to 2002, 2008, or 2015. Starting in 2002, changes to the NSDUH methodology were implemented and affected the comparability of the 2002 estimates with prior surveys. The 2008 NSDUH started a baseline for many mental health measures, such as mental illness and suicide. The 2015 NSDUH underwent a partial redesign and data collection changes, creating another new baseline for many measures. For more information on the 2002, 2008, and 2015 surveys, see Appendix C of the 2004 NSDUH report on national findings (Office of Applied Studies [OAS], 2005), Section B.4.4 of the 2008 NSDUH report on national findings (OAS, 2009b), and Sections B.4.1 and B.4.2 in Section B of the 2015 NSDUH methodological summary and definitions (CBHSQ, 2016a).

2.2.1 2017 Questionnaire Changes

For the 2017 NSDUH, several changes were made to the questionnaire, primarily to improve the quality of its data. Notable changes, as well as the effects on the 2017 detailed tables (CBHSQ, 2018b), are summarized below. Descriptions of additional changes to the 2017 NSDUH questionnaire can be found in the 2017 questionnaire specifications available at https://www.samhsa.gov/data/. In summary, the changes to the questionnaire in 2017 did not cause a break in trends with 2016 estimates.

The logic was updated for identifying past month alcohol users to determine whether respondents were eligible to be asked questions in later sections of the interview. Respondents who originally reported last using alcohol in the past 30 days but who reported in a subsequent question that they used alcohol on 0 days in that period were no longer defined as being past month alcohol users and received questions that did not necessarily apply to them. Consequently, such respondents in 2017 were not asked questions about the misuse of prescription drugs with alcohol in the past 30 days or questions in the consumption of alcohol section that applied to past month alcohol users. Published estimates for the misuse of prescription drugs with alcohol in the past 30 days are not presented in NSDUH reports and tables. An impact assessment determined there were not significant differences between 2016 and 2017 for estimates of the misuse of prescription drugs with alcohol in the past 30 days among past month prescription drug misusers aged 12 or older.

Additionally, text was added to questions about the misuse of "any other" prescription pain reliever, stimulant, or sedative to remind respondents not to include over-the-counter (OTC) drugs (e.g., Tylenol®, Dexatrim®, Sominex®). Corresponding text was not added for the misuse of any other tranquilizer because OTC drugs are not available for tranquilizers in the United States. Contrary to expectations, the addition of the instructions for respondents not to include OTC drugs when answering questions about the misuse of any other pain reliever, stimulant, or sedative *increased* the reports of misuse of OTC drugs. An impact assessment determined the estimates of the misuse of "other" pain relievers and "other" stimulants increased between 2016 and 2017; however, this increase did not appear to affect overall estimates for the misuse of any prescription pain relievers or stimulants. There were no increases in "other" sedative misuse or the overall sedative estimates between 2016 and 2017. For more detailed information, see Section C.5.2.2 in the 2017 methodological summary and definitions (CBHSQ, 2018a).

Finally, a minor correction was made to the logic for determining respondents' eligibility to be asked questions about alcohol use disorder in the substance dependence and abuse section. The logic was updated so that respondents who estimated the number of days that they drank alcohol in the past 30 days were eligible to be asked the alcohol use disorder questions if they estimated that they drank alcohol on more than 5 days in the past 30 days instead of on more than 2 days in that period. After an impact assessment, it was determined that the change in logic did not affect estimates in the 2017 detailed tables.

2.2.2 2018 Questionnaire Changes

For the 2018 NSDUH, several changes were made to the questionnaire, including logic updates to improve data quality and adding more questions. Notable changes, as well as the effects on the 2018 detailed tables (CBHSQ, 2019c), are summarized below. Descriptions of additional changes to the 2018 NSDUH questionnaire can be found in the 2018 questionnaire

specifications that are available at https://www.samhsa.gov/data/. In summary, the changes to the questionnaire in 2018 did not cause a break in trends with 2017 and prior year estimates.

The logic for identifying past month initiates for the misuse of individual prescription drugs (e.g., PR30INIT1 for initiation of Vicodin® misuse in the past 30 days) was updated in 2018. Respondents who reported that they first misused specific prescription drugs in the past 30 days (e.g., first misuse of the pain reliever Vicodin® at the respondent's current age, and the respondent had his or her last birthday within 30 days of the interview) were intended to be skipped out of the question about the misuse of any prescription drug in that category (e.g., pain relievers) in the past 30 days. These respondents logically misused prescription drugs in the past 30 days. Therefore, the computer-assisted interviewing logic was updated to not permit these respondents to contradict these data by reporting that they did not misuse any prescription drug in that category in the past 30 days. Similar updates were made to the 30-day initiation variables for the remaining prescription drugs in the 2018 instrument. An impact assessment determined there were not significant differences between 2017 and 2018 for the initiation of prescription drug misuse in the past 30 days. NSDUH reports discuss initiation of prescription drug misuse in the past 12 months but not the past 30 days, so there was no impact on the published estimates.

Four new questions for adult respondents aged 18 or older were added to the end of the consumption of alcohol section for 2018, asking about a perceived substance use problem or a mental health problem and perceived recovery from each. An assessment of the data indicated no impact on responses to questions that come after these items because the section on market information for marijuana following these new questions was also new for 2018, having been removed from the 2015 to 2017 questionnaires. Adolescent estimates for the market information for marijuana section would not be affected by the addition of these four new questions because only adults 18 or older were asked the perceived problem and recovery questions. Based on review of the 6-month data, estimates for these items were not added to the 2018 detailed tables, but estimates were included as an appendix table in the 2018 first findings report (FFR) (CBHSQ, 2019b).

Additionally, the market information for marijuana section that was last included in the 2014 NSDUH was added back to the 2018 NSDUH. The content of this section did not change relative to 2014. To reduce the potential for context effects that could disrupt trends for other measures, the section was added after the consumption of alcohol section and before the back-end demographics section. The marijuana market has changed considerably since 2014. An impact assessment showed significant differences between estimates in 2014 and 2018, which more than likely are explained by the amount of time elapsed between estimates rather than by the different placement of these questions in the interview (i.e., after the special topics section in 2014 and after the consumption of alcohol section in 2018). The preceding context of the market information for marijuana section questions indicated no impact on how most respondents answered the back-end demographic questions. Market information for marijuana data were not included as estimates in the detailed tables or in the 2018 FFR (CBHSQ, 2019b).

3. Prevalence Estimates

The national prevalence estimates were computed using a multiprocedural package called SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2013). *The final, nonresponse-adjusted, and poststratified analysis weights were used in SUDAAN to compute unbiased design-based estimates.* Appendix A contains examples that demonstrate how to compute the prevalence estimates as defined below using SUDAAN (Exhibit A.1), Stata® (StataCorp LP, 2017) (Exhibit A.2), SAS® (SAS Institute Inc., 2017) (Exhibit A.3), and R (R Core Team, 2018) (Exhibit A.4). For categorical measures, Exhibits A.39 through A.42 demonstrate how to compute the prevalence estimates.

Prevalence estimates are the proportions of the population who exhibit characteristics of interest (such as substance use). Let $\hat{p}_d$ represent the prevalence estimate of interest for domain *d*. Then $\hat{p}_d$ would be defined as the ratio

$$\hat{p}_d = \frac{\hat{Y}_d}{\hat{N}_d},$$

where $\hat{Y}_d = \sum_{i \in S} w_i \delta_i y_i$ represents the estimated number of people exhibiting the characteristic of interest in domain *d*, $\hat{N}_d = \sum_{i \in S} w_i \delta_i$ represents the estimated population total for domain *d*, *S* represents the sample, w_i represents the analysis weight, δ_i is defined as 1 if the *i*th sample unit is in domain *d* and is equal to 0 otherwise, and y_i is defined as 1 if the *i*th sample unit exhibits the characteristic of interest and is equal to 0 otherwise.

For certain populations of interest, sample sizes may not be adequate to support inferences using only 1 year of survey data. In these cases, estimates can be produced from annual averages based on combined data from 2 or more survey years. The 2018 detailed tables (Center for Behavioral Health Statistics and Quality [CBHSQ], 2019c) did not present any combined data, but combined data may be presented in future detailed tables. *The annual averages can be derived by concatenating the data for the respective years and dividing the analysis weights by a factor that varies depending on the number of years of concatenated data (see Exhibits A.1 through A.4).* For example, the weight would be divided by a factor of 2 for 2 years of concatenated data and a factor of 4 for 4 years of concatenated data.

Prevalence estimates are presented in the 2018 detailed tables (CBHSQ, 2019c) in the form of numbers in thousands and percentages rounded to the nearest tenth of a percent. For percentages, rounding an estimate close to zero to the nearest tenth of a percent, which has not been suppressed per the National Survey on Drug Use and Health (NSDUH) suppression rules (see Chapter 10), may result in an estimate of 0.0 percent being displayed in a table. Consequently, the corresponding population total presented in thousands may result in a 0 (i.e., 499 or fewer individuals) being displayed in a table. *Thus, users are reminded that a percentage of 0.0 or a number in thousands of 0 are not exact zeros but are unsuppressed, nonzero estimates that should not be interpreted as no respondents in the population of interest.* Note that in other NSDUH publications, the unsuppressed rounded prevalence estimate of 0.0 percent may be shown as < 0.05 percent, and an unsuppressed rounded number in thousands estimate of 0 may

be shown as < 500. If an estimate is exactly a 0 value, corresponding to no respondents in the sample, the percentage and the number in thousands will be suppressed under the NSDUH suppression rule.

3.1 Adult Major Depressive Episode

The past year adult major depressive episode (MDE) estimates shown in the 2018 detailed tables (CBHSQ, 2019c) are based on the full sample as was done in the 2010 to 2014 mental health detailed tables (CBHSQ, 2012a, 2012c, 2013b, 2014c, 2015b) and the 2015 to 2017 detailed tables (CBHSQ, 2016c, 2017d, 2018b). This differs from the 2008 past year MDE estimates shown in the 2008 detailed tables (Office of Applied Studies [OAS], 2009a) and the 2009 mental health detailed tables (CBHSQ, 2010), which were based on only the sample of adult respondents who received the World Health Organization Disability Assessment Schedule (WHODAS) questions in the mental health questionnaire section that preceded the adult depression questionnaire section. The analysis of 2008 MDE data was restricted to only the WHODAS half sample because of apparent reporting differences (context effects) between the half sample of adult respondents who were administered the WHODAS and the other half sample of adult respondents who received the Sheehan Disability Scale (SDS) questions (Dean & LeBaron, 2009).

Both half samples had issues with context effects not seen in 2007 and previous years because of the revisions to the mental health questionnaire section preceding the adult depression questionnaire section. *To address the break in comparability of the adult MDE data beginning in 2008 and to estimate adult MDE based on the full sample of adults from 2008, adjusted versions of lifetime and past year MDE variables for adults were created retroactively for 2005 to 2008.* These variables were adjusted to make MDE estimates from the SDS half sample in 2008 and from all adult respondents for 2005 to 2007 comparable with the MDE estimates based on data from the half sample of adults who received the WHODAS in 2008 and from all adult respondents in later years (2009 onward). *The adjusted data from 2005 to 2008 can be used in conjunction with unadjusted data from later years to estimate trends in adult MDE over the entire period from 2005 onward.* More information about how the statistically adjusted adult MDE variables were created can be found in Section 3.4.8 in Chapter 3 from the 2018 NSDUH methodological summary and definitions (CBHSQ, 2019a) and in the report describing the adjustments (Aldworth, Kott, Yu, Mosquin, & Barnett-Walker, 2012).

The standard analysis weight, ANALWT, was used to generate all estimates of adult MDE in the 2018 detailed tables (CBHSQ, 2019c) and in the 2018 first findings report (FFR) (CBHSQ, 2019b). More information on the analysis weight can be found in Section 2.3.4 in Chapter 2 of the 2018 NSDUH methodological summary and definitions (CBHSQ, 2019a).

3.2 Serious Psychological Distress

The Kessler Psychological Distress Scale (K6) was used to create the serious psychological distress (SPD) variable. Before 2008, the K6 consisted of one set of questions that asked adult respondents about symptoms of psychological distress in the month when they were the most depressed, anxious, or emotionally distressed in the past year. Starting in 2008, the K6 consisted of two sets of questions that asked adult respondents how frequently they experienced symptoms of psychological distress during two different periods: (1) during the past 30 days, and (2) if applicable, the month in the past year when they were at their worst emotionally.

Respondents were asked about this second period only if they indicated that there was a month in the past 12 months when they felt more depressed, anxious, or emotionally stressed than they felt during the past 30 days. Because of this change, past year K6 and SPD estimates from years before 2008 were no longer comparable with estimates from 2008 onward. *To address this comparability issue, adjusted versions of the past year worst K6 total score and past year SPD variables were created for each of the years from 2005 to 2007 to make the 2005-2007 past year K6 scores and past year SPD estimates comparable with their 2008 and subsequent NSDUH counterparts.*

In the 2018 detailed tables (CBHSQ, 2019c), ANALWT was used to generate 2005-2018 estimates of past year SPD and 2017-2018 estimates of past month SPD. The 2018 FFR (CBHSQ, 2019b) did not present SPD estimates. More information about how the adjusted K6 and SPD variables were created can be found in the report describing these adjustments (Aldworth et al., 2012).

3.3 Mental Illness

The Substance Abuse and Mental Health Services Administration (SAMHSA) has been publishing estimates of the prevalence of past year serious mental illness (SMI) and any mental illness (AMI) among adults aged 18 or older since the release of the 2008 NSDUH national findings report (OAS, 2009b). Originally, estimates were based on a prediction model for mental illness developed using the 2008 data from the Mental Health Surveillance Study (MHSS), which was embedded in the 2008 NSDUH (referred to as the 2008 WHODAS model). Each respondent in a subsample of adults (about 1,500 in 2008) who had completed the NSDUH interview was administered the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP) (First, Spitzer, Gibbon, & Williams, 2002).[13] For more specific information on the MHSS sample design, see the sample design report in the *2013 NSDUH Methodological Resource Book* (CBHSQ, 2014d).

The 2008 NSDUH included a split sample, in which half the respondents (approximately 750 MHSS respondents) were administered the WHODAS and the other half were administered the SDS. These samples are referred to as the SDS half sample and the WHODAS half sample. Two models were used to predict SMI for 2008, one for each impairment scale (WHODAS and SDS). The 2008 models for SMI were chosen so that estimates from the WHODAS and SDS samples were approximately equal; hence, SMI estimates for 2008 were based on both samples. *The WHODAS model was determined to be a better predictor of SMI than the SDS model; therefore, starting in 2009, only the WHODAS impairment scale was administered in NSDUH and used for estimating all levels of mental illness. Levels of mental illness include SMI, AMI, low (mild) mental illness (LMI), moderate mental illness (MMI), serious or moderate mental illness (SMMI), and AMI excluding SMI; however, not all measures of mental illness are reported each year.*

Although SAMHSA continued to obtain clinical interviews after 2008, estimates of mental illness from the 2009, 2010, and 2011 NSDUHs were originally based on the WHODAS model developed from the 2008 clinical assessment sample (however, these estimates have since been updated based on a new model; see the next paragraph for details). The same model was

[13] DSM-IV-TR stands for the *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Text Revision (American Psychiatric Association, 2008).

applied to each year's NSDUH data to provide consistency in mental illness comparisons across the years. Producing a new model each year based on the small annual clinical samples (only 500 interviews in 2009 and 2010) would have resulted in large changes in the model parameters and corresponding prevalence estimates because of sampling error, making it impossible to detect real trends in mental illness over time. Furthermore, an evaluation of the 2008 model, using the 2009 NSDUH clinical data, found that the model could not be significantly improved with the additional 500-case 2009 clinical sample. The clinical follow-up study, which started in 2008 and continued until 2012, led to a nationally representative sample of approximately 5,000 cases assigned to the WHODAS questions that were used to develop an improved mental illness prediction model (referred to as the 2012 WHODAS model). This revised and improved model has been used for estimating all levels of mental illness starting with the 2012 NSDUH and incorporates the NSDUH respondent's age and indicators of past year suicidal thoughts and depression, along with the variables that were specified in the 2008 model (e.g., variables for the K6 scale and the WHODAS), leading to more accurate estimates of mental illness (see below for details on the 2012 model and revised methodology).

For the 2012 through 2018 detailed tables (CBHSQ, 2013b, 2014c, 2015b, 2016c, 2017d, 2018b, 2019c),[14] the 2008 and later year mental illness estimates were based on the revised model. As of October 2013, the 2008 to 2011 detailed tables (OAS, 2009a; CBHSQ, 2010, 2012a, 2012c) containing estimates for past year mental illness for adults have been revised based on the 2012 model because the estimates were initially based on the 2008 model. *Thus, long-term trends are available for mental illness measures from the 2008 NSDUH and onward.*

For detailed information on model revisions to the mental illness items, see Section 3.4.7 in Chapter 3 of the 2018 methodological summary and definitions (CBHSQ, 2019a). The SMI measure available for years before 2004 is not comparable with the SMI measure based on the 2012 model, which was the case for the 2008 model SMI measures as well. No mental illness measures are available for the 2004 NSDUH. It should also be noted that there are limitations to the analyses of the mental illness variables that are based on the 2012 prediction model. For more information on this, see the "Using Mental Illness Variables in Analysis" section below.

2012 SMI Prediction Model

The 2012 model is a prediction model for mental illness, and it was used to predict SMI and to estimate prevalence of SMI for the 2018 NSDUH. The prediction model is a weighted logistic regression. The response variable Y was defined so that $Y = 1$ when an SMI diagnosis was positive based on the clinical interview; otherwise, $Y = 0$. If $\mathbf{X}$ is a vector of realized explanatory variables, then the response probability $\pi = \Pr(Y = 1 | \mathbf{X})$ can be estimated using a weighted logistic regression model. Further technical details on the 2012 prediction models and the impact of the revised model on the 2008-2011 estimates are available in other reports (see the *2012 MHSS: Design and Estimation Report* [CBHSQ, 2014a], Section 3.4.7 in Chapter 3 of the 2018 methodological summary and definitions [CBHSQ, 2019a], or the report on revisions to the 2008 estimation procedures [CBHSQ, 2015a]).

The 2012 SMI prediction model was fit with data from 4,912 WHODAS MHSS respondents from 2008 through 2012, excluding one case from 2008 and one case from 2009 that

[14] Mental health detailed tables were published separately for the 2009 to 2014 NSDUHs. The mental health and substance use detailed tables were combined starting with the 2015 NSDUH.

were dropped because of data errors. The final WHODAS calibration model for the 2012 prediction model for SMI was determined as

$$\text{logit}(\hat{\pi}) = \log[\hat{\pi}/(1-\hat{\pi})] = -5.972664 + 0.0873416X_k + 0.3385193X_w + 1.9552664X_s + 1.1267330\,X_m + 0.1059137X_a \quad (1)$$

or

$$\hat{\pi} = \frac{1}{1+\exp[-(-5.972664 + 0.0873416X_k + 0.3385193X_w + 1.9552664X_s + 1.1267330\,X_m + 0.1059137X_a)]},$$

where $\hat{\pi}$ refers to the estimate of the SMI response probability π. The covariates in equation (1) came from the main NSDUH interview data:

X_k = *Alternative Past Year K6 Score*: Past year K6 score of less than 8 recoded as 0; past year K6 score of 8 to 24 recoded as 1 to 17.

X_w = *Alternative WHODAS Score*: WHODAS item score of less than 2 recoded as 0; WHODAS item score of 2 to 3 recoded as 1, then summed for a score ranging from 0 to 8.

X_s = *Serious Thoughts of Suicide in the Past Year*: Coded as 1 if "yes"; coded as 0 otherwise.

X_m = *Past Year MDE*: Coded as 1 if the criteria for past year MDE were met;[15] coded as 0 otherwise.

X_a = *Recoded Age*: Coded as age minus 18 if aged 18 to 30; coded as 12 otherwise.

A cut point probability π_0 was determined, so that if $\hat{\pi} \geq \pi_0$ for a particular respondent, then he or she was predicted to be SMI positive; otherwise, he or she was predicted to be SMI negative. The cut points were chosen so that the weighted numbers of false positives and false negatives in the MHSS dataset were as close to equal as possible. The predicted SMI status for all adult NSDUH respondents was used to compute prevalence estimates of SMI. In the 2012 SMI WHODAS prediction model, the respondent is classified as having past year SMI if the predicted probability of SMI is greater than or equal to 0.260573529 (SMI cutoff point). A respondent is classified as having past year AMI if the predicted probability of SMI is greater than or equal to 0.0192519810 (AMI cutoff point). See Table 3.1 for the model specifications. Table 3.2 contains the cutoff points for other mental illness levels.

[15] In this situation, the past year MDE measure is from the main NSDUH interview (i.e., not from the SCID-I/NP). See Section 3.4.8 in Chapter 3 of the 2018 NSDUH methodological summary and definitions (CBHSQ, 2019a).

Table 3.1 Final SMI Prediction Models in the 2008-2012 MHSS

Sample/Model Parameter	Beta	Beta SE	*T* Statistic	*P* Value	*df*	Wald *p* Value[1]
WHODAS Sample (2008A-2012)						
Intercept	-5.9726640	0.3201	-18.6586	0.0000		
Alt PY K6	0.0873416	0.0248	3.5247	0.0009	1	0.0009
Alt WHODAS	0.3385193	0.0349	9.7034	0.0000	1	0.0000
PY Suicidal Thoughts	1.9552664	0.2164	9.0342	0.0000	1	0.0000
PY MDE	1.1267330	0.2196	5.1308	0.0000	1	0.0000
Age1830	0.1059137	0.0244	4.3380	0.0001	1	0.0001
WHODAS and SDS Samples (2008-2012)[2]						
Intercept	-5.7736246	0.3479	-16.5960	0.0000		
Alt PY K6	0.1772067	0.0190	9.3251	0.0000	1	0.0000
PY Suicidal Thoughts	1.8392433	0.1941	9.4781	0.0000	1	0.0000
PY MDE	1.6428623	0.2119	7.7528	0.0000	1	0.0000
Age1830	0.1231266	0.0259	4.7482	0.0000	1	0.0000

2008A = 2008 WHODAS half sample; Age1830 = recoded age variable; Alt = alternative; *df* = degrees of freedom; K6 = six-item Kessler Psychological Distress Scale; MDE = major depressive episode; MHSS = Mental Health Surveillance Study; PY = past year; SDS = Sheehan Disability Scale; SE = standard error; SMI = serious mental illness; WHODAS = eight-item World Health Organization Disability Assessment Schedule.

[1] The Wald *p* value is obtained from the overall model fitting.

[2] The model is fit over the WHODAS and SDS samples in 2008-2012 but is used only to produce predictions for the 2008 SDS sample.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2012.

Table 3.2 Cut Point Probabilities for SMI, AMI, and SMMI, by 2012 Model

Sample/Mental Illness Level	Cut Point Probability
WHODAS Sample (2008A-2012)	
SMI	0.260573529000
AMI	0.019251981000
SMMI	0.077686285365
WHODAS and SDS Samples (2008-2012)[1]	
SMI	0.236434000
AMI	0.019182625
SMMI	0.066163980

2008A = 2008 WHODAS half sample; AMI = any mental illness; SDS = Sheehan Disability Scale; SMI = serious mental illness; SMMI = serious or moderate mental illness; WHODAS = World Health Organization Disability Assessment Schedule.

[1] The model is fit over the WHODAS and SDS samples in 2008-2012, but the cut point predictions are used only to produce predictions for the 2008 SDS sample.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2012.

Additional levels of mental illness are created using a combination of the defined mental illness measures. These additional levels include MMI, LMI, and AMI excluding SMI. Respondents were defined as having past year MMI if they were classified as having SMMI but classified as not having SMI. Respondents were defined as having past year LMI if they were classified as having AMI but classified as not having SMMI. Note that MMI and LMI are no longer shown in the detailed tables starting with the 2016 NSDUH. Respondents were defined as having past year AMI excluding SMI if they were classified as having LMI or MMI. In some documentation, AMI excluding SMI is referred to as LMI or MMI.

Modified 2012 Model for the 2008 SDS Half Sample

As noted previously, the 2008 NSDUH data included a split sample. Similar to the 2008 model, the revised 2012 model also has an alternative model for the SDS data that was fit with data from the complete 2008-2012 MHSS clinical sample that contains 5,653 MHSS respondents, excluding 4 cases from 2008 (1 from the WHODAS half sample and 3 from the SDS half sample) and 1 case from 2009 that were dropped because of data errors.

The modified 2012 SMI prediction model for the SDS half sample was

$$\text{logit}(\hat{\pi}) = \log[\hat{\pi} / (1 - \hat{\pi})] = -5.7736246 + 0.1772067X_k + 1.8392433X_s + 1.6428623X_m + 0.1231266X_a \qquad (2)$$

or

$$\hat{\pi} = \frac{1}{1 + \exp[-(-5.7736246 + 0.1772067X_k + 1.8392433X_s + 1.6428623X_m + 0.1231266X_a)]}.$$

All the covariates in equation (2) also appeared in equation (1).

Similar to the WHODAS model, a cut point probability π_0 was determined, so if $\hat{\pi} \geq \pi_0$ for a particular respondent, then he or she was predicted to be SMI positive; otherwise, he or she was predicted to be SMI negative. The cut points were chosen so that the weighted numbers of false positives and false negatives in the MHSS dataset were as close to equal as possible. In the 2012 SMI SDS half sample prediction model, the respondent is classified as having past year SMI if the predicted probability of SMI is greater than or equal to 0.236434 (SMI cutoff point). Although the SDS half sample prediction model was fit across all years, and the cutoff points were determined based on all years, the cutoff points were used only for the main study respondents in the 2008 SDS sample B to predict the SMI positives. See Tables 3.1 and 3.2.

Weights Used for Estimates of Mental Illness

For the 2008 NSDUH, although SMI data for both half samples (SDS and WHODAS) could be analyzed together when using the 2008 model, the AMI, SMMI, LMI, MMI, and AMI excluding SMI data from the two half samples could not be combined for analysis. Under the 2012 model, both the 2008 half samples can be combined to analyze SMI and the other levels of mental illness because the 2012 models were generated so that the estimates would be comparable between the two half samples.

Mental illness measures (e.g., SMI, AMI, SMMI, AMI excluding SMI)[16] *that are defined based on the 2012 model should be analyzed using the standard analysis weight, ANALWT, for all survey years 2008 through 2018.* With the revised 2012 model, the WHODAS and SDS 2008 half samples can be combined to form single estimates and use ANALWT.[17]

Standard Errors for Mental Illness Estimates

For the 2018 FFR and the 2018 detailed tables (CBHSQ, 2019b, 2019c), standard errors (SEs) for mental illness estimates (SMI, AMI, and AMI excluding SMI) were computed using the NSDUH dichotomous variable values without taking into account any variance introduced through using a model based on the clinical subsample data. *This ignores the added error resulting from fitting the 2012 SMI model, which can be very large.* See the *2012 MHSS: Design and Estimation Report* (CBHSQ, 2014a) for details. These *conditional* SEs (conditional on the model predictions being correct) are useful when making comparisons across years and across subpopulations within years because the errors due to model fitting are nearly the same across the estimates being compared.

Using Mental Illness Variables in Analysis

The mental illness measures (e.g., SMI, AMI, AMI excluding SMI) that were defined based on the 2012 model were examined to determine how they were associated with the mental health predictor variables in the 2012 model. It was found that the 2012 model significantly overestimated the proportion of adults aged 18 or older with SMI (and those with AMI) who had suicidal thoughts in the past year and the proportion of adults who had MDE in the past year (as compared with the clinical interview estimates of the same categories). *Therefore, it is recommended that the mental illness measures derived from the 2012 model should not be used when analyzing past year suicidal thoughts, past year MDE, or other associated variables (including past year suicide attempts, suicide plans, medical treatment for suicide attempts, or lifetime MDE). For example, mental illness estimates should not be generated by whether a respondent has serious thoughts of suicide; likewise, it is not recommended to generate serious thoughts of suicide estimates by levels of mental illness. Similarly, it is recommended that model-based mental illness measures should not be used in conjunction with the K6 variables (including SPD) or WHODAS variables in any analysis (CBHSQ, 2014a).* Note that age is a predictor in the mental illness models; however, it is not an issue to show estimates of mental illness by any age group.

3.4 Substance Use Disorders

Because of the changes to the sections for hallucinogens, inhalants, methamphetamine, and prescription psychotherapeutic drugs in the 2015 NSDUH questionnaire, the sets of

[16] The mental illness measure for AMI excluding SMI was added during the 2014 NSDUH and is based on the 2012 model. Because AMI excluding SMI is a composite of the LMI and MMI measures, the same analysis issues apply.

[17] This differs from the initial recommendation for analyzing measures of mental illness besides SMI based on the 2008 model. Because of the 2008 split sample, an adjusted mental health sample weight, MHSAMPWT, was created so that the WHODAS and SDS half samples were separately representative of the civilian, noninstitutionalized population aged 18 or older. However, this weight should not be used to analyze 2008 mental illness data based on the 2012 model.

respondents who were eligible to be asked the questions about substance use disorders (SUDs) also changed. *Therefore, new baselines started in 2015 for the SUD estimates for hallucinogens, inhalants, methamphetamine, and prescription psychotherapeutic drugs, as well as the aggregate categories of illicit drug use disorder and for overall SUD (i.e., having illicit drug use disorder or alcohol use disorder). Because the alcohol, marijuana, cocaine, and heroin sections did not change in the 2015 NSDUH questionnaire, a new baseline was not necessary for the alcohol use disorders and the SUDs for marijuana, cocaine, and heroin.* Starting with the 2018 NSDUH, estimates for the combination of tranquilizer or sedative use disorders were added to the detailed tables. For more detailed information on how the SUDs are defined and affected from the 2015 survey redesign, see Section 3.4.3 in Chapter 3 of the 2018 NSDUH methodological summary and definitions (CBHSQ, 2019a).

Before the 2015 NSDUH, all dependence or abuse recodes (except nicotine dependence) were created using edited data, and all values of unknown were treated as a "no" response. *Starting with the 2015 NSDUH, the SUD variables where a trend was not broken continued to treat all missing item responses as a "no" response; however, the SUD variables where a trend was broken because of the questionnaire changes (hallucinogens, inhalants, methamphetamine, prescription pain relievers, prescription tranquilizers, prescription stimulants, and prescription sedatives) had all missing item response data imputed.* Because trends were expected to be broken for these SUD measures due to the questionnaire redesign, a decision was made to implement the modified predictive mean neighborhood (modPMN) procedure for imputation. The imputation process for the SUD variables is described in detail in the *2018 Methodological Resource Book's* editing and imputation report (CBHSQ, 2020a), and the rationale for using modPMN imputation versus zero-fill imputation for SUD measures is provided in Chapter 10 of the evaluation of imputation methods report (CBHSQ, 2017b).

3.5 Substance Use Treatment

Changes to the questionnaire sections for hallucinogens, inhalants, methamphetamine, and prescription psychotherapeutic drugs in the 2015 NSDUH might have affected the sets of respondents who were eligible to be asked questions about treatment for substance use. The potentially affected treatment measures include

- receipt of treatment for illicit drug or alcohol use,
- substances for which respondents last received or were currently receiving treatment,
- perceived need for treatment for illicit drug or alcohol use in the past 12 months, and
- specific substances for which respondents perceived a need for treatment in the past 12 months.

Analysis conducted as part of the 2015 NSDUH redesign impact assessment report (RIAR) (CBHSQ, 2017c) indicated no evidence of a break in comparability between 2015 and earlier years for the three overall substance use treatment variables (alcohol use treatment, illicit drug use treatment, alcohol or illicit drug use treatment). However, given the major changes in who was asked the treatment questions based on who answered the substance use sections and the possible effects of the questionnaire redesign on different subgroups and other substance use treatment measures, the 2015 to 2017 detailed tables (CBHSQ, 2016c, 2017d, 2018b) did not

show any multiyear trend tables for the substance use treatment measures.[18] Multiyear trend tables are available for these measures starting with the 2018 NSDUH reports and tables, using the new baseline starting with the 2015 NSDUH.

Note that the presence of a SUD in the past year is an important component for defining individuals as needing treatment for their illicit drug or alcohol use. *Because new baselines began with the 2015 NSDUH for illicit drug use disorder and for overall SUD (i.e., having either illicit drug use disorder or alcohol use disorder), analysts should consider analytic goals carefully before trending the need for treatment measures with data prior to 2015.* For more information on various types of need for substance use treatment, see Section 3.4.4 in Chapter 3 of the 2018 NSDUH methodological summary and definitions (CBHSQ, 2019a).

3.6 Perceptions of Risk and Availability

A survey redesign carries the risk that preceding changes to the questionnaire will affect how respondents answer later questions (e.g., context effects; see Section C.6.2 of the 2015 NSDUH methodological summary and definitions [CBHSQ, 2016a]). Although the questions on perceptions of the risk of harm from using different substances and the perceived availability of specific illicit drugs did not change in the 2015 NSDUH, initial data quality checks on preliminary data showed deviations from the expected trends for these measures. These deviations from the expected trends continued to persist in all the 2015 data on perceived risk and availability measures from all four quarters. It was unclear whether the changes seen in the perceived risk and availability measures can be attributed to questionnaire or other survey changes with the 2015 NSDUH or if these changes reflect true changes in the population. The set of questions preceding the risk and availability section in the questionnaire had undergone several significant changes that could have affected the way in which respondents answered the perceived risk and availability questions. Further analysis of the 2015 data and the first two quarters of the 2016 data showed a continued deviation from the expected trend based on data before 2015.

As a result of this deviation, the affected risk and availability variables are considered not comparable with similar variables in years before 2015 and therefore were renamed starting with the 2015 NSDUH. Thus, the 2015 to 2017 detailed tables (CBHSQ, 2016c, 2017d, 2018b) did not include any multiyear trend tables for the risk and availability measures. Multiyear trend tables are available for these measures starting with the 2018 NSDUH reports and tables, using the new baseline starting with the 2015 NSDUH.

3.7 Prescription Drug Subtypes

Starting with the 2015 NSDUH, new tables showing any use and misuse of various types of prescription drug subtypes were added to the detailed tables. For the 2015 and 2016 detailed tables, a small number of respondents reported past year misuse of "any other" prescription pain reliever, tranquilizer, stimulant, or sedative, but they specified (a) only the misuse of prescription drugs that corresponded to existing prescription drug subtypes from the NSDUH questionnaire,

[18] The 2015 RIAR looked at general subpopulations only and did not complete analyses among more specific subpopulations or for other measures in the substance use treatment section. Analytic goals should be considered before pooling or comparing substance use treatment data from 2015 and later years with prior years. For more information on specific treatment measures, see Section 5.3 of the 2015 RIAR (CBHSQ, 2017c).

or (b) only the misuse of prescription drugs that corresponded to existing subtypes *and* the misuse of over-the-counter (OTC) drugs. In 2016, for example, this issue affected about 40 respondents for the past year misuse of any other prescription pain reliever, about 10 respondents for the misuse of any other prescription tranquilizer, about 10 respondents for the misuse of any other prescription stimulant, and about 5 respondents for the misuse of any other prescription sedative. In the 2015 and 2016 detailed tables, these respondents were included in estimates for "any other" prescription drug *and* for the relevant prescription drug subtype. For example, if a respondent in 2015 or 2016 specified Vicodin® as the only "other" prescription pain reliever that he or she had misused in the past year (or specified only Vicodin® and an OTC drug, such as Advil®), then the respondent was counted twice (i.e., counted in estimates for the past year misuse of hydrocodone and as a past year misuser of other pain relievers).

Beginning with the 2017 detailed tables, however, respondents were no longer counted as having misused "any other" prescription drug if the only drugs that they specified corresponded to prescription drug subtypes for that psychotherapeutic category (with or without other reports of OTC drugs). Using the previous example, a respondent in 2017 who specified Vicodin® as the only "other" prescription pain reliever that he or she had misused in the past year was counted in estimates for the past year misuse of hydrocodone products but not for the past year misuse of "any other" pain reliever.

The detailed tables that present these measures are 2-year tables showing the most recent year and the year prior. For the 2017 detailed tables, this change to the recoding procedures was applied to the 2016 and 2017 estimates for the past year misuse of any other prescription pain reliever, tranquilizer, stimulant, or sedative. Consequently, the 2016 estimates in the 2017 detailed tables for these measures may differ from previously published estimates in the 2016 detailed tables. For the 2018 detailed tables, this change continued to be applied to 2018 estimates for the past year misuse of any other prescription pain reliever, tranquilizer, stimulant, or sedative. For more information on this revision, see the 2018 public use data file codebook (CBHSQ/SAMHSA, 2019).

3.8 Adult Mental Health Outpatient Treatment

For adults aged 18 or older, mental health service utilization was defined as receiving treatment or counseling for any problem with emotions, nerves, or mental health in the 12 months prior to the interview in any inpatient or outpatient setting or the use of prescription medication for treatment of any mental or emotional condition that was not caused by the use of alcohol or drugs. During the 2017 NSDUH, outpatient mental health service measures from the 2010 to 2016 NSDUHs were recoded to be consistent with data prior to 2010 by excluding data on outpatient service locations that respondents wrote in for other alternative sources of mental health services. Because of this coding change, estimates in 2017 and 2018 NSDUH reports and tables for the receipt of outpatient mental health services among adults in 2010 to 2016 may differ slightly from previously published estimates for 2010 to 2016. This coding update was applied only to the outpatient mental health measures for the 2010 to 2016 NSDUHs; thus, measures derived from the outpatient mental health measure remain unchanged for 2010 to 2016. Starting with the 2017 NSDUH, the updated outpatient mental health measure is the standard for all derived measures. For more information on this revision, see the 2018 public use data file codebook (CBHSQ/SAMHSA, 2019).

3.9 Youth Reasons for Receiving Mental Health Services

The 2018 detailed tables present estimates for reasons for receiving mental health services in the past year among youths aged 12 to 17 who received specific mental health services. Youths aged 12 to 17 were asked about the reasons for receiving mental health services in two separate questions. These "reason" measures are not mutually exclusive, meaning that respondents could report multiple reasons for receiving the mental health services.

Starting with the 2017 detailed tables, a data quality improvement included a reclassification of three other, specify levels that are actually defined disorders and are now included as "self-reported mental disorder." Previously, these levels were included as "some other reason." Another improvement allowed for respondents who entered a valid reason for a service type other than "some other reason" in the first question to be assigned a "no" value for the unselected service types in the first question, regardless of how the respondent answered the second question asking about reasons for receiving treatment. These coding improvements had little impact on the estimates, and the measures are considered comparable with previous years. In the 2017 detailed tables, these coding changes were retroactively applied to the 2016 data; therefore, the 2016 estimates presented in the 2017 detailed tables may differ from previously published 2016 estimates. The 2017 and 2018 estimates presented in the 2018 detailed tables contain this coding improvement.

In the detailed tables, mental health services for youths are divided into specialty services (e.g., outpatient, inpatient/residential) or nonspecialty services (e.g., education, general medicine, child welfare). In addition to the coding improvements noted earlier, the code for the specialty mental health and education, general medicine, or child welfare measures was revised to assign some respondents who indicated receiving specialty mental health services and were known to have not received education, general medicine, or child welfare services for the specified reason to the "no" category. Previously, these respondents were assigned a system missing code. This issue occurred only when, in addition to the single nonspecialty mental health service they reported, respondents indicated receiving specialty mental health services and had either missing data for the specific reason or indicated receiving specialty mental health services for the specific reason. Because of the number of respondents recategorized by this recode, these measures in 2016 and onward are not comparable with those in 2015 and prior years. In the 2017 detailed tables, this coding revision was applied retroactively to the 2016 data; therefore, the 2016 estimates for the specialty mental health and education, general medicine, or child welfare measures presented in the 2017 detailed tables may differ from previously published 2016 estimates. The 2017 and 2018 estimates presented in the 2018 detailed tables contain this coding improvement. For more information about testing between years for these measures, see Section 7.2. For more detailed information on this revision, see the 2018 public use data file codebook (CBHSQ/SAMHSA, 2019).

3.10 Measures of Benzodiazepine Use and Misuse

Starting with the 2018 NSDUH, estimates of past year use and misuse of benzodiazepines were added, and trend data are shown for these measures back to 2015. Benzodiazepines can be prescribed as tranquilizers or sedatives. Respondents were classified as having used any benzodiazepine tranquilizer or sedative in the past 12 months if they reported any use of one or more of the benzodiazepines in that period in the tranquilizers screener section or the sedatives

screener section.[19] Respondents who did not report using the specific benzodiazepines asked about in the tranquilizers screener section or sedatives screener section in the past 12 months, but specified a benzodiazepine as one of the other tranquilizers or other sedatives they misused in later sections (i.e., tranquilizers main section, sedatives main section) were logically defined as having used benzodiazepines in the past 12 months because respondents who reported misusing benzodiazepines logically must have also used them for any reason. Similarly, respondents who reported they misused a benzodiazepine tranquilizer or sedative in the past 12 months—either from a response to a direct question (e.g., the direct question about misuse of Xanax® in the past 12 months) or as some other tranquilizer or sedative they misused in that period—were classified as having misused any benzodiazepine in the past 12 months.

Because of the potential for respondents to report the misuse of benzodiazepine tranquilizers as other sedatives or vice versa, measures for the past year misuse of any miscellaneous prescription benzodiazepine also were created for the 2018 NSDUH detailed tables. Respondents were classified into this miscellaneous prescription benzodiazepine category if they reported the misuse of benzodiazepine tranquilizers they specified as other sedatives or the misuse of benzodiazepine sedatives they specified as other tranquilizers. However, respondents in this miscellaneous prescription benzodiazepine category also could fall into other benzodiazepine tranquilizer or sedative categories in the detailed tables. For example, respondents who reported in the tranquilizers section that they misused Xanax® in the past year and specified the past year misuse of Xanax® as some other sedative were counted as having misused benzodiazepine tranquilizers and alprazolam products because of their report of Xanax® misuse in the tranquilizers section; these respondents also were included in the miscellaneous prescription benzodiazepine measure because of their report that they misused Xanax® as some other sedative. Nevertheless, the miscellaneous prescription benzodiazepine estimates give data users an indication of the extent of reporting of benzodiazepines across the respective categories for tranquilizers and sedatives. However, respondents who reported benzodiazepine use or misuse in the tranquilizers and sedatives sections of the interview were counted only once in aggregate estimates for the use or misuse of any prescription benzodiazepine in the past 12 months.

Respondents were asked about their use and misuse of benzodiazepines only for the past year reference period; therefore, there are no lifetime or past month measures for benzodiazepines. The following measures were also not created for benzodiazepines: initiation of benzodiazepine misuse in the past year and a SUD attributable to the misuse of benzodiazepines in the past year. These measures were not created because the interview sections for tranquilizers and sedatives also included drugs that are not benzodiazepines.

For example, suppose respondents misused a sedative in the past 12 months that was not a benzodiazepine (e.g., Ambien®) and misused a benzodiazepine sedative in the past 12 months (e.g., Halcion®). These respondents were asked about their misuse in the past 30 days and their SUD symptoms in the past 12 months for "prescription sedatives." Consequently, it could not be determined unambiguously for these respondents whether they misused a benzodiazepine sedative in the past 30 days or whether their SUD symptoms were attributable to their misuse of benzodiazepine sedatives or sedatives that were not benzodiazepines.

[19] The 2018 NSDUH questionnaire is available at https://www.samhsa.gov/data/.

NSDUH respondents beginning with the 2015 survey were asked about the initiation of misuse of prescription psychotherapeutic drugs for only the individual prescription drugs they had misused in the past 12 months. If respondents misused benzodiazepine sedatives and sedatives that were not benzodiazepines in the past 12 months and reported past year initiation of misuse for all the sedatives they misused in that period, then they were asked whether they ever misused any prescription sedative more than 12 months prior to the interview. Respondents who reported they misused "any prescription sedative" prior to the past 12 months would not be past year initiates for the misuse of any prescription sedative but could still have initiated the misuse of benzodiazepine sedatives (or any benzodiazepine) in the past year.

3.11 Measures of Tranquilizer or Sedative Use and Misuse

Starting in 2018, the following measures for the misuse of tranquilizers or sedatives were included in NSDUH reports or tables: misuse of tranquilizers or sedatives in the past 12 months (i.e., past year) and misuse of tranquilizers or sedatives in the past 30 days (i.e., past month). Respondents were classified as having misused prescription tranquilizers or sedatives in the past 12 months if they reported the misuse of prescription tranquilizers, prescription sedatives, or both in that period. A similar principle applied to the classification of respondents as having misused tranquilizers or sedatives in the past 30 days. Lifetime estimates of tranquilizer or sedative use or misuse are not reported due to a change in emphasis on past year rather than lifetime misuse of specific prescription drugs as part of the partial redesign of the 2015 NSDUH questionnaire. This change appears to have affected the validity of estimates for lifetime misuse of prescription psychotherapeutic drugs (see Section C in the methodological summary and definitions report for the 2015 NSDUH; CBHSQ, 2016a).

3.12 Decennial Census Effects on NSDUH Substance Use and Mental Health Estimates

As discussed in Chapter 2, the person-level weights in NSDUH were calibrated to population estimates (or control totals) obtained from the U.S. Census Bureau. For the weights in 2002 through 2010, annually updated control totals based on the 2000 census were used.[20] Beginning with the 2011 weights, however, the control totals from the U.S. Census Bureau are based on the 2010 census. Two investigations were implemented at the national level to assess the effects of using control totals based on the 2010 census instead of the 2000 census. One of these investigations focused specifically on measures of substance use that are used in the 2011 national findings report (CBHSQ, 2012e) and detailed tables (CBHSQ, 2012b), whereas a separate analysis was conducted to evaluate the impact of the weighting changes on mental health estimates in the 2011 mental health findings report (CBHSQ, 2012d) and associated mental health detailed tables (CBHSQ, 2012c). Because the 2017 and 2018 NSDUH estimates are based on weights that were poststratified to population control totals that were in turn based on projections from the 2010 census, 2-year trend comparisons between 2017 and 2018 are not subject to census effects. *However, trends between 2010 (or earlier years) and 2011 (or later years) may be influenced by census effects, especially for particular subgroups (e.g., people reporting two or more races for both investigations, people reporting American Indian or Alaska*

[20] In addition to the standard 2010 analysis weights poststratified to 2000 census control totals, special weights that were poststratified to 2010 census control totals are available on the 2010 NSDUH public use file (CBHSQ/SAMHSA, 2012).

Native or Native Hawaiian or Other Pacific Islander). An additional investigation was done at the state level to evaluate the impact of census effects on model-based small area estimation (SAE).

For more information on the impact of decennial census effects on NSDUH substance use direct estimates, see Section B.4.3 in Appendix B of the 2011 national findings report (CBHSQ, 2012e). For more information on the impact of the decennial census effects on NSDUH mental health direct estimates, see Appendix A of the 2011 mental health findings report (CBHSQ, 2012d). For more information on the impact of the decennial census effects on NSDUH model-based small area estimates, see the 2011-2012 NSDUH SAE guide (CBHSQ, 2013a) and, for greater detail, a currently unpublished NSDUH report (CBHSQ, 2014b). Additionally, for more information on the sampling weight calibration in the 2011 NSDUH, see the person-level sampling weight calibration report (Chen et al., 2013).

3.13 Using Revised Estimates for 2006 to 2010

During regular data collection and processing checks for the 2011 NSDUH, data errors were identified. *These errors affected the data for Pennsylvania (2006-2010) and Maryland (2008-2009). Cases with erroneous data were removed from the data files, and the remaining cases were reweighted to provide representative estimates.* The errors had minimal impact on the national estimates and no effect on direct estimates for the other 48 states and the District of Columbia. In reports where model-based SAE techniques were used, estimates for all states may have been affected, even though the errors were concentrated in only two states. However, in reports that did not use model-based estimates, the only estimates appreciably affected are estimates for Pennsylvania, Maryland, the mid-Atlantic division, and the Northeast region. The 2018 FFR and 2018 detailed tables (CBHSQ, 2019b, 2019c) did not include state-level, model-based, or division-level estimates. However, the detailed tables did include estimates for the Northeast region. *Estimates for the Northeast region based on 2006-2010 data may therefore differ from previously published estimates.* Tables and estimates based only on 2011 or later data are unaffected by these data errors. All affected tables (i.e., tables with estimates based on 2006-2010 data) contain a note to indicate this to the user.

Caution is advised when comparing estimates from older reports with data from more recent reports that are based on corrected data files. As discussed previously, comparisons of estimates for Pennsylvania, Maryland, the mid-Atlantic division, and the Northeast region are of most concern, whereas comparisons of national data or data for other states and regions are essentially still valid. A selected set of corrected versions of reports and tables has been produced. In particular, a set of modified detailed tables that include revised 2006-2010 estimates for the mid-Atlantic division and the Northeast region for certain key measures has been released. Given the change noted previously, comparisons between unrevised 2006-2010 estimates and estimates based on 2011-2018 data for the areas of most concern are not recommended.

This page intentionally left blank

4. Missingness

4.1 Potential Estimation Bias Due to Missingness

In the 2018 National Survey on Drug Use and Health (NSDUH), many variables, including the main drug and various demographic variables, had missing item response values imputed. The imputation process treats the imputed value as a true response and therefore may underestimate the variance, but the difference is small enough to be considered ignorable. See the 2018 NSDUH editing and imputation report (Center for Behavioral Health Statistics and Quality [CBHSQ], 2020a) and the predictive mean neighborhood evaluation report (CBHSQ, 2017b) for further details on the imputation process and the evaluation on the impact of imputation on the variance.

The missing item responses of many other variables were not imputed, and these missing responses may lead to biased estimates in the 2018 detailed tables (CBHSQ, 2019c). In addition, another source of potential uncertainty about some estimates may occur because of the way unknown item responses (e.g., blank, "don't know," "refused") were coded for different variables. *For example, some recoded variables (i.e., variables created from one or more source variables) classified unknown item responses in the source variable(s) as missing values, whereas others did not.* See Ruppenkamp, Emrich, Aldworth, Hirsch, and Foster (2006) for further details. Exhibits A.31 through A.34 demonstrate how to compute prevalence estimates for variables with missing data using SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2013), Stata® (StataCorp LP, 2017), SAS® (SAS Institute Inc., 2017), and R (R Core Team, 2018).

Recall from Chapter 3 that prevalence estimates are defined as the proportions of the population who exhibit characteristics of interest. Let $\hat{p}_d$ represent the estimated prevalence estimate of interest for domain *d*, with $\hat{p}_d$ defined as

$$\hat{p}_d = \frac{\hat{Y}_d}{\hat{N}_d},$$

where $\hat{Y}_d$ = estimated number of people exhibiting the characteristic of interest in domain *d*, and $\hat{N}_d$ = estimated population total for domain *d*.

The variable defining the characteristic of interest (e.g., illicit drug use) is referred to as the *analysis* variable, and the variable defining the domain of interest (e.g., receipt of past year mental health treatment/counseling) is referred to as the *domain* variable. Suppose that the analysis variable has all its missing values imputed, but the domain variable does not employ the imputation of missing values. In such cases, the estimates $\hat{N}_d$ and $\hat{Y}_d$ may be negatively biased, and the $\hat{p}_d$ estimates also may be biased. To see this, suppose that the domain variable has *D* levels, and define

$$\hat{N} = \sum_{d=1}^{D} \hat{N}_d + \hat{N}_m,$$

where $\hat{N}$ = estimated population total, $\hat{N}_d$ = estimated population total for domain *d*, $d = 1, 2, ..., D$, and $\hat{N}_m$ = estimated population total corresponding to the missing values of the domain variable. Thus, if $\hat{N}_m$ is positive (i.e., there are missing domain-variable responses), then at least one of the $\hat{N}_d$ estimates will be negatively biased. The presence of negative bias in at least one of the $\hat{Y}_d$ estimates can be similarly demonstrated if $\hat{Y}_m$ is positive, where $\hat{Y}_m$ = the estimated number of people exhibiting the characteristic of interest and corresponding to the missing values of the domain variable. *If either of $\hat{N}_m$ and $\hat{Y}_m$ is positive, then $\hat{p}_d$ may be biased by some unknown amount.*

Suppose instead that the domain variable has all its missing values imputed, but the analysis variable does not employ the imputation of missing values. In such cases, at least one of the $\hat{N}_d$ estimates will be negatively biased. If all missing values for the analysis variable in the domain do not have the condition of interest, $\hat{Y}_d$ would have no bias. Otherwise, $\hat{Y}_d$ will be negatively biased. Thus, $\hat{p}_d$ may be biased by some unknown amount. Likewise, $\hat{p}_d$ may be biased in the case that the domain and analysis variables do not employ the imputation of missing values.

In the 2018 detailed tables (CBHSQ, 2019c), *potential bias in the $\hat{N}_d$, $\hat{Y}_d$, or $\hat{p}_d$ estimates were not treated, although footnotes included on the tables provide detailed information about which estimates included or excluded missing values.* This problem may be illustrated by the following example, which corresponds to information presented in Tables 9.9A and 9.9B of the 2015 detailed tables.[21] Table 9.9A presents estimates of the past year use of several types of illicit drugs among youths aged 12 to 17 for 2014 and 2015. These analysis variables are grouped into a two-level domain variable that is categorized according to whether a respondent had a past year major depressive episode (MDE). Table 12.1A of the 2015 detailed tables (CBHSQ, 2016c) shows the population estimate of youths aged 12 to 17 as approximately 24,893,000. However, the subdomain population estimates summed to approximately 24,259,000, resulting in an estimate of $\hat{N}_m$ = 634,000 (approximately 2.5 percent of the total population). This number represents the estimated population not assigned to either domain. This negative bias can extend to various analysis variables, such as "Illicit Drugs." In 2015, the total estimate of youths aged 12 to 17 who used illicit drugs in the past year was approximately 4,346,000. However, the 2015 estimates of youths aged 12 to 17 who used illicit drugs in the past year among the valid subdomains (where past year MDE status was not missing) summed to 4,206,000, resulting in an estimate of $\hat{Y}_m$ = 140,000 (approximately 3.2 percent of the total population aged 12 to 17 who used illicit drugs in the past year).

Table 9.9B in the 2015 detailed tables (CBHSQ, 2016c) presents prevalence estimates of the past year use of several types of illicit drugs among youths aged 12 to 17 for 2014 and 2015. Because $\hat{N}_m$ is positive and $\hat{Y}_m$ is positive for the "Illicit Drugs" analysis variable, the prevalence estimates for this variable may be biased by some unknown amount across the two

[21] Although this example references estimates from the 2015 detailed tables, similar examples can be found in the 2016 detailed tables onward.

domains. The 2015 prevalence estimates of illicit drug use reported in Table 9.9B for youths who had or did not have past year MDE are 31.5 and 15.3 percent, respectively. By recoding the item missingness of the domain variable MDE as having or not having MDE, the approximate range of possible bias values for each of these estimates is as follows: between -4.41 and 3.03 percent and between -0.35 and 0.55 percent, respectively.

As mentioned previously, some recoded variables classify unknown item responses in source variables as missing values, whereas others do not. That is, for some variables, item missingness is zero imputed (i.e., missing items are imputed as not having the condition or event of interest). Some examples of zero-imputed variables include various substance use treatment variables, select dependence and abuse variables (dependence and abuse variables for prescription drugs, inhalants, methamphetamine, and hallucinogens are imputed using the modified predictive mean neighborhood procedure), and serious psychological distress variables.[22] Respondents with missing data that are not imputed are generally excluded from the relevant analyses. For the detailed tables, investigations are performed to look at these rates of missingness. Rates of missingness are evaluated separately for each subpopulation within a table to allow for detection of variations in missingness rates among different subpopulations.

For the 2018 detailed tables (CBHSQ, 2019c), there were no new analysis variables with missing data to investigate. There were, however, new subpopulations to evaluate for existing analysis variables, including new age groups, substance use measures, and substance use disorders. These evaluations found that the rates of missingness for these new subpopulations were similar to the rates for existing subpopulations. The results from the 2014 through 2017 investigations for all existing analysis variables and subpopulations were assumed to still hold (CBHSQ, 2016d, 2017e, 2018c, 2019e), which concluded overall that missing data were not a concern for most topics presented in these tables. *However, items on perceived availability of various illicit drugs and source of prescription drugs obtained for most recent use in the detailed tables generally have larger rates of missing data.* For example, the maximum weighted rate of missing data for the source of prescription drugs obtained for most recent use was 13.8 percent with about half of the subpopulations considered for these measures having a weighted missingness rate of greater than 5.0 percent in 2015.

For nonimputed recoded variables where unknown item responses were treated as negative responses and not as missing values, there is also potential bias. Assuming that unknown item responses are negative responses, a negative bias is created with magnitude dependent on the percentage of respondents with missing data and on the magnitude of the estimate. Specifically, higher levels of nonresponse paired with high estimates induce a larger negative bias. A lower level of nonresponse paired with lower prevalence estimates induces a smaller negative bias. Intermediate combinations induce a moderate negative bias.

The approximate range of bias can be illustrated with Table 5.16 of the 2015 detailed tables (2016c), which presents prevalence estimates of the past year receipt of substance use treatment among people aged 12 or older by various demographic characteristics for 2015.[23] Because the unknown responses for the analysis variable are treated as negative responses, the

[22] This is not an exhaustive list of zero-imputed variables. For more information on specific variables, see the 2018 public use data file codebook (CBHSQ/Substance Abuse and Mental Health Services Administration, 2019).

[23] Although this example references estimates from the 2015 detailed tables, similar examples can be found in the 2016 detailed tables onward.

full population is used in the table (267,694,000, from Table 12.1A of the 2015 detailed tables) (CBHSQ, 2016c). Table 5.16A shows that 3,679,000 people aged 12 or older received substance use treatment in the past year for illicit drug or alcohol use (1.4 percent of the total population; Table 5.16B) in 2015. If unknown responses are excluded from the analysis, the estimated total population would be 264,778,000, resulting in a prevalence estimate of 1.4. (Note that there is a slight difference between the two prevalence estimates not seen because of rounding.) However, if the unknown responses are treated as positive responses, then the estimated number of people aged 12 or older who received illicit drug treatment in the past year would be 6,595,000 (2.5 percent of the total population). Thus, there is an approximate range of bias based on the 2015 data between 0 and -1.1 percent.

4.2 Variance Estimation in the Presence of Missingness

SUDAAN (RTI International, 2013) uses the number of strata (see Chapter 6 for more information) and number of primary sampling units (PSUs) in its variance calculations, even if there are some PSUs in which a variable is entirely missing for all sample members associated with that PSU. The rationale behind this approach is that there may be individuals in the target population who have nonmissing values in PSUs where no sample members have nonmissing values.

To illustrate how this is operationalized in SUDAAN, consider the following example. Suppose there is interest in calculating the mean of some variable (say, X), but there are missing values associated with variable X. SUDAAN then creates an internal subpopulation indicator variable (say, δ), where $\delta = 1$ if variable X is not missing, and $\delta = 0$ if variable X is missing. SUDAAN then internally calculates the mean and variance of variable X by using δX, assuming that the full sample mean is the same as the nonmissing sample mean.

For the variance estimator based on the Taylor series linearization approach, one of the terms in the variance estimator consists of the sum of squared deviations of PSU-level totals about their stratum-level means, divided by the number of PSUs in the stratum minus 1. Therefore, if SUDAAN encounters an incorrect number of PSUs within a stratum, then this term is incorrectly calculated. In addition, if there is only one PSU in a stratum, then the denominator for the variance term associated with that stratum becomes 0, which causes the overall variance estimate to return an error message in SUDAAN. *By including all PSUs in a stratum, whether the PSU has reported values, SUDAAN computes the variances appropriately; that is, PSUs with nothing but missing values for a variable should never be excluded from an input file. Thus, users are encouraged to use the full NSDUH dataset when running analyses in order to keep the complete data structure for variance estimation. Subsetting of the data to populations of interest should be done within SUDAAN (e.g., using SUDAAN's SUBPOPN statement).*

5. Sampling Error

In sampling, statistics from different samples will vary and can differ from the true population parameter. Sampling error is the error caused by using statistics based on a sample instead of a complete census. Standard errors (SEs) are commonly used to measure how much these statistics differ from the true parameter. This measure is incorporated in common statistical methods such as significance testing (see Chapter 7) and confidence intervals (see Chapter 8). *As were the prevalence estimates, all the variance estimates for prevalence (including those for prevalence based on annual averages from combined data) were calculated using a method in SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2013) that is unbiased for linear statistics.* This method is based on multistage clustered sample designs where the first-stage (primary) sampling units are drawn with replacement.

Because of the complex nature of the sampling design for the National Survey on Drug Use and Health (NSDUH) (specifically, the use of stratified cluster sampling), key nesting variables were created for use in SUDAAN to capture explicit stratification and to identify clustering. Starting with the 2005 NSDUH,[24] a change was made in the way the key nesting variables were defined. Each state sampling region (SSR) appears in a different variance estimation stratum every quarter. This method has the effect of assigning the regions to strata in a pseudo-random fashion while ensuring that each stratum consists of four SSRs from four different states.

Two replicates per year are defined within each variance stratum (VEREP). Each variance replicate consists of four segments, one for each quarter of data collection. One replicate consists of those segments that are "phasing out" or will not be used in the next survey year. The other replicate consists of those segments that are "phasing in" or will be fielded again the following year, thus constituting the 50 percent overlap between survey years. A segment stays in the same VEREP for the 2 years it is in the sample. This simplifies computing SEs for estimates based on combined data from adjacent survey years.

Although the SEs of estimates of means and proportions can be calculated appropriately in SUDAAN using a Taylor series linearization approach, the actual SEs of estimates of totals may be smaller in situations where the domain size is poststratified to data from the U.S. Census Bureau. Because of the potential for gains in precision, alternatives for estimating SEs of totals were implemented in all of the 2018 detailed tables (Center for Behavioral Health Statistics and Quality [CBHSQ], 2019c), where appropriate.

Estimates of means or proportions, $\hat{p}_d$, such as drug use prevalence estimates for a domain *d*, can be expressed as a ratio estimate:

$$\hat{p}_d = \frac{\hat{Y}_d}{\hat{N}_d},$$

[24] The new design variables were created retroactively for 1999 through 2004; however, the old design variables continue to be used to generate 2002-2004 estimates in multiyear trend detailed tables and first findings reports (FFRs) for consistency with previously published estimates. Analyses beyond the detailed tables and FFRs typically use the new design variables for all available years.

where $\hat{Y}_d$ is a linear statistic estimating the number of substance users in the domain *d*, and $\hat{N}_d$ is a linear statistic estimating the total number of people in domain *d* (users and nonusers). The SUDAAN software package is used to calculate direct estimates of $\hat{Y}_d$ and $\hat{N}_d$ and can be used to estimate their respective SEs. A Taylor series approximation method implemented in SUDAAN provides estimates for $\hat{p}_d$ and its SE.

When the domain size, $\hat{N}_d$, is free of sampling error, the following formula is an alternative to using SUDAAN to estimate the SE for the total number of persons with a characteristic of interest (e.g., substance users):

$$\mathrm{SE}(\hat{Y}_d) = \hat{N}_d \mathrm{SE}(\hat{p}_d).$$

This alternative SE estimation method is theoretically correct when the domain size estimates, $\hat{N}_d$, are fixed (i.e., among those domains forced to match their respective U.S. Census Bureau population estimates through the weight calibration process). In these situations, $\hat{N}_d$ is not subject to a sampling error induced by the NSDUH design. For more information on the person-level sampling weight calibration in the 2018 NSDUH, see CBHSQ (2020b).

For an estimated number $\hat{Y}_d$, where the domain $\hat{N}_d$ is nonfixed (i.e., where domain size estimates are not forced to match the U.S. Census Bureau population estimates), this alternative SE estimation method still may provide a good approximation if it can be assumed that the sampling variation in $\hat{N}_d$ is negligible relative to the sampling variation in $\hat{p}_d$. This is a reasonable assumption for most estimates in NSDUH.

For various subsets of estimates, using this alternative SE estimation method where domain sizes are nonfixed yielded an underestimate of the variance of a total because $\hat{N}_d$ was subject to considerable variation. Because of this underestimation, the alternative SE estimation method was not implemented when $\hat{N}_d$ was nonfixed. In 2000, an approach was implemented to reflect more accurately the effects of the weighting process on the variance of total estimates. This approach consisted of calculating SEs of totals for all estimates in a particular detailed table using the alternative SE estimation method when a majority of estimates in a table were among domains in which $\hat{N}_d$ was fixed during weighting or if it could be assumed that the sampling variation in $\hat{N}_d$ was negligible. Detailed tables in which the majority of estimates were among domains where $\hat{N}_d$ was subject to considerable variability were calculated directly in SUDAAN.

To improve on the accuracy of the SEs, a "mixed" method approach was implemented in which tables might include more than one method of SE estimation. This mixed approach was applied to selected tables in the 2004 NSDUH, and it was implemented across all tables starting with the 2005 NSDUH and continuing in subsequent years. *This approach assigns the method of SE calculation to domains within tables so that all estimates among a select set of domains with fixed* $\hat{N}_d$ *were calculated using the alternative SE estimation method, and all other estimates*

were calculated directly in SUDAAN, regardless of other estimates within the same table. The set of domains with a fixed $\hat{N}_d$ was restricted to main effects and two-way interactions to maintain continuity between years.[25] Domains consisting of three-way interactions may be fixed in one year but not necessarily in preceding or subsequent years. Using such SEs did not affect the SE estimates for the corresponding proportions presented in the same sets of tables because all SEs for means and proportions are calculated directly in SUDAAN. Appendix A contains SUDAAN, Stata® (StataCorp LP, 2017), SAS® (SAS Institute Inc., 2017), and R (R Core Team, 2018) examples that demonstrate how to compute SEs of proportions and both types of SEs of totals (see Exhibits A.1 through A.8).

Table 5.1 contains a list of domains used in the 2018 detailed tables that employ the alternative SE estimation method for the restricted-use data file.[26] This table includes the main effects and two-way interactions and can be used to identify the method of SE calculation employed for estimates of totals in the 2018 detailed tables (CBHSQ, 2019c). An example from the 2015 detailed tables would be Table 1.30, which presents estimates of illicit drug use among adults aged 18 or older within the domains of gender, Hispanic or Latino (referred to as "Hispanic" hereafter) origin and race, education, and current employment. Estimates among the total population (age main effect), males and females (age by gender interaction), and Hispanics and non-Hispanics (age by Hispanic origin interaction) used the alternative SE estimation method to calculate the SEs. The SEs for all other estimates, including white and black or African American (age by Hispanic origin by race interaction), were calculated directly from SUDAAN.[27] It is important to note that estimates presented in the detailed tables for racial groups are among non-Hispanics, unless noted otherwise. For instance, the domain for whites is actually non-Hispanic whites and is therefore a two-way interaction.

Table 5.1 Demographic and Geographic Domains Shown in the First Findings Reports and Detailed Tables That Use the Alternative Standard Error Estimation Method for Calculating Standard Errors of the Estimated Number of Persons (Totals), 2018

Main Effects	Two-Way Interactions
Age Group 12-17 18-25 26-34 35-49 50-64 65 or Older Collapsed Age Group Categories from Above[1] **Gender** Male Female	**Age Group × Gender** (e.g., Males Aged 12 to 17) **Age Group × Hispanic Origin[1]** (e.g., Hispanics or Latinos Aged 18 to 25)

(continued)

[25] In some years, not all the race domains in Table 5.1 are forced to fully match the U.S. Census Bureau population estimates due to models not converging. When this occurs, the sampling variation in $\hat{N}_d$ for these domains is considered negligible. Therefore, the race domains are considered fixed for every year.

[26] See the variance estimation of totals section in the 2018 public use data file introduction for a list of domains that employ the alternative SE estimation method for the 2018 public use data file (CBHSQ/Substance Abuse and Mental Health Services Administration, 2019).

[27] Although this example references estimates from the 2015 detailed tables, similar examples can be found in the 2016 detailed tables onward.

Table 5.1 Demographic and Geographic Domains Shown in the First Findings Reports and Detailed Tables That Use the Alternative Standard Error Estimation Method for Calculating Standard Errors of the Estimated Number of Persons (Totals), 2018 (continued)

Main Effects	Two-Way Interactions
Hispanic Origin	
Hispanic or Latino	**Age Group** × **Geographic Region**[1]
Not Hispanic or Latino	(e.g., People Aged 12 to 25 in the Northeast)
Race[2]	
White	
Black or African American	**Gender** × **Hispanic Origin**
Others	(e.g., Not Hispanic or Latino Males)
Geographic Region	
Northeast	
Midwest	**Hispanic Origin** × **Race**
South	(e.g., Not Hispanic or Latino Whites)
West	

NOTE: The alternative standard error (SE) estimation method for estimated number of persons (totals), $SE\left(\hat{Y}_d\right) = \hat{N}_d \mathrm{SE}\left(\hat{p}_d\right)$, is applied when the domain size estimates, $\hat{N}_d$, are among those forced to match their respective U.S. Census Bureau population estimates through the weight calibration process.

NOTE: The alternative SE estimation method does not affect the SE for the corresponding means and proportions. SEs for means and proportions are calculated directly in SUDAAN, whereas the alternative SE estimation method is computed outside of SUDAAN using the formula above.

NOTE: This table shows only the domains and domain combinations used in the first findings reports and detailed tables. Other domains and domain combinations are omitted that also use this alternative SE estimation method but are not included in these specific reports or tables. For example, methodological studies or special requests often include a wider variety of domains and survey years. This variation requires the SE method to be assessed for each individual analysis. For a detailed list of domains for NSDUH that were forced to match their respective U.S. Census Bureau population estimates through the weight calibration process, see Center for Behavioral Health Statistics and Quality (2020b).

[1] Main effect age group categories shown in the table can be collapsed to form broader age group categories (e.g., 12 or older, 50 or older, 18 to 49, 26 to 49). Collapsed main effect age group categories and two-way interactions with other main effect demographic or geographic domains shown (e.g., males aged 50 or older) also use the alternative SE estimation method because the collapsed main effects will sum to the census totals for the category being defined. However, broader age groups that include only a subset of the main effect age groups (e.g., 12 to 20, 21 or older, 15 to 44), age groups that are finer than the main effect age groups (e.g., 12 to 13, 18 to 20), or two-way interactions of these types of collapsed age categories with other main effect domains (e.g., females aged 15 to 44) should not use the alternative SE estimation method.

[2] Race is included as a main effect in this table for completeness; however, race groups presented here include all persons within a given race category, regardless of whether they are Hispanic or not Hispanic. In contrast, all other groups presented in the detailed tables are indented underneath the non-Hispanic ethnicity heading. For example, the domain for whites in the detailed tables is actually non-Hispanic whites and is therefore a two-way interaction. Thus, any additional domains crossed with non-Hispanic whites (e.g., whites aged 18 to 25) represent three-way interactions that do not use the alternative SE estimation method.

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

Some of the three-way interactions in Tables 5.22 to 5.25 of the 2016 detailed tables (CBHSQ, 2017d) had the alternative SE method inadvertently applied. Specifically, the overall race categories along with the three-way interaction of age by Hispanic origin by race were treated as fixed domains. As noted previously, the standard practice for the detailed tables is to calculate the SE of three-way interactions directly in SUDAAN. Starting with the 2017 detailed tables, SEs were calculated accordingly; therefore, SEs of the 2016 total estimates in the 2017 detailed tables for the non-Hispanic white and non-Hispanic black or African American race

categories may differ from those presented in the 2016 detailed tables. The magnitude of change for SEs of the total estimates for other impacted domains was not large enough to be seen in the detailed table estimates. The standard practice for computing SEs was also applied to corresponding Tables 5.20 to 5.22 in the 2018 detailed tables.

Although not reported in the 2018 detailed tables, additional geographic interactions are also treated as domains with fixed $\hat{N}_d$ for other NSDUH analyses. Similar to geographic region, geographic division and individual states are treated as domains with fixed $\hat{N}_d$, and two-way interactions with state and gender, Hispanic origin, quarter, and age group (12 to 17, 18 to 25, and 26 or older), and the two-way interaction between geographic region and age group, would all employ the alternative SE estimation method. Additionally, quarter, although not used in the 2018 detailed tables, is treated as a domain with fixed $\hat{N}_d$ as well as the two-way interaction with state. Other two-way interactions with quarter such as gender or age group are not treated as a domain with fixed $\hat{N}_d$.

This page intentionally left blank

6. Degrees of Freedom

6.1 Background

To determine whether the observed difference between estimates is statistically significant, the degrees of freedom (*df*) are needed to locate the corresponding probability level (*p* value) of the test statistic. The test statistic is computed from the sample data and represents a numerical summary of the difference between the estimates under consideration; it is a random variable that has a predetermined distribution (such as Student's *t*, chi-square, or *F*). *The df characterize the amount of variation expected in the estimation of sampling error and are used in conjunction with the test statistic to determine probabilities and evaluate statistical significance.* In statistics, the number of *df* refers to the number of independent units of information in a sample relevant to the estimation of a parameter or calculation of a statistic. In general, the *df* of a parameter estimate are equal to the number of independent observations that go into the estimate minus the number of other parameters that need to be estimated as an intermediate step. The *df* are also used to compute the confidence intervals (CIs) discussed in Chapter 8. The upper and lower limits of the CIs are defined by a constant value that is chosen to yield a level of confidence based on the *df*.

In practice, beyond a certain value, which *df* value is used has little impact. For example, the 97.5th percentile of the *t*-distribution is used in the National Survey on Drug Use and Health (NSDUH) to create 95 percent CIs and for two-sided hypothesis tests, and this does not change much once there are about 50 *df*. Thus, results with 50 *df* are similar to results with the 900 *df* used for the 2002 to 2013 NSDUHs and the 750 *df* used for the 2014 to 2022 NSDUHs (Exhibit 6.1). In addition, Table 6.1 shows the large sample 95 percent CI for a "typical" estimate (e.g., the percentage of past month users of alcohol in 2015) for different *df*. The CIs are similar.[28]

The *df* for NSDUH vary based on the sample design. Table 6.2 shows the *df* for specific states per the NSDUH sample designs.[29] Starting with the 2005 NSDUH, a change in the definition of the variance estimation strata had the effect of increasing the number of *df* for the state-level estimates fourfold while preserving the number of *df* for the national estimates. Revised design variables were created retroactively for years before 2005 (see footnote 24). *When producing 2002 to 2013 NSDUH estimates at the national level, there are 900 df.* If an analysis involves individual states, the *df* are determined by the number of strata in which the state is included. In the 2002 to 2013 surveys, there were two sample size groups. Large sample states (i.e., California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas) have 192 *df* because each large state is in 192 strata. Small sample states (i.e., all other states including the District of Columbia) have 48 *df* because each small state is in 48 different strata.

[28] Although this example references estimates from the 2015 detailed tables, similar examples can be found in the 2016 detailed tables onward.

[29] Users of the 2018 public use file (Center for Behavioral Health Statistics and Quality [CBHSQ]/ Substance Abuse and Mental Health Services Administration [SAMHSA], 2019) may find inconsistencies with the specific *df* presented in this report because the specific information referenced is based on the restricted-use dataset that was used to create the 2018 detailed tables (CBHSQ, 2019c) and the 2018 first findings report (FFR) (CBHSQ, 2019b).

Exhibit 6.1 97.5th Percentiles of *t*-Distributions for Varying Degrees of Freedom

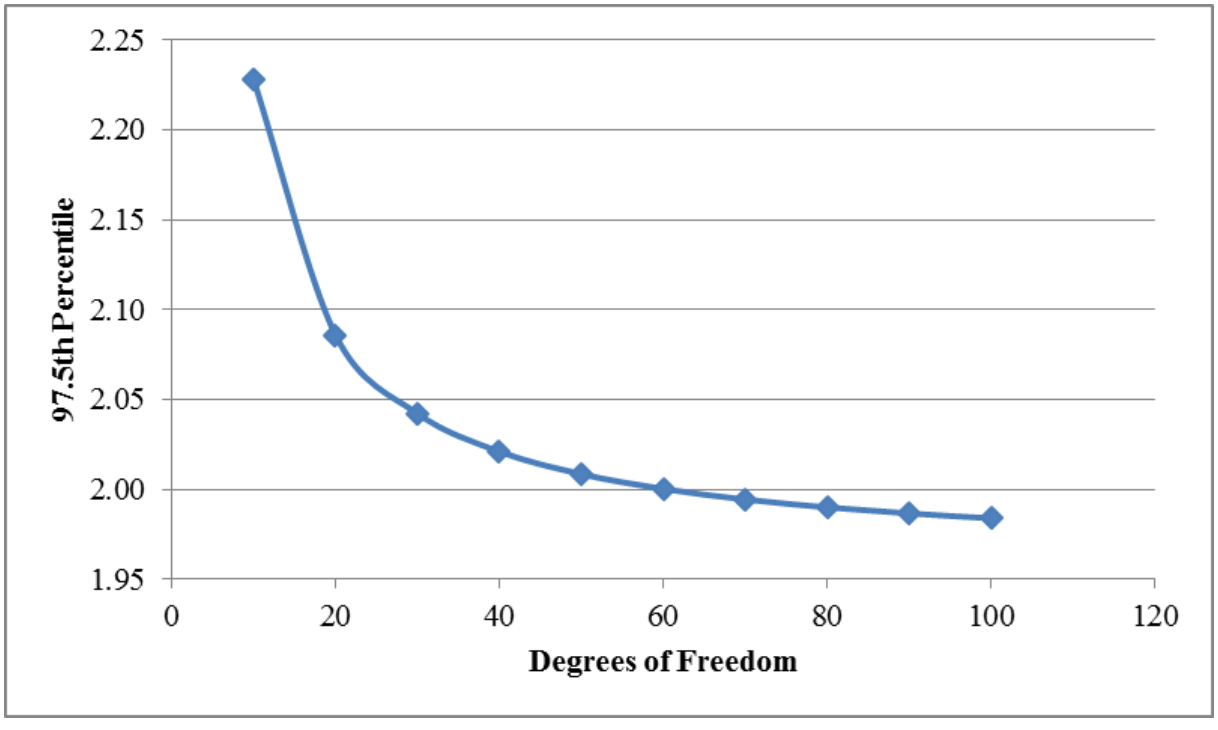


Table 6.1 Ninety-Five Percent Confidence Intervals for the Percentage of Past Month Users of Alcohol, Using Different Degrees of Freedom, 2015

Degrees of Freedom	Critical Value of the *t*-Distribution	95% Confidence Interval	
		Lower Limit	Upper Limit
10	2.2281	50.96	52.38
20	2.0860	51.01	52.34
30	2.0423	51.02	52.32
40	2.0211	51.03	52.31
50	2.0086	51.03	52.31
60	2.0003	51.03	52.31
70	1.9944	51.04	52.31
80	1.9901	51.04	52.30
90	1.9867	51.04	52.30
100	1.9840	51.04	52.30
500	1.9647	51.05	52.30
750	1.9631	51.05	52.30
900	1.9626	51.05	52.30
1,800	1.9613	51.05	52.30

NOTE: The percentage of past month users of alcohol used to produce the data in this table is 51.67 percent, with a corresponding standard error of 0.32, both rounded to 2 decimal places.

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

Table 6.2 Degrees of Freedom for Specific States per the NSDUH Sample Design Based on the Restricted-Use Dataset

States	Sample Design Years[1]	Degrees of Freedom[2]
California	2014-2022	144
	2005-2013	192
	2002-2004	192
Florida, New York, and Texas	2014-2022	120
	2005-2013	192
	2002-2004	192
Illinois, Michigan, Ohio, and Pennsylvania	2014-2022	96
	2005-2013	192
	2002-2004	192
Georgia, New Jersey, North Carolina, and Virginia	2014-2022	60
	2005-2013	48
	2002-2004	48
Remaining 38 states and the District of Columbia	2014-2022	48
	2005-2013	48
	2002-2004	48

[1] The NSDUH sample design variables were revised in 2005 and 2014. The 2005 revisions were applied retroactively to the 1999 through 2004 NSDUHs. Because of survey improvements in the 2002 NSDUH, the 2002 data constitute a new baseline, so this table does not include information before 2002.

[2] The degrees of freedom in this table are based on the new sample design variables. If using the old sample design variables for NSDUH years 2002 to 2004, the state degrees of freedom listed in this table would be divided by 4.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2002-2018.

Changes were made to the 2014 through 2017 NSDUHs and continued with the 2018 through 2022 sample allocation in order to increase the sample in the original 43 small states to improve precision of the state and substate estimates while moving closer to a proportional allocation in the larger states. This design moved the sample from two state sample size groups (large and small) to five state sample size groups. In this design, sampling strata called state sampling regions (SSRs) were formed within each state. *The partitioning divided the United States into a total of 750 SSRs, which result in 750 df for national estimates.* States in sample size group 1 (i.e., California) have 144 *df*, states in sample size group 2 (i.e., Florida, New York, and Texas) have 120 *df*, states in sample size group 3 (i.e., Illinois, Michigan, Ohio, and Pennsylvania) have 96 *df*, states in sample size group 4 (i.e., Georgia, New Jersey, North Carolina, and Virginia) have 60 *df*, and states in sample size group 5 (i.e., the remaining 38 states and the District of Columbia) have 48 *df*.

Appendix A contains examples that demonstrate how to define the *df* in SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2013), Stata® (StataCorp LP, 2017), SAS® (SAS Institute Inc., 2017), and R (R Core Team, 2018) to compute design-based estimates.

Under the NSDUH sample designs, for an analysis of a group of states, the *df* would be less than or equal to the sum of the *df* for each individual state due to overlap of strata. *Therefore, the specific number of df should be computed by counting the unique values of VESTR (variance estimation [pseudo] stratum) for the particular geographic area of interest.* For these

types of specific state analyses (or other subpopulations of interest), the *df* can be calculated outside SUDAAN, and this value entered manually into SUDAAN for use in testing; otherwise, the *df* are computed using the entire dataset. Similar methods can be used to compute appropriate *df* for any geographic region comprising counties. Using this technique with the public use file will give similar, but not always exact, results.

The technique of counting the number of unique values of VESTR (see above) can also be used to compute the number of *df* for analyses based on combining survey data across years. *An alternative technique for computing the df for analyses that use data combined (or pooled) across NSDUH sample design years involves summing the df from each sample design year (see Table 6.2) to determine the df for the NSDUH years and states of interest) because each sample design (i.e., 2002-2004, 2005-2013, 2014-2022) contains unique variance strata.* For example, when pooling 2013 and 2014 NSDUH data, the *df* for California would be 192 (2013) + 144 (2014) = 336 because the years being pooled come from two different sample designs. However, if pooling 2012 and 2013 NSDUH data, which both come from the same sample design, the *df* would simply be 192. Pooled data examples can be found in Exhibits A.1 through A.4 showing how the *df* can be designated.

6.2 Degrees of Freedom Used in Key NSDUH Analyses

The current practices for applying *df* to NSDUH data depend on the type of analyses. Table 6.3 summarizes key types of NSDUH analyses and the *df* used for these analyses for the various survey design years. *The detailed tables and FFRs use the national df for the most current survey year (including census region and division and estimates for all years including pooled years), with the exception of estimates for the mean age of first use (AFU) and the average number of days used.*[30] The current year *df* are used because when conducting significance testing between estimates with different *df* (e.g., 2014 vs. 2013), the lower *df* provide a more conservative test and are used. For all the currently analyzed years of NSDUH data, the current year's *df* have always been less than or equal to the previous years' *df*.

AFU and average number of days used estimates are treated differently because of the possibility of smaller sample sizes (i.e., the sample sizes for AFU estimates are typically the number of past year initiates); therefore, they belong to fewer variance estimation strata. Based on the NSDUH suppression rules, the sample size threshold for suppression of an average estimate is 10, whereas for prevalence estimates, it is 100. Thus, it is possible for nonsuppressed average estimates to have smaller sample sizes than prevalence estimates. For example, the subpopulation for estimates of mean AFU includes only past year initiates of prescription drugs and lifetime users of other drugs, which could be small for drugs with low prevalence estimates of use. An impact assessment was done using 2012-2013 data to determine whether the results of statistical comparisons between the means for the 2 years would be affected if the *df* were changed from the national *df* (900 in 2013) to the number of nonempty strata (the number of strata containing respondents with valid data to each specific question within the subpopulation). This latter value would produce more conservative tests. *After the impact assessment, a decision was made to use the number of nonempty strata as the df for the detailed tables that include*

[30] Beginning with the 2018 NSDUH, estimates for mean ages at first use are no longer presented in the detailed tables or the FFRs.

estimates of mean AFU. This decision was expanded to include estimates for all average estimates starting with the 2016 NSDUH.

Unlike the detailed tables and the FFRs, which use the national *df* for estimates by geographic subgroups (census region and division), special analyses and methodological reports follow the procedures described in Section 6.1 for these subgroups. The *df* used for key NSDUH analyses are summarized in Table 6.3. *For NSDUH analyses that compare two subpopulations (including those that compare subpopulations with the full population), the standard practice is to use the smaller of the two values for df to err on the side of being conservative.* For analyses where the subpopulation is not geographic in nature (e.g., members of a certain race or age category, past year users of a certain drug), the standard practice is to use the same *df* value that is used for analyses involving the whole population.

Table 6.3 Key NSDUH Analyses and Degrees of Freedom for the Restricted-Use Data File and the Public Use Data File, by Sample Design Years, 2002-2022

Analyses	Sample Design Years[1]	Degrees of Freedom for Restricted-Use (Public Use) Data File[2]
Special analyses involving the whole population or a nongeographic subpopulation[3]	2014-2022	750 (50)
	2005-2013	900 (60)
	2002-2004	900 (60)
Special analyses involving a single state	See Table 6.2	See Table 6.2
Special analyses involving other geographic subpopulations[3]	Any	Count of the unique values of VESTR (variance estimation [pseudo] stratum) for the particular geographic area of interest[4]
Detailed tables (including mental health in years before 2015) or first findings reports (FFRs) with estimates of averages, including mean age at first use[5]	2014-2022	Number of nonempty[6] strata (for each estimate/subpopulation)
	2005-2013	900 (60)
	2002-2004	900 (60)
All other detailed tables (including mental health in years before 2015) and FFRs (including geographic subpopulations)	2014-2022	750 (50)
	2005-2013	900 (60)
	2002-2004	900 (60)

[1] The NSDUH sample design variables were revised in 2005 and 2014. The 2005 revisions were applied retroactively to the 1999 through 2004 NSDUHs. Because of survey improvements in the 2002 NSDUH, the 2002 data constitute a new baseline, so this table does not include information before 2002.

[2] The degrees of freedom shown first in this column are based on the restricted-use data files, and the degrees of freedom in parentheses are based on the public use data file. State is not available on the public use data file; thus, only information on the degrees of freedom based on the restricted-use data files is provided.

[3] Some analyses capped the degrees of freedom at 900, regardless of year combinations across the study year groups. This rule is not consistently applied to all special analyses and reports.

[4] Users of the 2018 public use file (Center for Behavioral Health Statistics and Quality/Substance Abuse and Mental Health Services Administration, 2019) may find inconsistencies in the counts when comparing them with published data.

[5] Beginning with the 2018 NSDUH, estimates for mean ages at first use are no longer presented in the detailed tables or FFRs.

[6] A stratum or primary sampling unit (PSU) is *empty* for a given subpopulation if the respondent pool contains no subpopulation members in the stratum or PSU.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2002-2022.

This page intentionally left blank

7. Statistical Significance of Differences

Once the degrees of freedom (*df*) have been determined as described in Chapter 6, various methods used to compare prevalence estimates may be employed. This chapter describes the impact on significance testing from the 2014 sample redesign, the 2016 questionnaire changes, the 2017 data quality improvements, and the 2018-2022 continued sample design, as well as the methods used to compare prevalence estimates, examples showing how to compute the comparison of estimates between years, and the impact of rounding in interpreting testing results.

Customarily, the observed difference between estimates is evaluated in terms of its statistical significance. Statistical significance is based on the size of the test statistic and its corresponding *p* value, which refers to the probability that a difference as large as that observed would occur because of random variability in the sample estimates if there were no differences in the population prevalence values being compared. The significance of observed differences is generally reported at the 0.05 and 0.01 levels when the *p* value is defined as less than or equal to the designated significance level.[31]

Significance tests were conducted on differences between prevalence estimates from the 2018 National Survey on Drug Use and Health (NSDUH) and previous years of NSDUH back to 2002. *Because of survey design changes implemented in 2002, data from the 2002 NSDUH and onward should not be compared with data from survey years before 2002. Additionally, questionnaire changes in 2015 and 2016 caused many estimates to break trend. When there is a trend break in estimates, that year's estimates should not be compared with prior year estimates. For example, in the case of the 2017 questionnaire changes, 2017 estimates should not be compared with prior year estimates (see Section 2.2.1 for more information).* In some years, significance tests are also conducted on differences between prevalence estimates from combined years of survey data (e.g., 2012-2013 vs. 2014-2015); however, the 2017 and subsequent detailed tables (CBHSQ, 2018b) did not show any combined-year estimates. Within-year tests were conducted on differences between prevalence estimates for various populations (or subgroups) of interest using data from the 2018 survey. In addition to comparing subpopulations, linear trend tests for all data points across all years of interest were performed. Tests against the national average were also conducted, comparing individual subgroups with the full population for certain demographics such as region.

7.1 Impact of the 2014 Sample Redesign on Significance Testing between Years

The 2014-2017 NSDUH sample was redesigned and continues to be used for 2018-2022. The primary purpose of the redesign was to redistribute the sample sizes by state and by age group, so the sample size in each state was more proportional to the state population, and similarly for age groups (i.e., youths aged 12 to 17 and young adults aged 18 to 25 were oversampled less, and older adults aged 50 or older were undersampled less). *The change in sample design with regard to states resulted in greater precision (i.e., smaller standard errors*

[31] Starting with the 2018 detailed tables (Center for Behavioral Health Statistics and Quality [CBHSQ], 2019c), the significance of observed differences was only reported at the .05 level; the .01 level of reporting was dropped.

[SEs]) overall, and the change in sample design with regard to age groups resulted in slightly decreased precision for youths and young adults but increased precision for older adults; the increase in precision for older adults was much larger than the decrease in precision for youths and young adults.

Other sample design changes that started in 2014 included (a) using the 2010 census data (instead of projections from the 2000 census), the 2006 to 2010 American Community Surveys, and Claritas to provide more up-to-date information for constructing the sampling frame and thereby slightly increasing precision; (b) reducing the number of state sampling regions so that national, regional, and state *df* were typically reduced (e.g., from 900 in 2013 and earlier to 750 in 2014 and beyond for national estimates), but the effect on critical values of the *t*-distribution was small (i.e., relative changes all less than 1 percent); and (c) increasing the average cluster (i.e., segment) size while simultaneously reducing the number of clusters, which did not result in a significant loss of precision.

Changes (mainly increases) in the precision of estimates due to the 2014 sample redesign are likely to affect significance testing. For example, suppose an estimate in 2013 is identical to that in 2014, but the 2014 estimate is more precise; it is then possible that a test between 2013 and 2012 estimates may not be significant, but the same test between 2014 and 2012 estimates may be significant because the 2014 estimate has a smaller SE.

7.2 Impact of 2017 Data Quality Improvements on Significance Testing between Years

In 2017, several data quality improvements were made for various measures (see Sections 3.7 to 3.9 for details). Most of these data quality improvements resulted in measures that were comparable across years.[32] However, that was not the case for all the measures regarding youth reasons for receiving mental health care. Initial changes made to these measures had minimal impact on the estimates; thus, the recoded variables were revised for the 2 years presented in the 2017 detailed tables and not for years prior to 2016. Despite the recoded variables not being revised for years prior to 2016, significance tests can still be performed because the recodes are considered comparable. Users also have the option to recreate the recodes prior to 2016 as described in the 2017 public use data file codebook for significance testing (CBHSQ/Substance Abuse and Mental Health Services Administration [SAMHSA], 2018). However, the specialty mental health and education, general medicine, or child welfare reason measures had further revisions applied with a larger impact on the estimates; thus, these measures are not considered comparable with prior years, so the recoded variables were renamed for 2016 and 2017. Therefore, significance tests can currently be performed for these measures between 2016 and later years only. To perform tests between 2017 and years prior to 2016, a user would need to recreate the recoded variables prior to 2016 as described in the 2017 public use data file codebook (CBHSQ/SAMHSA, 2018), if applicable.

[32] Revised variables are not included on currently available data files prior to the 2017 NSDUH but will be included on future releases of the public use data files.

7.3 Impact of 2018-2022 Sample Design Continuation on Significance Testing between Years

The 2014-2017 sample design is a coordinated design that facilitates 50 percent overlap in third-stage units (area segments) within each successive 2-year period from 2014 through 2022. This designed sample overlap slightly increases the precision of estimates of year-to-year trends because of the expected small but positive correlation resulting from the overlapping sample between successive survey years. Because the 2018-2022 sample design is a continuation of the 2014-2017 sample allocation, there is no impact on testing across years.

7.4 Comparing Prevalence Estimates between Years

When comparing prevalence estimates, one can test the null hypothesis (no difference in the population) against the alternative hypothesis (there is a difference in the population) using the standard *t* test (with the appropriate *df*) for the difference in proportions test, expressed as

$$t_{df} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\text{var}(\hat{p}_1) + \text{var}(\hat{p}_2) - 2\text{cov}(\hat{p}_1, \hat{p}_2)}}, \quad (1)$$

or

$$t_{df} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\text{var}(\hat{p}_1) + \text{var}(\hat{p}_2) - 2\rho(\hat{p}_1, \hat{p}_2)\text{SE}(\hat{p}_1)\text{SE}(\hat{p}_2)}}, \quad (2)$$

where in both formulas, *df* = the appropriate degrees of freedom, $\hat{p}_1$ = the first prevalence estimate, $\hat{p}_2$ = the second prevalence estimate, $\text{var}(\hat{p}_1)$ = the variance of the first prevalence estimate, and $\text{var}(\hat{p}_2)$ = the variance of the second prevalence estimate. In the first formula, $\text{cov}(\hat{p}_1, \hat{p}_2)$ = covariance between $\hat{p}_1$ and $\hat{p}_2$. In the second formula, the covariance between $\hat{p}_1$ and $\hat{p}_2$ is displayed as the product of the correlation between $\hat{p}_1$ and $\hat{p}_2$ and the SEs of $\hat{p}_1$ and $\hat{p}_2$, where $\rho(\hat{p}_1, \hat{p}_2)$ = the correlation between $\hat{p}_1$ and $\hat{p}_2$ and $\text{SE}(\hat{p}_1)\text{SE}(\hat{p}_2)$ = the product of the SEs for $\hat{p}_1$ and $\hat{p}_2$ (i.e., the two formulas are equivalent; the first formula is defined in terms of the covariance, and the second is defined in terms of the correlations and SEs). Generally, the correlations between estimates in adjacent years are very small and positive; thus, ignoring the correlation in the second formula will usually result in a slightly more conservative test outcome, which is a test that is less likely to reject the null hypothesis that there is no difference in the two estimates. However, a negative correlation is possible and would result in a liberal test, which means it would be more likely to reject the null hypothesis that there is no difference in the two estimates. Additionally, the second (simplified) formula can be used in the case of two independent (i.e., uncorrelated) samples, as in the case of comparing two nonadjacent year estimates. Note that the first and second prevalence estimates may take the form of prevalence estimates from two survey years (e.g., 2017 and 2018, respectively), prevalence estimates from sets of combined survey data (e.g., 2015-2016 annual averages and 2017-2018 annual averages, respectively), or prevalence estimates for different populations of interest within a single survey year. Quick tests (where the correlation of 0 is assumed) are great

tools for gaining a better understanding of published estimates; however, the results of these quick tests should be confirmed using NSDUH data and appropriate software.

Under the null hypothesis, the test statistic *t* is a random variable that asymptotically follows a *t*-distribution. Therefore, calculated values of *t*, along with the appropriate *df*, can be used to determine the corresponding probability level (i.e., *p* value). *Whether testing for differences between years or from different populations within the same year, the covariance term in the formula for t (see formula 1 earlier) will, in general, not be equal to 0.* SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2013) is used to compute estimates of *t* along with the associated *p* values such that the covariance term is calculated by taking the sample design into account. A similar procedure and formula for *t* are used for estimated totals; however, it should be noted that because it was necessary to calculate the SE indirectly outside SUDAAN using the mean that was computed using SUDAAN for domains forced by the weighting process to match their respective U.S. Census Bureau population estimates, the corresponding test statistics also were computed indirectly outside SUDAAN. SUDAAN along with auxiliary SAS® code (SAS Institute Inc., 2017), Stata® (StataCorp LP, 2017), SAS, and R (R Core Team, 2018) examples showing the computational methods for generating *p* values of estimates of *t* for means and totals can be found in Appendix A (Exhibits A.13 through A.30).

Under the null hypothesis, the test statistic with known variances asymptotically follows a standard normal (*Z*) distribution. However, because the variances of the test statistic are estimated, its distribution is more accurately described by the *t*-distribution for finite sample sizes. A sufficiently large sample size is required for the asymptotic properties to take effect, and this is usually determined through the suppression criteria applied to the estimates (see Chapter 10). As the *df* approach infinity, the *t*-distribution approaches the *Z* distribution. *That is, because most of the statistical tests performed have 750 df (see Chapter 6), the t tests performed produce approximately the same numerical results as if a Z test had been performed.*

If SUDAAN is not available to compute the standard t test, using published estimates can provide similar pairwise testing results. When comparing prevalence estimates shown in the detailed tables with their SEs, independent *t* tests for the difference of proportions can be performed and usually will provide the same results as tests performed in SUDAAN (see Sections 7.5 and 7.6). However, where the *p* value is close to the predetermined level of significance, results may differ for two reasons: (1) the covariance term is included in the SUDAAN tests, whereas it is not included in independent *t* tests; and (2) the reduced number of significant digits shown in the published estimates may cause rounding errors in the independent *t* tests.

7.5 Example of Comparing Prevalence Estimates between Years

The following example reproduces the difference in the proportions tested between 2014 and 2015 for a measure shown in Table 2.1B of the 2015 detailed tables (CBHSQ, 2016c).[33] Table 2.1B displays the prevalence for lifetime, past year, and past month tobacco and alcohol use. This example will test the difference between 2014 and 2015 past month tobacco product use. Past month tobacco product use shown in Table 2.1B has a prevalence estimate of

[33] Although this example references estimates from the 2015 detailed tables, similar examples can be found in the 2016 detailed tables onward.

25.2 percent in 2014 and 23.9 percent in 2015. The corresponding SEs shown in Table 2.1D are 0.28 percent for 2014 and 0.26 percent for 2015. Assuming that the source data are not available and/or the user does not have access to appropriate software (i.e., SUDAAN), the second *t* test formula provided earlier in this chapter can be used with the assumption that the correlation is 0.

Note that

$$\text{var}(\hat{p}_i) = (\text{SE}(\hat{p}_i))^2 ,$$

$$t_{750} = \frac{25.2 - 23.9}{\sqrt{0.28^2 + 0.26^2 - 2(0)(0.28)(0.26)}} = 3.4023 .$$

Using a *t* test to find the corresponding *p* value when $t = 3.4023$ and $df = 750$ results in *p* value = 0.0007. This is very close to the SUDAAN-calculated *p* value of 0.0002 provided in Table 2.1P. This example confirms that the difference between the 2014 estimate of 25.2 percent and the 2015 estimate of 23.9 percent is statistically significant at the 0.01 level as indicated by footnote b included on the 2014 estimate in Table 2.1B. Note that the calculated *p* value assuming the correlation is 0 is larger than the actual *p* value, which supports the earlier assertion that assuming the correlation is 0 results in a more conservative *p* value. Note, however, that this calculation could produce a smaller *p* value due to the use of rounded estimates from the table. (If the unrounded estimates had been available, the formula would yield a slightly larger *p* value than what is published in the tables.)

The following example uses the same formula with the unrounded estimates and the covariance from SUDAAN. The extra digits and the covariance change the *t*-score slightly, resulting in the published *p* value of 0.0002. Note that the *t* statistic from the below formula gives the same results as the test in SUDAAN.

$$t_{750} = \frac{25.23322447 - 23.89257397}{\sqrt{(0.27760137)^2 + (0.2610112)^2 - 2(0.0889702487627662)\ (0.27760137)(0.2610112)}} = 3.6859 .$$

Also note that the correlations between estimates in adjacent years are generally very small and positive, but a negative correlation is possible. Estimates with negative correlations will also be close to 0; thus, the differences in SUDAAN-calculated *p* values and *p* values calculated from published estimates using the second *t* test formula provided earlier in this chapter (where the correlation is assumed to be 0) would still be minimal, such as the small differences shown in this section. *However, where the p value is close to the predetermined level of significance, results may differ.*

7.6 Example of Comparing Prevalence Estimates between Years in Excel

Using the same numbers presented in Section 7.5, this example uses Excel functions to produce the same *p* value produced in the previous example. The same assumption is made about the correlation (i.e., it is 0) and that $\text{var}\left(\hat{p}_i\right) = \left[\text{SE}\left(\hat{p}_i\right)\right]^2$. The correlation of 0 results in the simplified formula shown below (additionally, the variances have been replaced by SEs squared).

$$t_{df} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{(\mathrm{SE}(\hat{p}_1))^2 + (\mathrm{SE}(\hat{p}_2))^2}} .$$

Excel can be used to set up a simple table (shown below) to compare prevalence estimates. Cells A2 through E2 are the known values input by the user. Cells F2 and G2 contain functions. This table could extend over several rows to aid in comparing many different pairs of prevalence estimates (i.e., data for columns A through E would have to be entered for each row, then the formulas in columns F and G could be copied for all rows).

	A	B	C	D	E	F	G
1	p_1	p_2	SE(p_1)	SE(p_2)	*df*	*t*	*p* value
2	25.2	23.9	0.28	0.26	750	3.4023	0.0007

The standardized test statistic is found using the simplified formula for t_{df}.

	A	B	C	D	E	F	G
1	p_1	p_2	SE(p_1)	SE(p_2)	*df*	*t*	*p* value
2	25.2	23.9	0.28	0.26	750	=(A2-B2)/SQRT(C2^2+D2^2)	0.0007

The Excel T.DIST.2T function then calculates the two-tailed Student's *t*-distribution, a continuous probability distribution.

	A	B	C	D	E	F	G
1	p_1	p_2	SE(p_1)	SE(p_2)	*df*	*t*	*p* value
2	25.2	23.9	0.28	0.26	750	3.4023	=T.DIST.2T(ABS(F2),E2)

Alternatively, the Excel NORM.S.DIST function can be used to calculate the standard normal cumulative distribution function because the *t*-distribution approaches the *Z* distribution as the *df* approach infinity. Tests performed having 750 *df* produce approximately the same numerical results as if a *Z* test had been performed. Note that this function refers to the test statistic as *Z* and does not require the *df* input.

	A	B	C	D	E	F	G
1	p_1	p_2	SE(p_1)	SE(p_2)	*df*	*Z*	*p* value
2	25.2	23.9	0.28	0.26	750	3.4023	=2*(1-NORMSDIST(ABS(F2)))

The T.DIST.2T and NORM.S.DIST functions yield the same *p* value, 0.0007. Although not generated in all NSDUH publications, some publications do include sampling error in the form of 95 percent confidence intervals (CIs). In terms of testing for differences between prevalence estimates shown with 95 percent CIs, it is important to note that two overlapping 95 percent CIs do not imply that their estimates are statistically equivalent at the 5 percent level of significance. For additional information, see Schenker and Gentleman (2001) and Payton, Greenstone, and Schenker (2003).

7.7 Comparing Prevalence Estimates in Categorical Subgroups

In addition to examining estimates between years, significance testing is also used when comparing population subgroups defined by three or more levels of a categorical variable within a given year. In this type of situation, log-linear chi-square tests of independence of the subgroup and the prevalence variables were conducted first to control the error level for multiple comparisons. Although these tests are generally not published in the detailed tables, they can aid in report writing for NSDUH publications to verify statements implying significance, such as claiming that the prevalence for a measure of interest varies by age groups. In Appendix A, see Exhibit A.47 for example SUDAAN code, Exhibit A.48 for Stata code, Exhibit A.49 for SAS code, and Exhibit A.50 for R code showing this type of testing. *If Shah's Wald F test (transformed from the standard Wald chi-square) indicated overall significant differences, the significance of each particular pairwise comparison of interest was tested using SUDAAN analytic procedures to properly account for the sample design (RTI International, 2013).* Individual pairwise tests are also used in report writing for NSDUH publications to verify statements implying significance, such as claiming that a particular age group has the highest prevalence for a measure of interest. See Exhibits A.51 through A.54 for pairwise testing examples.

Significance testing can also compare individual subgroups with the full population (e.g., adults employed full time vs. all adults). Because this testing involves two overlapping domains, a stacked dataset that includes two records for each respondent in the overlap is needed for analysis. This type of testing was included for demographics (race/Hispanicity and region) commonly compared in the 2018 detailed tables (CBHSQ, 2019c). Tests against the national average are generally not published in the detailed tables, but they can aid in report writing for NSDUH publications to verify statements implying significance, such as claiming that the prevalence for a measure of interest is higher or lower among a certain region when compared with the national average. In Appendix A, see Exhibit A.43 for example SUDAAN code, Exhibit A.44 for Stata code, Exhibit A.45 for SAS code, and Exhibit A.46 for R code showing this type of testing.

7.8 Comparing Prevalence Estimates to Identify Linear Trends

In addition to comparing subpopulations for one year versus another year, it can also be useful to test the linear trend for all data points across all years of interest. *Linear trend testing can inform users about whether prevalence use has decreased, increased, or remained steady over the entire span of the years of interest or about changes in specific measures.* Various methods can be used to test a linear trend. Linear trend testing is produced for the detailed tables as applicable, but it is only used to aid in NSDUH report writing and is not published. These linear trend tests are implemented using the SUDAAN procedure DESCRIPT with CONTRAST statements looking across years to evaluate change over time. In Appendix A, see Exhibit A.55 for example SUDAAN code, Exhibit A.56 for example Stata code, Exhibit A.57 for SAS code, and Exhibit A.58 for R code showing this type of linear trend testing.

For linear testing within the detailed tables, the DESCRIPT procedure is used in the mass production of detailed tables *only* to aid in report writing regarding whether a particular measure has remained stable, increased, or decreased over time. This method uses the *t* test, similar to the pairwise method used when testing means between years and between demographic levels within the detailed tables. Instead of using PAIRWISE statements, type I errors (incorrectly producing

significant differences) are controlled by using orthogonal polynomial coefficients in the CONTRAST statement. Although pairwise testing gives detailed information for testing between 2 years, it does not perform as well for overall trend information and increases type I errors. Orthogonal polynomial coefficients are used for not only linear trend testing but also for higher-order trend testing within the detailed tables, such as for quadratic trends (when the trend changes at a certain time point).

The DESCRIPT procedure for linear testing within the detailed tables is a good approximation to a model-based approach. The 2014 redesign impact assessment report (RIAR) (CBHSQ, 2015c) and the 2015 RIAR (CBHSQ, 2017c) also include linear trend testing and implemented the testing using a model-based approach—specifically, linear regression, logistic regression, and multinomial logistic regression models—to determine whether there were breaks in trends for the most current year. Models were also run and stratified by age and state group. The more complex model-based approach was used to incorporate more information about the outcome into the models (i.e., what type of data are being modeled) and to allow for multiple covariates, which helped determine whether there was a break in trend. This model-based approach was specific to the RIARs, of which the 2015 RIAR was featured as part of the *2015 Methodological Resource Book*. In Appendix A, see Exhibit A.59 for example SUDAAN code, Exhibit A.60 for example Stata code, Exhibit A.61 for example SAS code, and Exhibit A.62 for example R code showing the model-based linear trend testing.

The model-based method used in the RIARs is more flexible to measure a change in measurement over time when controlling for multiple covariates as needed. The modeling method can be used to estimate more specific measures, such as testing a year effect in a trend model that adjusts for seasonal effects and redesign effects or comparing an estimate with an estimated forecast using data up to a specified year. The modeling method may yield a slightly different result from the DESCRIPT method under similar settings. Because the purpose of the testing for the detailed tables is to test whether any observed difference across years is significant without consideration of other covariates, the DESCRIPT method was used for its simplicity to be incorporated into the table generation software under the given time constraints.

7.9 Impact of Rounding in Interpreting Testing Results

Prevalence estimates in the form of percentages are presented in the annual detailed tables and first findings reports and are rounded to the nearest 10th of a percent. *Testing between two rounded prevalence estimates can indicate significant or nonsignificant differences involving seemingly identical estimates.* The following examples using data from the 2015 detailed tables are provided to aid users in interpreting significance testing results:[34]

1. Differences between the estimate in a given year (e.g., 2014) and the estimate in the current year (e.g., 2015) are shown as statistically significant, but the percentages appear to be identical. For example, in Table 1.18B of the 2015 detailed tables (CBHSQ, 2016c), the estimate for lifetime crack use among youths aged 12 to 17 was 0.1 percent for 2014 and 2015 and was indicated as significantly different. Although the rounded estimates appear the same, the unrounded estimates were 0.1390 percent for 2014 and 0.0501 percent for 2015.

[34] Although these examples reference estimates from the 2015 detailed tables, similar examples can be found in the 2016 detailed tables onward.

2. The difference between the estimate in prior year A (e.g., 2002) and the estimate in the current year (e.g., 2015) is statistically significant, but the difference between the estimate in prior year B (e.g., 2004) and the estimate in the current year (e.g., 2015) is not significant, but the estimates for prior years A and B appear to be identical. For example, in Table 7.3B of the 2015 detailed tables (CBHSQ, 2016c), the estimate for past month crack use among people aged 12 or older is 0.2 percent for 2002, 2004, 2007, 2009, and 2012, but only the 2002 and 2007 estimates are significantly different from the 2015 estimate of 0.1 percent. Although the rounded estimates for 2002, 2004, 2007, 2009, and 2012 appear the same, the unrounded estimates were 0.2411 for 2002, 0.1940 for 2004, 0.2464 percent for 2007, 0.1973 percent for 2009, and 0.1705 percent for 2012.

This page intentionally left blank

8. Confidence Intervals

In some National Survey on Drug Use and Health (NSDUH) publications, sampling error has been quantified using 95 percent confidence intervals (CIs). CIs provide a scale to judge how close the sample statistic is likely to be to the true population parameter under repeated sampling. A 95 percent CI, which varies for each sample, is expected to capture the true population parameter in 95 percent of samples. The interval provides a value above and below the estimate and is determined by using the sampling distribution and standard error (SE). The sampling distribution translates the confidence level into the appropriate multiplier, and the SE measures how much statistics differ from the parameter because of sampling variability. Samples with more variability will result in a larger spread in the CI. *Symmetric CIs for small proportions may lead to the undesirable result of a lower CI limit that is less than 0.* Frequently, NSDUH estimates are small percentages (i.e., close to 0), and in that case, a logit transformation of the estimate provides favorable properties. For example, the logit transformation yields asymmetric interval boundaries between 0 and 1 that are more balanced with respect to the true probability that the true value falls below or above the interval boundaries. This is partly because for values close to 0, the distribution of a logit-transformed estimate approximates the normal distribution more closely than the standard estimate.

To illustrate the logit transformation method, let the proportion P_d represent the true proportion for a particular analysis domain *d*. Then the logit transformation of P_d, commonly referred to as the "log odds," is defined as

$$L = \ln[P_d / (1 - P_d)],$$

where "ln" denotes the natural logarithm.

Letting $\hat{p}_d$ be the estimate of the domain proportion, the log odds estimate becomes

$$\hat{L} = \ln[\hat{p}_d / (1 - \hat{p}_d)].$$

The lower and upper confidence limits of *L* are formed as

$$A = \hat{L} - K\left[\frac{\sqrt{\text{var}(\hat{p}_d)}}{\hat{p}_d(1 - \hat{p}_d)}\right],$$

$$B = \hat{L} + K\left[\frac{\sqrt{\text{var}(\hat{p}_d)}}{\hat{p}_d(1 - \hat{p}_d)}\right],$$

where $\text{var}(\hat{p}_d)$ is the variance estimate of $\hat{p}_d$, the quantity in brackets is a first-order Taylor series approximation of the SE of $\hat{L}$, and *K* is the critical value of the *t*-distribution associated with a specified level of confidence and degrees of freedom (*df*). For example, to produce 95 percent confidence limits for the 2018 national estimates, the value of *K* would be 1.96 based

on 750 *df*. See Chapter 6 for more details on what *df* should be used for various subpopulations in order to determine *K* appropriately.

Although the distribution of the logit-transformed estimate, $\hat{L}$, is asymptotically normal, the variance term in the CI is estimated, and a critical value from the *t*-distribution is therefore appropriate when calculating CIs. A sufficiently large sample size is required for the asymptotic properties to take effect, and this is usually determined through the suppression criteria applied to the estimates (see Chapter 10).

Applying the inverse logit transformation to *A* and *B* earlier yields a CI for $\hat{p}_d$ as follows:

$$\hat{p}_{d,lower} = \frac{1}{1+\exp(-A)},$$

$$\hat{p}_{d,upper} = \frac{1}{1+\exp(-B)},$$

where "exp" denotes the inverse log transformation. The lower and upper CI endpoints for percentage estimates are obtained by multiplying the lower and upper endpoints of $\hat{p}_d$ by 100.

The CI for the estimated domain total, $\hat{Y}_d$, as estimated by

$$\hat{Y}_d = \hat{N}_d \cdot \hat{p}_d,$$

is obtained by multiplying the lower and upper limits of the proportion CI by $\hat{N}_d$. For domain totals $\hat{Y}_d$, where $\hat{N}_d$ (weighted population total) is nonfixed, the CI approximation assumes that the sampling variation in $\hat{N}_d$ is negligible relative to the sampling variation in $\hat{p}_d$.

The following examples illustrate how to compute and use CIs of prevalence estimates. *Note that CIs of totals cannot be computed using published data from the detailed tables because this computation requires the weighted sum of the measures, which is most often not a published estimate.* In Appendix A, see Exhibit A.35 for example SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2013) code, Exhibit A.36 for Stata® code (StataCorp LP, 2017), Exhibit A.37 for SAS® code (SAS Institute Inc., 2017), and Exhibit A.38 for R code (R Core Team, 2018) on how to compute the CIs of the totals. The example in Section 8.1 computes CIs using the formulas shown earlier, the Section 8.2 example computes CIs using Excel, the Section 8.3 example shows how to use the CIs to compute SEs, and the Section 8.4 example shows how to use Excel to compute the SE from the CIs.

8.1 Example of Calculating Confidence Intervals Using Published Prevalence Estimates and Standard Errors

The following example illustrates how to determine the 95 percent CI using the prevalence estimates and SEs provided for measures shown in the detailed tables. This example uses estimates from Table 1.1B of the 2015 detailed tables (Center for Behavioral Health Statistics and Quality, 2016c), which displays the prevalence for lifetime, past year, and past month illicit drug use.[35] This example focuses on 2015 past year pain reliever use. Pain reliever use shown in Table 1.1B has a prevalence estimate of 4.7 percent in 2015. The corresponding SE shown in Table 1.1D is 0.11 percent for 2015. This example uses the formulas shown earlier to determine the 95 percent CI for the prevalence estimate of past year pain reliever use in 2015. Note that

$$\text{var}\left(\hat{p}_d\right) = \left[\text{SE}\left(\hat{p}_d\right)\right]^2 \text{; thus, } \sqrt{\text{var}\left(\hat{p}_d\right)} = \text{SE}\left(\hat{p}_d\right).$$

The log odds estimate can be defined as follows:

$$\hat{L} = \ln\left[0.047 / \left(1 - 0.047\right)\right] = -3.0095.$$

The upper and lower confidence limits of the log odds can then be defined:

$$A = -3.0095 - 1.96\left[\frac{0.0011}{0.0448}\right] = -3.0577 \text{, and}$$

$$B = -3.0095 + 1.96\left[\frac{0.0011}{0.0448}\right] = -2.9613.$$

Applying the inverse logit transformation yields the CIs' *p*:

$$\hat{p}_{d,lower} = \frac{1}{1 + \exp(3.0577)} = 0.0449 \text{, and}$$

$$\hat{p}_{d,upper} = \frac{1}{1 + \exp(2.9613)} = 0.0492.$$

Rounding to two significant digits, the 95 percent CI is therefore 4.5 to 4.9 percent.

The same CI calculated using SUDAAN is also 4.4 to 4.9 percent. The slight difference is a rounding error due to the reduced number of significant digits shown in the published estimates. However, the results are usually close. Producing the CIs for totals requires the

[35] Although this example references estimates from the 2015 detailed tables, similar examples can be found in the 2016 detailed tables onward.

weighted sum, which is generally not published. For examples to calculate CIs for means and totals, see Exhibits A.35 and A.38, respectively.

8.2 Example of Calculating Confidence Intervals in Excel Using Published Prevalence Estimates and Standard Errors

Using the same estimates presented in Section 8.1, this example uses Excel functions to produce the same CIs produced in the previous example. Recall that $\text{var}(\hat{p}_d) = \left[\text{SE}(\hat{p}_d)\right]^2$; thus, $\sqrt{\text{var}(\hat{p}_d)} = \text{SE}(\hat{p}_d)$. Excel can be used to set up a simple table (shown below) to produce the CI. Cells A2 through D2 are the known values input by the user. Cells E2 and F2 contain functions. This table could extend over several rows to aid in producing many CIs (i.e., data for columns A through D would have to be entered for each row, then the formulas in columns E and F could be copied for all rows).

	A	B	C	D	E	F
1	p_d	SE(p_d)	α	*df*	$p_{d,lower}$	$p_{d,upper}$
2	0.047	0.0011	0.05	750	0.0449	0.0492

The lower confidence limit is determined using the extended formula for $\hat{p}_{d,lower}$.

	A	B	C	D	E	F
1	p_d	SE(p_d)	α	*df*	$p_{d,lower}$	$p_{d,upper}$
2	0.047	0.0011	0.05	750	=1/(1+EXP(-(LN(A2/(1-A2)) - T.INV.2T(C2,D2)*(B2/(A2*(1-A2)))))))	0.0492

The upper limit is determined using the extended formula for $\hat{p}_{d,upper}$.

	A	B	C	D	E	F
1	p_d	SE(p_d)	α	*df*	$p_{d,lower}$	$p_{d,upper}$
2	0.047	0.0011	0.05	750	0.0449	=1/(1+EXP(-(LN(A2/(1-A2)) + T.INV.2T(C2,D2)*(B2/(A2*(1-A2)))))))

The 95 percent CI is 4.4 to 4.9 percent.

In the Excel formulas for $\hat{p}_{d,lower}$ and $\hat{p}_{d,upper}$, the Excel function T.INV.2T calculates the inverse of the two-tailed Student's *t*-distribution, a continuous probability distribution. The function arguments are T.INV.2T (probability, *df*), where probability is the probability (between 0 and 1) for which the user would want to evaluate the inverse of the two-tailed Student's *t*-distribution. This is also sometimes referred to as the alpha level. For 95 percent CIs, the alpha level is always 0.05. The example uses 750 *df* for a national estimate, but this could be adjusted for smaller areas of estimation.

8.3 Example of Calculating Standard Errors Using Published Confidence Intervals

This example illustrates how to determine the SE for an estimate when only the prevalence and 95 percent CI are provided. If a NSDUH publication provided only the prevalence estimate for 2015 past year pain reliever use (4.7 percent) and the 95 percent CI (4.4 to 4.9 percent), the reader may want to determine the SE for use in significance testing. This example uses formulas provided earlier to determine the SE for the prevalence estimate of past year pain reliever use in 2015.[36] Note that

$$\text{var}\left(\hat{p}_d\right) = \left[\text{SE}\left(\hat{p}_d\right)\right]^2 ; \text{ thus, } \sqrt{\text{var}\left(\hat{p}_d\right)} = \text{SE}\left(\hat{p}_d\right).$$

The following formula can be used to calculate A (lower CI for log odds estimate) by using the lower CI of the prevalence estimate (p).

$$\hat{p}_{d,lower} = \frac{1}{1+\exp\left(-A\right)}; \text{ thus, } A = \ln\left(\frac{\hat{p}_{d,lower}}{1-\hat{p}_{d,lower}}\right).$$

$$\ln\left(\frac{0.047}{1-0.047}\right) = -3.0095.$$

Below is the formula for A (lower limit of the log odds ratio). To get the SE, this formula can be converted as follows.

$$A = \hat{L} - K\left[\frac{\sqrt{\text{var}\left(\hat{p}_d\right)}}{\hat{p}_d\left(1-\hat{p}_d\right)}\right]; \text{ thus, } \text{SE}\left(\hat{p}_d\right) = \frac{\left(A-\hat{L}\right)\left[\hat{p}_d\left(1-\hat{p}_d\right)\right]}{-K}.$$

Recall from the Section 8.1 example that $L = -3.0095$. Thus, the SE is computed as follows:

$$\text{SE}\left(\hat{p}_d\right) = \frac{\left(-3.0577+3.0095\right)\left[0.047\left(1-0.047\right)\right]}{-1.96} = 0.0011 \text{ or } 0.11 \text{ percent.}$$

Using similar steps, the SE can be produced from the upper CI with the formulas below. Note that the denominator is positive in the SE formula when using the upper CI.

$$B = \ln\left(\frac{\hat{p}_{d,upper}}{1-\hat{p}_{d,upper}}\right), \text{ and } \text{SE}\left(\hat{p}_d\right) = \frac{\left(B-\hat{L}\right)\left[\hat{p}_d\left(1-\hat{p}_d\right)\right]}{K}.$$

[36] Although this example references estimates from the 2015 detailed tables, similar examples can be found in the 2016 detailed tables onward.

$$B = -2.9613, \text{ and } \mathrm{SE}(\hat{p}_d) = 0.0011 \text{ or } 0.11 \text{ percent.}$$

As previously mentioned, the 2015 NSDUH's Table 1.1D shows that the actual SE when calculated in SUDAAN is 0.11 percent, which is the same as the calculated 0.11 percent. *Note that the reduced number of significant digits shown in the published estimates may cause rounding errors when producing SEs from the lower or upper limits of the CIs.* This can result in SE estimates that differ when compared with the SUDAAN-calculated SE. However, SEs calculated from the lower or upper limits usually will provide the same testing results as tests performed in SUDAAN, except that results may differ when the *p* value is close to the predetermined level of significance.

8.4 Example of Calculating Standard Errors in Excel Using Published Confidence Intervals

Using the same estimates presented in Section 8.3, this example uses Excel functions to produce the same SEs from the previous example (i.e., the SUDAAN-generated SE from the 2015 NSDUH's Table 1.1D). Recall that $\mathrm{var}(\hat{p}_d) = \left[\mathrm{SE}(\hat{p}_d)\right]^2$; thus, $\sqrt{\mathrm{var}(\hat{p}_d)} = \mathrm{SE}(\hat{p}_d)$. Excel can be used to set up a simple table (shown below) to produce the SE from the upper and lower limits of the CI. Cells A2 through D2 are the known values input by the user. Cell E2 contains the function to determine the SE. This table could extend over several rows to aid in producing many SEs (i.e., data for columns A through D would have to be entered for each row, then the formula in column E could be copied for all rows). Note that once the methods used in this example have determined the SE from the CI, the methods shown in the Section 7.6 example can be used to perform independent *t* tests for differences of reported estimates in Excel.

Calculate the SE from the lower limit of the CI:

	A	B	C	D	E
1	p_d	$p_{d,lower}$	α	*df*	$SE(p_d)$
2	0.047	0.0449	0.05	750	0.0011

$\mathrm{SE}(\hat{p}_d) = 0.0011$ or 0.11 percent.

Similar to the Section 8.2 example, the Excel function T.INV.2T is used in the formula to determine the SE.

	A	B	C	D	E
1	p_d	$p_{d,lower}$	α	*df*	$SE(p_d)$
2	0.047	0.0449	0.05	750	=(((LN(B2/(1-B2)))-(LN(A2/(1-A2))))*(A2*(1-A2)))/(-T.INV.2T(C2,D2))

Calculate the SE from the upper limit of the CI:

	A	B	C	D	E
1	p_d	$p_{d,upper}$	α	*df*	$SE(p_d)$
2	0.047	0.0492	0.05	750	0.0011

$\mathrm{SE}(\hat{p}_d) = 0.0011$ or 0.11 percent.

This also requires the use of the Excel function T.INV.2T (see details in Section 8.2).

	A	B	C	D	E
1	p_d	$p_{d,upper}$	α	*df*	$SE(p_d)$
2	0.047	0.0492	0.05	750	=(((LN(B2/(1-B2)))-(LN(A2/(1-A2))))*(A2*(1-A2)))/(T.INV.2T(C2,D2))

Remember that the reduced number of significant digits shown in the published estimates may cause rounding errors when producing SEs. This can result in SE estimates that differ when using the lower or upper limit when compared with the SUDAAN-calculated SE. However, SEs calculated from the lower or upper limits usually will provide the same testing results as tests performed in SUDAAN, except results may differ when the *p* value is close to the predetermined level of significance.

This page intentionally left blank

9. Initiation Estimates

Since its inception in the 2004 National Survey on Drug Use and Health (NSDUH), past year initiation refers to respondents whose date of first use of a particular substance (or misuse of psychotherapeutic drugs) was within the 12 months before their interview date. Beginning in 2015, based on questionnaire changes regarding use and misuse of psychotherapeutic drugs (pain relievers, tranquilizers, stimulants, and sedatives), past year initiation for these psychotherapeutic drugs now refers to the first time that misuse occurred rather than a respondent's first use.[37] Past year initiation is determined by self-reported past year use, age at first use, year and month of recent new use, and the interview date.[38]

Since 1999, the survey questionnaire has collected year and month of first use for recent initiates (i.e., individuals who used a particular substance for the first time at their current age or the year before their current age). Month, day, and year of birth also are obtained directly or are imputed for item nonrespondents as part of the data postprocessing. Additionally, the computer-assisted interviewing instrument records and provides the date of the interview.

The calculation of past year initiation does not take into account whether the respondent initiated substance use while a resident of the United States. This method of calculation has little effect on past year estimates and provides direct comparability with other standard measures of substance use because the populations of interest for the measures will be the same (i.e., both measures examine all possible respondents and do not restrict to those only initiating substance use in the United States).

One important note for initiation estimates is the relationship between a main substance category and subcategories of substances (e.g., hallucinogens would be a main category, and lysergic acid diethylamide [LSD], phencyclidine [PCP], and Ecstasy would be subcategories in relation to hallucinogens). For most measures of substance use, any member of a subcategory is by necessity a member of the main category (e.g., if a respondent is a past month user of Ecstasy, then he or she is also a past month user of any hallucinogen). However, this is not the case with regard to estimates for the initiation of substances. *Because an individual can be an initiate of a particular substance category (main or subcategory) only a single time, a respondent with lifetime use of a subcategory may not, by necessity, be included as an initiate of the corresponding main category, even if he or she were an initiate for a different subcategory.* For example, an individual can initiate use of any hallucinogen, LSD, PCP, or Ecstasy only once. A respondent who initiated use of any hallucinogen more than 12 months ago by definition is not a past year initiate of hallucinogen use, even if he or she initiated use of LSD, PCP, or Ecstasy in the past year. For prescription drugs, see Section 9.1 for specifics on how initiation is defined.

In addition to estimates of the number of people initiating use of a substance in the past year, estimates of the mean age of past year first-time users of these substances can be

[37] For brevity, "misuse" is not repeated in every instance that text refers to first use. Readers are advised that terms such as "past year use" and "first use" that are used in the remainder of this chapter for substance use in general refer to *misuse* for prescription psychotherapeutic drugs.

[38] "Self-reported" refers to responses provided by the respondents within the questionnaire. Responses are imputed for respondents who do not self-report for these items. Day-of-first-use data are imputed because this information is not asked in the questionnaire.

computed. The 2018 detailed tables did not show any mean age of past year initiates, but *in some detailed tables prior to the 2018 NSDUH, estimates of mean age at initiation in the past 12 months have been restricted to people aged 12 to 49 so that the mean age estimates reported are not influenced by those few respondents who were past year initiates at age 50 or older.* As a measure of central tendency, means are influenced heavily by the presence of extreme values in the data, and this age constraint of 12 to 49 should increase the utility of these results to health researchers and analysts by providing a better picture of the substance use initiation behaviors among the U.S. civilian, noninstitutionalized population. This constraint was applied only to estimates of mean age at first use and does not affect estimates of initiation.

9.1 Initiation of Misuse of Prescription Psychotherapeutic Drugs

Starting in the 2015 NSDUH, respondents were asked about the initiation of misuse of prescription psychotherapeutic drugs for only the individual prescription drugs that they had misused in the past 12 months. If respondents reported initiation of one or more prescription drugs at an age or in a year and month that was more than 12 months before the interview date, they logically were not past year initiates for misuse of any drug in that psychotherapeutic category (e.g., pain relievers). If respondents reported only past year initiation of the drugs that they misused in the past 12 months, they were asked a follow-up question to determine whether they ever misused any drug in that category more than 12 months before the interview.[39] *Therefore, unlike the situation for other substances in NSDUH (see below), respondents' status as past year initiates of misuse of any psychotherapeutic drug in an overall category was determined principally through their answers to the relevant follow-up question.*

If respondents answered the follow-up question as "yes," then they were defined as *not* being past year initiates for the overall category; the affirmative response indicated that respondents had misused one or more other drugs in the category more than 12 months ago. Respondents who answered the follow-up question as "no" *were* defined as past year initiates for the overall entire category; the negative response indicated that these respondents did not misuse any other drug in that category more than 12 months ago. If respondents answered the follow-up question on initiation as "don't know" or "refused," then their status as a past year initiate (or not) was resolved through imputation.

Because of this question structure for identifying individuals who initiated misuse of any psychotherapeutic drug in a given category in the past year, measures of the age and date of first misuse of any psychotherapeutic drug in that category were created only for respondents who were past year initiates. If past year initiates had no missing data for the age, year, and month when they first misused any drug in that category, then the age, year, and month of first misuse logically were assigned from the earliest reports.[40] If past year initiates did not know or refused to report the age when they first misused some drugs in that category, but they reported first misuse of at least one psychotherapeutic drug in the category at the age that was 1 year younger

[39] Respondents also were asked the follow-up question if the sum of the reports of past year initiation plus missing data for initiation equaled the number of specific drugs that they misused in the past year (i.e., and there were no reports of initiation of misuse more than 12 months before the interview date).

[40] The questionnaire included items for the age, year, and month of first misuse for each individual psychotherapeutic drug that respondents misused in the past year. A day of first misuse was imputed for past year initiates.

than their current age, then it nevertheless could be logically inferred that this was the age when these past year initiates first misused any drug in that category. Similarly, if past year initiates did not know or refused to report the year when they first misused some drugs in that category but they reported first misuse of at least one psychotherapeutic drug in the previous calendar year (e.g., 2017 for respondents in the 2018 NSDUH), then it could be logically inferred that respondents initiated misuse of any drug in that category in the previous calendar year. If it was not possible to assign a definite age, year, and month of first misuse for a past year initiate based on the respondent's questionnaire data, then these values were assigned through imputation.

The total number of past year initiates of misuse of any psychotherapeutic drug in a category can be used in the estimation of percentages among (1) all individuals in the population (or all individuals in a subgroup of the population, such as individuals in a given age group) and (2) individuals who were past year users of the substance. The 2018 NSDUH detailed tables show estimates for these two percentages (Center for Behavioral Health Statistics and Quality [CBHSQ], 2019c). Because of the change in focus starting with the 2015 NSDUH questions for specific psychotherapeutic drugs from the lifetime to the past year period, respondents who last misused any prescription psychotherapeutic drug in a category more than 12 months ago may underreport misuse. This is especially true if they are not presented with examples of drugs that formerly were available by prescription in the United States but are no longer available. These respondents who did not report misuse that occurred more than 12 months ago would be misclassified as still being "at risk" for initiation of misuse of prescription drugs in that psychotherapeutic category (i.e., individuals who initiated misuse more than 12 months ago are no longer at risk for initiation). *For this reason, starting with the 2015 NSDUH, the detailed tables do not show percentages for initiation of misuse of psychotherapeutic drugs among individuals who were at risk for initiation.* For more information on the impact of the 2015 survey changes on the initiation of the prescription drug misuse, see Section A.4.3 in Appendix A of the report on prescription drug use and misuse in the United States (Hughes et al., 2016).

9.2 Initiation of Use of Substances Other Than Prescription Psychotherapeutic Drugs

For substances other than prescription psychotherapeutic drugs (i.e., cigarettes, smokeless tobacco, cigars, alcohol, cocaine, crack cocaine, heroin, hallucinogens, inhalants, and methamphetamine), past year initiation among people using a substance in the past year can be viewed as an indicator variable defined as follows:

$$I_{(\text{Past Year Initiate})} \text{ if } [(\text{MM/DD/YYYY})_{\text{Interview}} - (\text{MM/DD/YYYY})_{\text{First Use of Substance}}] \leq 365,$$

where $(\text{MM/DD/YYYY})_{\text{Interview}}$ denotes the month, day, and year of the interview, and $(\text{MM/DD/YYYY})_{\text{First Use of Substance}}$ denotes the date of first use. The total number of past year initiates can be used in the estimation of different percentages. For these substances, denominators for the percentages vary according to whether estimates are being calculated for (1) all individuals in the population (or all individuals in a subgroup of the population, such as individuals in a given age group), (2) individuals who are at risk for initiation because they have

not used the substance of interest before the past 12 months, or (3) past year users of the substance. The detailed tables show all three of these percentages.

Note that the 12-month reference period (i.e., 365 days) is set up on the calendar at the beginning of the audio computer-assisted self-interviewing portion of the computer-assisted interview. For example, if the date of the interview (DOI) is December 1, 2016 (12/01/2016), then 365 days earlier would be December 1, 2015 (12/01/2015). If a respondent's date of first use is the same as the DOI, then the respondent is considered a past year initiate (because I = 0). Additionally, in this example, a respondent interviewed on 12/01/2016 could have used for the first time as far back as 12/01/2015 and be considered a past year initiate.

Potential Undercoverage of Past Year Initiates

Because NSDUH respondents are aged 12 or older at the time of the interview, younger individuals (younger than 12 years) in the sample dwelling units are not eligible for selection into the NSDUH sample. *Some of these younger people may have initiated substance use during the past year. As a result, past year initiate estimates suffer from undercoverage when one can think of the estimates as reflecting all initial users regardless of current age.* For substance use estimates in 2018 that are comparable with those in earlier years,[41] data can be obtained retrospectively based on the age at and date of first use. As an example, people who were 12 years old on the date of their interview in the 2018 survey may have reported initiating use of cigarettes between 1 and 2 years ago; these people would have been past year initiates reported in the 2017 survey had people who were 11 years old on the date of the 2017 interview been allowed to participate in the survey. Similarly, estimates of past year use by younger people (aged 10 or younger) can be derived from the current survey, but they apply to initiation in prior years—not the survey year.

To get a rough estimate of the potential undercoverage of individuals younger than 12 years in the current year, reports of substance use initiation reported by people aged 12 or older were estimated for the years in which these people would have been 1 to 11 years younger. These estimates do not necessarily reflect behavior by people who were 1 to 11 years younger in the current survey. Instead, the data for the 11-year-olds reflect initiation in the year before the current survey, the data for the 10-year-olds reflect behavior between the 12th and 23rd month before this year's survey, and so on. A crude way to adjust for the difference in the years that the estimate pertains to without considering changes to the population is to apply an adjustment factor to each age-based estimate of past year initiates. The adjustment factor can be based on a ratio of lifetime users aged 12 to 17 in the current survey year to the same estimates for the prior applicable survey year. To illustrate the calculation, consider past year use of alcohol. In the 2015 survey, 73,115 youths who were 12 years old were estimated to have initiated use of alcohol between 1 and 2 years earlier.[42] These youths would have been past year initiates in the 2014 survey conducted on the same dates had the 2014 survey covered younger people. The estimated number of lifetime users currently aged 12 to 17 was 7,074,614 for 2015 and 7,375,125 for 2014, indicating fewer overall initiates of alcohol use among people aged 17 or

[41] Briefly, CBHSQ considers estimates in 2015 through 2018 to be comparable with those in 2002 to 2014 for cigarettes, cigars, alcohol (any use), marijuana, cocaine, crack cocaine, and heroin.

[42] Although this example references estimates from the 2015 detailed tables, similar examples can be found in the 2016 detailed tables onward.

younger in 2015. Thus, an adjusted estimate of initiation of alcohol use by people who were 11 years old in 2015 is given by

$$(\text{Estimated Past Year Initiates Aged 11})_{2014} \times \frac{(\text{Estimated Lifetime Users Aged 12 to 17})_{2015}}{(\text{Estimated Lifetime Users Aged 12 to 17})_{2014}}.$$

This yielded an adjusted estimate of 70,136 people who were 11 years old on a 2015 survey date and initiated use of alcohol in the past year:

$$73{,}115 \times \frac{7{,}074{,}614}{7{,}375{,}125} = 70{,}136.$$

A similar procedure was used to adjust the estimated number of past year initiates among respondents who would have been 10 years old on the same month and day of the month as the interview date in 2013 and for younger individuals in earlier years. The overall adjusted estimate for past year initiates of alcohol use by youths aged 11 or younger on the date of the interview in 2013 was 123,673, or about 2.6 percent of the estimate based on past year initiation by people aged 12 or older only (123,673 ÷ 4,760,846 = 0.0260). Based on similar analyses, the estimated undercoverage of past year initiates by youths aged 11 or younger on the date of the 2013 interview was 2.6 percent for cigarettes and 1.2 percent for marijuana.

The undercoverage of past year initiates aged 11 or younger also affects the mean age-at-first-use estimate. An adjusted estimate of the mean age at first use was calculated using a weighted estimate of the mean age at first use based on the 2015 survey and the numbers of youths aged 11 or younger in the past year obtained in the aforementioned analysis for estimating undercoverage of past year initiates. Analysis results based on the 2014 and 2015 NSDUHs showed that the mean age at first use was changed from 17.6 to 17.4 for alcohol, from 17.9 to 17.6 for cigarettes, and from 19.0 to 18.9 for marijuana. The decreases reported earlier are comparable with results generated in prior survey years and results generated for the 2016 survey year. Based on the 2016 NSDUH, alcohol changed from 17.4 to 17.1, cigarettes from 18.0 to 17.8, and marijuana from 19.3 to 19.2 as described in more detail in Section B.4.2 of the 2016 methodological summary and definitions (CBHSQ, 2017a).

Analysis results based on the 2018 data for undercoverage of past year initiates similar to those shown earlier can be found in Section 3.4.2 in Chapter 3 of the 2018 methodological summary and definitions (CBHSQ, 2019a). Similar analysis results for inhalants, methamphetamine, Ecstasy, and overall hallucinogens are not available for 2015 because changes to the questions affected the comparability of estimates between 2014 and 2015. Specific information about undercoverage for inhalants using 2018 data can also be found in Section 3.4.2 of the 2018 methodological summary and definitions (CBHSQ, 2019a).

This page intentionally left blank

10. Suppression of Estimates with Low Precision

Direct survey estimates that were considered to be unreliable because of unacceptably large sampling errors were not reported but rather were noted by an asterisk (*). *The criteria used to assess the need to suppress direct survey estimates were based on prevalence (for proportion estimates), the relative standard error (RSE) (defined as the ratio of the standard error [SE] over the estimate), nominal (actual) sample size, and effective sample size for each estimate.*

Proportion estimates ($\hat{p}$) within the range $0 < \hat{p} < 1,$ and corresponding estimated numbers of users, were suppressed if

$$\text{RSE}[-\ln(\hat{p})] > .175 \text{ when } \hat{p} \leq .5$$

or

$$\text{RSE}[-\ln(1-\hat{p})] > .175 \text{ when } \hat{p} > .5\,.$$

The choice of .175 is arbitrary, but it roughly marks the tails of the distribution.

Based on a first-order Taylor series approximation of $\text{RSE}[-\ln(\hat{p})]$ and $\text{RSE}[-\ln(1-\hat{p})],$ the following equation was derived and used for computational purposes when applying a suppression rule dependent on effective sample sizes:

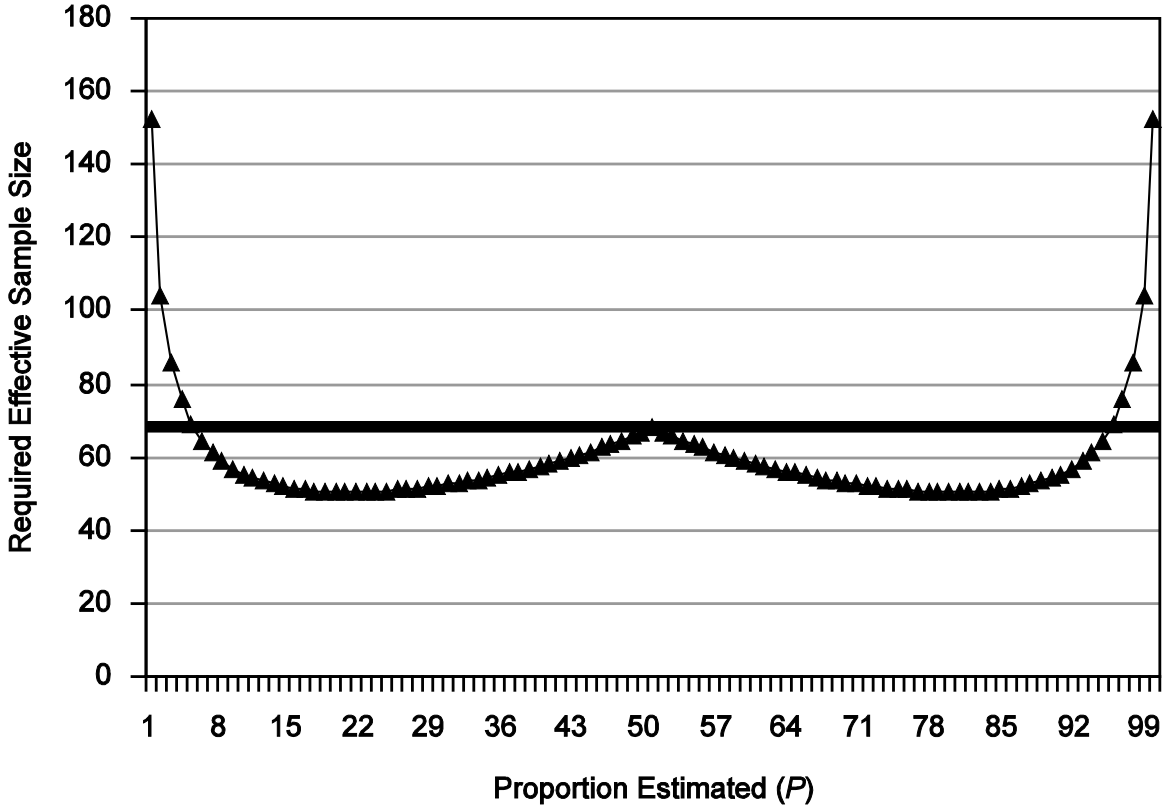
$$\frac{\text{SE}(\hat{p})/\hat{p}}{-\ln(\hat{p})} > .175 \text{ when } \hat{p} \leq .5$$

or

$$\frac{\text{SE}(\hat{p})/(1-\hat{p})}{-\ln(1-\hat{p})} > .175 \text{ when } \hat{p} > .5\,.$$

The separate formulas for $\hat{p} \leq .5$ and $\hat{p} > .5$ produce a symmetric suppression rule; that is, if $\hat{p}$ is suppressed, $1-\hat{p}$ will be suppressed as well. See Exhibit 10.1 for a graphical representation of the required minimum effective sample sizes as a function of the proportion estimated. When $.05 < \hat{p} < .95,$ the symmetric properties of the rule produce local minimum effective sample sizes at $\hat{p} = .2$ and again at $\hat{p} = .8,$ such that an effective sample size of greater than 50 is required; this means that estimates would be suppressed for these values of $\hat{p}$ unless the effective sample sizes were greater than 50. Within this same interval of $.05 < \hat{p} < .95,$ a local maximum effective sample size of 68 is required at $\hat{p} = .5.$

Exhibit 10.1 Required Effective Sample in the 2018 NSDUH as a Function of the Proportion Estimated



These varying effective sample size requirements sometimes produced unusual occurrences of suppression for a particular combination of prevalence estimates. For example, in some cases, lifetime prevalence estimates near $\hat{p} = .5$ were suppressed (effective sample size was less than 68 but greater than 50), while not suppressing the corresponding past year or past month estimates near $\hat{p} = .2$ (effective sample sizes greater than 50). *To reduce the occurrence of this type of inconsistency and to maintain a conservative suppression rule, estimates of $\hat{p}$ between 0.05 and 0.95, which had effective sample sizes below 68, were suppressed starting with the 2000 National Survey on Drug Use and Health (NSDUH).*

The effective sample size for a domain is a function of the nominal sample size and the design effect (i.e., nominal sample size/design effect). During the original development of this suppression rule, the design effect was calculated outside SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2013) in SAS® (SAS Institute Inc., 2017). Since the 2005 NSDUH analysis, the direct SUDAAN design effect was used to provide a more precise and accurate reflection of the design effect (because of the removal of several possible rounding errors) when compared with the SAS method used in the past. The differences between the direct SUDAAN design effects and the SAS-calculated design effects occur only at approximately the 10th decimal place or later; however, previously published estimates that were on the borderline of being suppressed or unsuppressed because of the effective sample size suppression rule may potentially change from suppressed to unsuppressed, or vice versa.

Design effects range widely among the measures and domains found in the detailed tables. Potential problems with suppression occur only if large design effects are combined with small domains. Large estimates of design effects when resulting from small sample sizes (variability of the variance estimate) should be suppressed on effective sample size alone, and the rule mentioned earlier achieves this. *But to protect against unreliable estimates caused by small design effects and small nominal sample sizes, a minimum nominal sample size suppression criterion (n = 100) was employed starting with the 2000 NSDUH.* Table 10.1 shows a formula for calculating design effects. *Prevalence estimates also were suppressed if they were close to 0 or 100 percent (i.e., if $\hat{p} < 0.00005$ or if $\hat{p} > 0.99995$).*

Table 10.1 Summary of 2018 NSDUH Suppression Rules

Estimate	Suppress if:
Prevalence estimate, $\hat{p}$, with nominal sample size, *n*, and design effect, *deff* $\left(deff = \frac{n[\mathrm{SE}(\hat{p})]^2}{\hat{p}(1-\hat{p})} \right)$	(1) The estimated prevalence estimate, $\hat{p}$, is < 0.00005 or > 0.99995,[1] or (2) $\frac{\mathrm{SE}(\hat{p}) \ / \ \hat{p}}{-\ln(\hat{p})} > 0.175$ when $\hat{p} \le 0.5$, or $\frac{\mathrm{SE}(\hat{p}) \ / \ (1-\hat{p})}{-\ln(1-\hat{p})} > .175$ when $\hat{p} > 0.5$, or (3) Effective *n* < 68, where $Effective\ n = \frac{n}{deff} = \frac{\hat{p}(1-\hat{p})}{\left[\mathrm{SE}(\hat{p})\right]^2}$, or (4) *n* < 100. Note: The rounding portion of this suppression rule for prevalence estimates will produce some estimates that round at one decimal place to 0.0 or 100.0 percent but are not suppressed from the tables.[2]
Estimated number (numerator of $\hat{p}$)	The estimated prevalence estimate, $\hat{p}$, is suppressed. Note: In some instances when $\hat{p}$ is not suppressed, the estimated number may appear as a 0 in the tables. This means that the estimate is greater than 0 but less than 500 (estimated numbers are shown in thousands). Note: In some instances when totals corresponding to several different means that are displayed in the same table and some, but not all, of those means are suppressed, the totals will not be suppressed. When all means are suppressed, the totals will also be suppressed.
Means not bounded between 0 and 1 (i.e., mean age at first use, mean number of drinks), $\bar{x}$, with nominal sample size, *n*	(1) $\mathrm{RSE}(\bar{x}) > 0.5$, or (2) *n* < 10.

deff = design effect; NSDUH = National Survey on Drug Use and Health; RSE = relative standard error; SE = standard error.

NOTE: The suppression rules included in this table are used for detecting unreliable estimates and are sufficient for confidentiality purposes in the context of NSDUH's first findings reports and detailed tables.

[1] Starting with the 2015 NSDUH, the close to 100 percent portion of the rule was changed to $\hat{p} > 0.99995$ instead of the old rule, which was greater than or equal to 0.99995. This was done so the close to 0 and close to 100 rules were both strict inequalities.

[2] See Chapters 3 and 7 of this report for more information on rounding.

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

Beginning with the 1991 survey, the suppression rule for proportions based on $RSE[-\ln(\hat{p})]$ described earlier replaced an older rule in which data were suppressed whenever $RSE(\hat{p}) > .5$. This rule was changed because the older rule imposed a very stringent application for small $\hat{p}$ but a very lax application for large $\hat{p}$. The new rule ensured a more uniformly stringent application across the whole range of $\hat{p}$ (i.e., from 0 to 1). The old rule also was asymmetric in the sense that suppression only occurred in terms of $\hat{p}$; that is, there was no complementary rule for $(1 - \hat{p})$, for which the new suppression rules now account.

Estimates of totals were suppressed if the corresponding prevalence estimates were suppressed. Given this rule, a user may encounter a couple of unexpected results after applying the suppression rules. One such unexpected result may occur when equivalent estimates of totals corresponding to different prevalence estimates, $\hat{p}$, are suppressed differently. To demonstrate, consider a table presenting estimates of past month substance use among different age groups (e.g., 12-17, 18-25, 12 or older), where the 12 or older prevalence estimate is not suppressed but the 12-17 estimate is suppressed. Thus, the estimated total would be displayed for the 12 or older age group only and would be suppressed for the 12-17 age group. However, if both prevalence estimates were suppressed, then both of the estimated totals would be suppressed as well. Another unexpected result may occur when $\hat{p}$ is not suppressed, but the estimated total is displayed as a zero. Because the estimated totals are shown in numbers of thousands, a zero actually represents an estimated total greater than zero but less than 500, which is appropriately displayed because the prevalence estimate was not suppressed.

Estimates of means not bounded between 0 and 1 (e.g., mean age at first use, mean number of drinks consumed) were suppressed if the RSEs of the estimates were larger than .5 or if the sample sizes were smaller than 10 respondents. This rule was based on an empirical examination of the estimates of mean age of first use and their SEs for various empirical sample sizes. *Although arbitrary, a sample size of 10 appears to provide sufficient precision and still allow reporting by year of first use for many substances.* In these cases, the totals (e.g., total number of drinks consumed) were suppressed if the corresponding mean estimates were suppressed.

Section 4 of the detailed tables demonstrates an exception to the rule that indicates the totals are suppressed when their corresponding means are suppressed. Some tables in Section 4 of the detailed tables show estimates of initiation among different populations. Specifically, these Section 4 tables display the number of initiates among three populations: the total population, people at risk for initiation, and past year users.[43] In these tables, some mean estimates may be suppressed, whereas the total estimate is not suppressed. When at least one mean estimate in the table is not suppressed, one can assume that the numerator (or total estimate) is not the cause for the suppression and the total estimate will not be suppressed. In contrast, when all mean estimates are suppressed, the total will also be suppressed.

Tables that show sample sizes and population counts do not incorporate the suppression rule for several reasons. One reason is that no mean is associated with these estimates; thus,

[43] Starting in 2015, the prescription pain reliever, prescription tranquilizer, prescription stimulant, and prescription sedative Section 4 tables do not show estimates for people at risk for initiation.

most of the components of the suppression criteria are not applicable. Also, because no behavior associated with the numbers is displayed, there is no risk of behavior disclosure.

The suppression criteria for various NSDUH estimates are summarized in Table 10.1, and sample SAS code based on both SAS and SUDAAN output, Stata® code (StataCorp LP, 2017), SAS code (SAS Institute Inc., 2017), and R code (R Core Team, 2018) demonstrating how to implement these rules can be found in Appendix A (Exhibits A.9 through A.12).

This page intentionally left blank

References

Aldworth, J., Kott, P., Yu, F., Mosquin, P., & Barnett-Walker, K. (2012). Analysis of effects of 2008 NSDUH questionnaire changes: Methods to adjust adult MDE and SPD estimates and to estimate SMI in the 2005-2009 surveys. In *2010 National Survey on Drug Use and Health: Methodological resource book* (Section 16b, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.108.005). Research Triangle Park, NC: RTI International.

American Psychiatric Association. (2008). *Diagnostic and statistical manual of mental disorders, 4th ed., text revision (DSM-IV-TR)*. Retrieved from https://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890420249.dsm-iv-tr

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

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

Center for Behavioral Health Statistics and Quality. (2010). *Results from the 2009 National Survey on Drug Use and Health: Mental health detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2012a). *Results from the 2010 National Survey on Drug Use and Health: Mental health detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2012b). *Results from the 2011 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2012c). *Results from the 2011 National Survey on Drug Use and Health: Mental health detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2012d). *Results from the 2011 National Survey on Drug Use and Health: Mental health findings* (HHS Publication No. SMA 12-4725, NSDUH Series H-45). Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2012e). *Results from the 2011 National Survey on Drug Use and Health: Summary of national findings* (HHS Publication No. SMA 12-4713, NSDUH Series H-44). Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2013a). *2011-2012 National Surveys on Drug Use and Health: Guide to state tables and summary of small area estimation methodology*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2013b). *Results from the 2012 National Survey on Drug Use and Health: Mental health detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2014a). 2012 Mental Health Surveillance Study: Design and estimation report. In *2012 National Survey on Drug Use and Health: Methodological resource book (Section 16a)*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2014b). *National Survey on Drug Use and Health: Impact of changing from 2000 to 2010 census data on comparisons of 2010-2011 and 2011-2012 model-based state estimates* (unpublished internal documentation). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2014c). *Results from the 2013 National Survey on Drug Use and Health: Mental health detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2014d). Sample design report. In *2013 National Survey on Drug Use and Health: Methodological resource book (Section 2)*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2015a). *Estimating mental illness among adults in the United States: Revisions to the 2008 estimation procedures*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2015b). *Results from the 2014 National Survey on Drug Use and Health: Mental health detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2015c). *Results from the 2014 National Survey on Drug Use and Health: Sample redesign impact assessment, final 12-month report*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2016a). *2015 National Survey on Drug Use and Health: Methodological summary and definitions*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2016b). *Key substance use and mental health indicators in the United States: Results from the 2015 National Survey on Drug Use and Health* (HHS Publication No. SMA 16-4984, NSDUH Series H-51). Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2016c). *Results from the 2015 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2016d). Statistical inference report. In *2014 National Survey on Drug Use and Health: Methodological resource book (Section 13)*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2017a). *2016 National Survey on Drug Use and Health: Methodological summary and definitions*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2017b). *Evaluation of imputation methods for the National Survey on Drug Use and Health*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2017c). *Questionnaire redesign impact assessment, final report (Volume 1)*. In *2015 National Survey on Drug Use and Health: Methodological resource book (Section 15)*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2017d). *Results from the 2016 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2017e). *Statistical inference report.* In *2015 National Survey on Drug Use and Health: Methodological resource book (Section 13)*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2018a). *2017 National Survey on Drug Use and Health: Methodological summary and definitions*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2018b). *Results from the 2017 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2018c). *Statistical inference report.* In *2016 National Survey on Drug Use and Health: Methodological resource book (Section 13)*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2019a). *2018 National Survey on Drug Use and Health: Methodological summary and definitions*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2019b). *Key substance use and mental health indicators in the United States: Results from the 2018 National Survey on Drug Use and Health* (HHS Publication No. PEP19-5068, NSDUH Series H-54). Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2019c). *Results from the 2018 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2019d). Sample design report. In *2018 National Survey on Drug Use and Health: Methodological resource book (Section 2)*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2019e). *Statistical inference report.* In *2017 National Survey on Drug Use and Health: Methodological resource book (Section 13)*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2020a). *Editing and imputation report.* In *2018 National Survey on Drug Use and Health: Methodological resource book (Section 10)*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2020b). *Person-level sampling weight calibration.* In *2018 National Survey on Drug Use and Health: Methodological resource book (Section 11)*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. (2012). *National Survey on Drug Use and Health: 2010 public use file codebook*. Retrieved from https://datafiles.samhsa.gov/study-dataset/national-survey-drug-use-and-health-2010-nsduh-2010-ds0001-nid13727

Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. (2018). *National Survey on Drug Use and Health: 2017 public use file codebook*. Retrieved from https://datafiles.samhsa.gov/study/national-survey-drug-use-and-health-nsduh-2017-nid17938

Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. (2019). *National Survey on Drug Use and Health: 2018 public use file codebook*. Retrieved from https://samhda.s3-us-gov-west-1.amazonaws.com/s3fs-public/field-uploads-protected/studies/NSDUH-2018/NSDUH-2018-datasets/NSDUH-2018-DS0001/NSDUH-2018-DS0001-info/NSDUH-2018-DS0001-info-codebook.pdf

Chen, P., Cribb, D., Dai, L., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (2013). Person-level sampling weight calibration. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, Phase II, Deliverable No. 39, RTI/0211838.207.004). Research Triangle Park, NC: RTI International.

Chromy, J. R., & Penne, M. (2002). Pair sampling in household surveys. In *Proceedings of the 2002 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, New York, NY [CD-ROM]* (pp. 552-554). Alexandria, VA: American Statistical Association. Retrieved from http://www.asasrms.org/Proceedings/index.html

Dean, E., & LeBaron, P. (2009, November). *2008 National Survey on Drug Use and Health: Context effects report* (prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-2004-00022, RTI/0209009.523.006.002). Research Triangle Park, NC: RTI International.

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

Hughes, A., Williams, M. R., Lipari, R. N., Bose, J., Copello, E. A. P., & Kroutil, L. A. (2016, September). *Prescription drug use and misuse in the United States: Results from the 2015 National Survey on Drug Use and Health*. NSDUH Data Review. Retrieved from https://www.samhsa.gov/data/

Lipari, R. N., Forsyth, B., Bose, J., Kroutil, L. A., & Lane, M. E. (2016, November). *Spouses and children of U.S. military personnel: Substance use and mental health profile*. NSDUH Data Review. Retrieved from https://www.samhsa.gov/data/

Lipari, R. N., Williams, M. R., Copello, E. A. P., & Pemberton, M. R. (2016, October). *Risk and protective factors and estimates of substance use initiation: Results from the 2015 National Survey on Drug Use and Health*. NSDUH Data Review. Retrieved from https://www.samhsa.gov/data/

Medley, G., Lipari, R. N., Bose, J., Cribb, D. S., Kroutil, L. A., & McHenry, G. (2016, October). *Sexual orientation and estimates of adult substance use and mental health: Results from the 2015 National Survey on Drug Use and Health*. NSDUH Data Review. Retrieved from https://www.samhsa.gov/data/

Office of Applied Studies. (2005). *Results from the 2004 National Survey on Drug Use and Health: National findings* (HHS Publication No. SMA 05-4062, NSDUH Series H-28). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Applied Studies. (2009a). *Results from the 2008 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Office of Applied Studies. (2009b). *Results from the 2008 National Survey on Drug Use and Health: National findings* (HHS Publication No. SMA 09-4434, NSDUH Series H-36). Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Park-Lee, E., Lipari, R. N., Hedden, S. L., Copello, E. A. P., & Kroutil, L. A. (2016, September). *Receipt of services for substance use and mental health issues among adults: Results from the 2015 National Survey on Drug Use and Health*. NSDUH Data Review. Retrieved from https://www.samhsa.gov/data/

Payton, M. E., Greenstone, M. H., & Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *Journal of Insect Science, 3*, 34. https://doi.org/10.1673/031.003.3401

Piscopo, K., Lipari, R. N., Cooney, J., & Glasheen, C. (2016, September). *Suicidal thoughts and behavior among adults: Results from the 2015 National Survey on Drug Use and Health*. NSDUH Data Review. Retrieved from https://www.samhsa.gov/data/

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

RTI International. (2013). *SUDAAN® language manual, release 11.0.1*. Research Triangle Park, NC: Author.

Ruppenkamp, J., Emrich, S., Aldworth, J., Hirsch, E., & Foster, M. (2006, February). *Missingness evaluation in the 2004 NSDUH* (draft report, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-03-9028, RTI/0208726.187.022). Research Triangle Park, NC: RTI International.

SAS Institute Inc. (2017). *SAS/STAT software: Release 14.1*. Cary, NC: Author.

Schenker, N., & Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *American Statistician, 55*(3), 182-186. https://doi.org/10.1198/000313001317097960

StataCorp LP. (2017). *Stata statistical software: Release 14*. College Station, TX: Author.

List of Contributors

This methodological report was prepared by the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality (CBHSQ), and RTI International (a registered trademark and a trade name of Research Triangle Institute). Work by RTI was performed under Contract No. HHSS283201300001C. Kathryn Piscopo served as the government project officer and as the contracting officer representative, and David Hunter served as the RTI project director.

This report was drafted by RTI and reviewed at SAMHSA. Production of the report at SAMHSA was managed by Rachel N. Lipari. Additional SAMHSA reviewers, listed alphabetically, include Rebecca D. Ahrnsbrak and Eunice Park-Lee.

Significant contributors to this report at RTI include Kat Asman, Kristen Gulledge Brown, Devon Cribb, Teresa R. Davis, Dexter McNutt, Lisa E. Packer, Jessica Roycroft, and Kathryn Spagnola. Report and web production staff at RTI, listed alphabetically, include Teresa F. Bass, Debbie F. Bond, Kimberly H. Cone, Margaret Johnson, and Pamela Tuck.

This page intentionally left blank

Appendix A: Documentation for Conducting Various Statistical Procedures: SUDAAN®, Stata®, SAS®, and R Examples

This appendix provides guidance concerning various options that should be specified in SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2013), Stata® (StataCorp LP, 2017), SAS® (SAS Institute Inc., 2017), and R (R Core Team, 2018) to correctly analyze the National Survey on Drug Use and Health (NSDUH) data. Additionally, example SUDAAN, Stata, SAS, and R code is provided to illustrate how the information in this report is applied to generate estimates (means, totals, and percentages, along with the standard errors [SEs]), implement the suppression rule, perform statistical tests of differences, handle missing data, calculate confidence intervals (CIs), test between overlapping domains, test independence of two variables, perform pairwise tests, and perform linear trend tests. Specifically, examples using 2002-2018 NSDUH data are included in this appendix that produce estimates using the statistical procedures documented within this report and implemented in the 2018 detailed tables (Center for Behavioral Health Statistics and Quality [CBHSQ], 2019c). The following examples are created using variable names found on the restricted-use dataset; thus, some variable names may differ when using the public use file (see footnote 5 in Chapter 1 for more details). Note that all the detailed tables are produced using survey analysis procedures in SUDAAN and accompanying auxiliary SAS code. However, the following Stata, SAS, and R survey analysis code replicates results from these tables. Note that results may vary slightly across software packages because of differences in precision or with regard to the default degrees of freedom. The exhibit number for each example, a description of the example, and a reference to the report chapter that addresses the example are provided in Table A.1.

Table A.1 Summary of SUDAAN, Stata, SAS, and R Exhibits

SUDAAN/ SAS Exhibit	Stata Exhibit	SAS Exhibit	R Exhibit	Description	Report Chapter
Exhibit A.1	Exhibit A.2	Exhibit A.3	Exhibit A.4	Produces estimates (including means, totals, and the respective SEs) using single- or combined-year (pooled) data.	Chapters 3, 5, and 6
Exhibit A.5	Exhibit A.6	Exhibit A.7	Exhibit A.8	Calculates the SE of the total for fixed domains using the alternative SE estimation method using the estimates produced in Exhibits A.1 through A.4.	Chapter 5
Exhibit A.9	Exhibit A.10	Exhibit A.11	Exhibit A.12	Creates suppression indicators for each estimate (i.e., suppression rule).	Chapter 10
Exhibit A.13	Exhibit A.14	Exhibit A.15	Exhibit A.16	Performs statistical tests of differences between means.	Chapter 7
Exhibit A.17	Exhibit A.18	Exhibit A.19	Exhibit A.20	Calculates the *p* value for the test of differences between totals of nonfixed domains (using estimates produced in Exhibits A.13 through A.16.	Chapter 7
Exhibits A.21, A.23, A.25, and A.27	Exhibits A.22, A.24, A.26, and A.28	Exhibit A.29	Exhibit A.30	Calculates the *p* value for the test of differences between fixed domains by producing the covariance matrix, pulling the relevant covariance components, and calculating the variances.	Chapter 7

See notes at end of table. (continued)

Table A.1 Summary of SUDAAN, Stata, SAS, and R Exhibits (continued)

SUDAAN/ SAS Exhibit	Stata Exhibit	SAS Exhibit	R Exhibit	Description	Report Chapter
Exhibit A.31	Exhibit A.32	Exhibit A.33	Exhibit A.34	Produces estimates where the variable of interest has missing values.	Chapter 4
Exhibit A.35	Exhibit A.36	Exhibit A.37	Exhibit A.38	Calculates a CI using estimates produced in Exhibits A.1 through A.4.	Chapter 8
Exhibit A.39	Exhibit A.40	Exhibit A.41	Exhibit A.42	Calculates percentages and the associated SEs.	Chapters 3 and 5
Exhibit A.43	Exhibit A.44	Exhibit A.45	Exhibit A.46	Performs statistical tests of differences between two groups when the two groups overlap.	Chapter 7
Exhibit A.47	Exhibit A.48	Exhibit A.49	Exhibit A.50	Performs tests of the independence of the prevalence variable and subgroup variable.	Chapter 7
Exhibit A.51	Exhibit A.52	Exhibit A.53	Exhibit A.54	Performs pairwise tests for each subgroup variable found significant in Exhibits A.47 through A.50.	Chapter 7
Exhibit A.55	Exhibit A.56	Exhibit A.57	Exhibit A.58	Performs linear trend test of significance across years using test statements.	Chapter 7
Exhibit A.59	Exhibit A.60	Exhibit A.61	Exhibit A.62	Performs linear trend test of significance across years using modeling.	Chapter 7

CI = confidence interval; SE = standard error.

Guide for Defining Options for Analyzing NSDUH Data

Before running the SUDAAN procedures, the input dataset must be sorted by the nesting variables (VESTR and VEREP), or the NOTSORTED option must be used for SUDAAN to create an internal copy of the input dataset properly sorted by the nesting variables. The SUDAAN procedure DESCRIPT can then be run to produce weighted (using ANALWT for restricted-use data files, ANALWT_C for public use files, or a pooled weight created for calculating annual averages) and unweighted sample sizes, means, totals, SEs of means and totals, and *p* values for testing of the means and totals.

Stata, SAS, and R commands can be run without the data being sorted. In these exhibits, the Stata commands svy: mean and svy: total will be used throughout, and it should be noted that Stata code is case sensitive. The SAS procedure SURVEYMEANS will be used in Exhibit A.3. Note that Stata and SAS still use VESTR and VEREP as nesting variables; however, as previously noted, the data do not need to be sorted.

The following options are specified within the SUDAAN, Stata, SAS, and R examples to correctly produce estimates using NSDUH data.

Design

Because of the complex NSDUH sample design, estimates are calculated using a method in SUDAAN that is unbiased for linear statistics. This method is based on multistage clustered sample designs where the first-stage (primary) sampling units are drawn with replacement. In SUDAAN, a user must specify DESIGN=WR (meaning with replacement). Note that with Stata and SAS, the design does not need to be indicated because the svyset command in Stata and the SURVEYMEANS procedure in SAS use Taylor linearized variance estimation as a default.

The R sample codes in this appendix mainly use the *survey* package. Users must install it before trying to run the sample codes. In addition to the *survey* package, three more R packages are used Exhibit A.4 (the first R sample code) starts with installing all necessary packages. The version information is as follows: The sample codes were tested in R version 3.5.1, with survey 3.36, haven 1.1.2, dplyr 0.7.8, and multcomp 1.4-10.

The R *survey* package allows the use of Taylor series linearization and replication weighting for variance estimation. The sample codes in this appendix use only the former one. The object, svydesign, specifies a complex survey design. The svydesign object combines data and all the survey design information needed to analyze the data.

Nesting Variables

The NSDUH nesting variables (VESTR and VEREP) are used to capture explicit stratification and to identify clustering with the NSDUH data, which are needed to compute the variance estimates correctly. Two replicates per year were defined within each variance stratum (VESTR). Each variance replicate (VEREP) consists of four segments, one for each quarter of data collection. One replicate consists of those segments that are "phasing out" or will not be used in the next survey year. The other replicate consists of those segments that are "phasing in" or will be fielded again the following year, thus constituting the 50 percent overlap between survey years. A segment stays in the same VEREP for the 2 years it is in the sample. This simplifies computing SEs for estimates based on combined data from adjacent survey years. In SUDAAN, users must use the NEST statement within one of the appropriate SUDAAN procedures. In the NEST statement, the variable for the variance stratum should be listed first, followed by the primary sampling unit variable; that is, the VESTR variable should be listed first, followed by the VEREP variable. In Stata, the nesting variables are specified in the svyset command. In SAS, users must use the STRATA and CLUSTER statements within one of the appropriate SAS procedures. VESTR should be listed in the STRATA statement, and VEREP should be listed in the CLUSTER statement. Unlike the svyset command in Stata where it needs to be called only once, the NEST statement in SUDAAN and the STRATA and CLUSTER statements in SAS will need to be used each time a user calls one of the appropriate SUDAAN or SAS procedures, respectively. Similar to Stata, once the survey design object is created with svydesign in R, the object is used subsequently with no need for repeating.

Degrees of Freedom

As described in Chapter 6 of this report, the degrees of freedom (DDF in SUDAAN and dof in Stata) are 750 for the 2018 national estimates: 144 in California; 120 each in Florida, New York, and Texas; 96 each in Illinois, Michigan, Ohio, and Pennsylvania; 60 each in Georgia, New Jersey, North Carolina, and Virginia; and 48 each in the remaining 38 states and the District of Columbia. For an analysis of a group of states, the degrees of freedom can be less than or equal to the sum of the degrees of freedom for each individual state due to overlap of variance strata. The specific number of degrees of freedom can be computed by counting the unique values of VESTR for the particular geographic area of interest. The technique of counting the number of unique values of VESTR can also be used for analyses combining survey data across years. When combining any years of data (i.e., 2017 and 2018), the degrees of freedom remain the same as if it were a single year (e.g., 750 for national estimates) when these years are

part of the same sample design. When comparing estimates in two domains with different degrees of freedom, one should err on the conservative side and use the smaller degrees of freedom. To specify the degrees of freedom in SUDAAN, the DDF = option on the procedure statement is used. This option should be used each time one of the appropriate SUDAAN procedures is called to ensure correct calculations. In Stata, the degrees of freedom are specified as a design option in the svyset command; that is, "dof(750)." If switching from national estimates to state estimates, the svyset command would need to be rerun with the updated degrees of freedom. In SAS, degrees of freedom can be specified by using the DF = option in the MODEL statement. For the Taylor series method, the default degrees of freedom equal the number of clusters minus the number of strata (in our case, 1500-750). In R, if DF are not provided, the DF are estimated from a model. A function, degf, is used to extract DF from a survey design. See an example of using degf in R sample codes.

Design Effect

The option DEFT4 within SUDAAN provides the correct measure of variance inflation due to stratification (or blocking), clustering, and unequal weighting in NSDUH estimation. Requesting deff srssubpop in Stata or DEFF="REPLACE" in R gives the same result as using DEFT4 in SUDAAN. The design effect cannot be output directly from the SURVEYMEANS procedure in SAS. In the following exhibits, the UNIVARIATE procedure with the VARDEFF=WGT option is used to correctly calculate the variance under simple random sampling.

The following examples apply the specific NSDUH options described previously to compute estimates, apply the suppression rule, and perform significance testing by using the data produced by the examples in Exhibit A.1 (using SUDAAN code), Exhibit A.2 (using Stata code), Exhibit A.3 (using SAS code), and Exhibit A.4 (using R code). Note that the data produced by the example in Exhibit A.3 are used only to calculate estimates and SEs and to apply the suppression rule.

Generation of Estimates

Exhibits A.1 through A.4 demonstrate how to compute various types of estimates for past month alcohol use by year and gender for single- or combined-year (pooled) data using the SUDAAN descript procedure, the Stata svy: mean and svy: total commands, the SAS SURVEYMEANS procedure, and the R svytotal and svymean commands, respectively. The SUDAAN example includes code to compute the prevalence estimate (MEAN), SE of the mean (SEMEAN), weighted sample size (WSUM), unweighted sample size (NSUM), weighted total (TOTAL), and SE of the totals (SETOTAL). The Stata svy: mean and svy: total commands, the SAS SURVEYMEANS procedure, and the R code svymean and svytotal commands will produce the same estimates. Whether the SETOTAL is taken directly from SUDAAN, Stata, SAS, or R depends on whether the specified domain (i.e., gender in this example) is fixed (i.e., domains forced to match their respective U.S. Census Bureau population estimates through the weight calibration process). See the next section in this appendix for additional information on SEs. For more information on how to create a pooled weight to use when producing annual averages of combined years of data, see Chapter 3.

Exhibit A.1 SUDAAN DESCRIPT Procedure (Estimate Generation: Single Year and Pooled Years of Data)

```
PROC SORT DATA=DATANAME; /*SAS code to sort output dataset by
Nesting Variables*/
BY VESTR VEREP;
RUN;

PROC DESCRIPT DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS DEFT4;
/*Alternatively, the DOF may change if using combined data based
on whether or not the combined years cross survey designs*/
NEST VESTR VEREP;
WEIGHT ANALWT;  /*Standard single-year, person-level analysis
weight. Alternatively, a created pooled weight could be used here
to produce annual averages based on combined years of data.*/
VAR ALCMON;  /*Past month alcohol analysis variable*/
SUBGROUP YEAR IRSEX;
     /*Year variable, where 2017=1 & 2018=2. Alternatively, the
     year variable could identify the combined years of data,
     i.e., 2015 and 2016 = 1 & 2017 and 2018 = 2*/
     /*Gender variable, where male=1 & female=2*/
LEVELS 2 2;
TABLES YEAR*IRSEX; /*Gender by year*/
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 YEAR AND GENDER";
RUN;
```

Note: The following CLASS statement could be used in place of SUBGROUP and LEVELS statements in the above example:

```
CLASS YEAR IRSEX;
```

Exhibit A.2 Stata COMMANDS svy: mean and svy: total (Estimate Generation: Single Year and Pooled Years of Data)

```
use using ".\\dataname.dta", clear

/*Ensure all variables are lower case*/
rename *, lower

/*ID Nesting variables (VESTR and VEREP) and weight variable (ANALWT -
standard single-year, person-level analysis weight). Alternatively, a
created pooled weight could be used here to produce annual averages
based on combined years of data. The DOF may also change if using
combined data based on whether or not the combined years cross survey
designs*/
svyset verep [pw=analwt], strata(vestr) dof(750)

gen total_out=.
gen setotal=.
```

Exhibit A.2 Stata COMMANDS svy: mean and svy: total (Estimate Generation: Single Year and Pooled Years of Data) (continued)

```
gen mean_out=.
gen semean=.
gen nsum=.
gen wsum=.
gen deffmean=.

/*Estimated means of past month alcohol use by year and gender*/

 /*Year variable, where 2017=1 & 2018=2. Alternatively, the year
variable could identify the combined years of data, i.e., 2015 and
2016 = 1 & 2017 and 2018 = 2*/
 /*Gender variable, where male=1 & female=2*/
svy: mean alcmon, over(year 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(deffsub) /*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 year and gender.*/
 local counter=1
  forvalues i=1/2 { /*number of years*/
   forvalues j=1/2 { /* number of gender categories*/
      replace mean_out=(M[1,`counter']) if year==`i' & irsex==`j'
      replace semean=(sqrt(S[`counter',`counter'])) ///

if year==`i' & irsex==`j'
      replace nsum=(N[1,`counter']) if year==`i' & irsex==`j'
      replace wsum=(W[1,`counter']) if year==`i' & irsex==`j'
      replace deffmean=(D[1,`counter']) if year==`i' & irsex==`j'
 local counter=`counter'+1
   }
  }

/*Estimated Totals*/
svy: total alcmon, over(year 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 year and gender.*/
```

Exhibit A.2 Stata COMMANDS svy: mean and svy: total (Estimate Generation: Single Year and Pooled Years of Data) (continued)

```
local counter=1
  forvalues i=1/2 { /*number of years*/
   forvalues j=1/2 { /* number of gender categories*/
    replace total_out=(M[1,`counter']) if year==`i' & irsex==`j'
    replace setotal=(sqrt(S[`counter',`counter'])) ///
if year==`i' & irsex==`j'
  local counter=`counter'+1
   }
  }

keep wsum mean_out semean total_out setotal nsum deffmean year irsex

duplicates drop year 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 year irsex wsum nsum mean_out semean total_out setotal

/*The output from this exhibit will be utilized in Exhibit A.26. Users
can either rerun the code presented in this exhibit or save the output
from this exhibit to a dataset using the following command.*/
save ".\\EXa2.dta" , replace
```

Exhibit A.3 SAS SURVEYMEANS Procedure (Estimate Generation: Single Year and Pooled Years of Data)

```
TITLE "ESTIMATES OF PAST MONTH ALCOHOL BY YEAR AND GENDER";
PROC SURVEYMEANS DATA=DATANAME SUMWGT NOBS MEAN SUM;
STRATA VESTR;  /*Nesting variable - strata*/
CLUSTER VEREP;  /*Nesting variable - PSU*/
WEIGHT ANALWT;  /*Standard single-year, person-level analysis
weight. Alternatively, a created pooled weight could be used here
to produce annual averages based on combined years of data.*/
VAR ALCMON;  /*Past month alcohol analysis variable*/
DOMAIN YEAR*IRSEX;  /*Gender by year*/
     /*Year variable, where 2017=1 & 2018=2. Alternatively, the
     year variable could identify the combined years of data,
     i.e., 2015 and 2016 = 1 & 2017 and 2018 = 2*/
     /*Gender variable, where male=1 & female=2*/
ODS OUTPUT DOMAIN=OUT.SASFILE;
RUN;
```

Exhibit A.4 R Code: svytotal and svymean (Estimate Generation: Single Year and Pooled Years of Data)

```
# set directory, install packages (only need to run once)
# note that R is case sensitive
#######################################################################
#install.packages("haven", lib="YOUR DIRECTORY")
#install.packages("survey", lib="YOUR DIRECTORY")
#install.packages("dplyr", lib="YOUR DIRECTORY")
#install.packages("multcomp", lib="YOUR DIRECTORY")

# setwd("YOUR DIRECTORY")
getwd()

library(haven, lib.loc="YOUR DIRECTORY")
library(survey, lib.loc="YOUR DIRECTORY")
library(dplyr, lib.loc="YOUR DIRECTORY")
library(multcomp, lib.loc="YOUR DIRECTORY")
# read NSDUH SAS files: 2017, 2018, 2-year data with selected
variables
#######################################################################
# used haven to read NSDUH SAS dataset
# used DATANAME as generic r dataframe object throughout
# YEAR variable, where 2017=1 & 2018=2. Alternatively, the year
variable
# could identify the combined years of data, i.e., 2015 and 2016 = 1
and
# 2017 and 2018 = 2
DATANAME<-read_sas("SAS DATASET PATH AND NAME",
col_select=c(QUESTID,VESTR,VEREP,ANALWT,YEAR,IRSEX,ALCMON,MRJMDAYS,MRJ
MON,CIGMON,IRWRKSTAT18,CATAG18))
#convert variable names to lower case
names(DATANAME)<-tolower(names(DATANAME))
```

Exhibit A.4 R Code: svytotal and svymean (Estimate Generation: Single Year and Pooled Years of Data) (continued)

```
# Setup for survey data analysis for Multiple Exhibits
#####################################################################

design <-
svydesign(
    id = ~ verep ,
    strata = ~ vestr ,
    data = DATANAME ,
    weights = ~ analwt ,
    nest = TRUE
)

#add new columns and change variable types
design <-
  update(design,
    one = 1 ,
    yearfactor =
      factor(
        year ,
        levels = 1:2 ,
        labels = c( "2017" , "2018" ) ) ,
    irsex =
      factor(
      irsex ,
      levels = 1:2 ,
      labels = c( "male" , "female")),
    mrjmdays =
      factor(
      mrjmdays,
      levels = 1:5 ,
      labels = c("1=1-2 days","2=3-5 days","3=6-19 days","4=20+
days","5=did not use in the past month")),

yearcombined=ifelse(year %in% c('2017', '2018'), 1, 0),
year0=ifelse(year==1, 1, 0),
year1=ifelse(year==2, 1, 0),
sexmale=ifelse(irsex=='male', 1, 0),

sexfemale=ifelse(irsex=='female', 1, 0)
  )

# degrees of freedom
degf( design )

# sample domain N
##pooled two year
sum(weights(design , "sampling" ) != 0 )
##each year (1=2017; 2=2018)
svyby( ~ one , ~ year , design , unwtd.count )
```

Exhibit A.4 R Code: svytotal and svymean (Estimate Generation: Single Year and Pooled Years of Data) (continued)

```
##sex in pooled two years
svyby( ~ one , ~ irsex , design , unwtd.count )
##sex by each year
svyby( ~ one , ~ year+irsex , design , unwtd.count )

# weighted sample domain N
##each year (1=2017; 2=2018)
svytotal(~one, design) %>% round #Pooled two years
svyby(~one, ~year, design, FUN=svytotal) %>% round
## sex in pooled two year
svyby( ~ one , ~ irsex , design , FUN=svytotal)
## by gender and year
svyby( ~ one , ~ year+irsex , design , FUN=svytotal )
# proportion estimate: past month Alcohol use
##pooled two years
svymean(~alcmon, design, deff = "replace") %>% round(2)

##each year
svyby(~alcmon, ~year, design, svymean, deff = "replace" )
##sex in pooled two years
svyby(~alcmon, ~irsex, design, svymean, deff = "replace" )
##sex by each year
svyby(~alcmon, ~year+irsex, design, svymean, deff = "replace" )

# count estimate: past month Alcohol drinker number total
svytotal(~alcmon, design) #pooled two years
svyby(~alcmon, ~year, design, svytotal )# by year
svyby(~alcmon, ~irsex, design, svytotal )#by gender in pooled two
years
svyby(~alcmon, ~year+irsex, design, svytotal ) #by gender year
```

Standard Errors

As discussed in Chapter 5 of this report, the SE for the mean (or proportion) comes directly out of SUDAAN, SAS, and R in the output variables SEMEAN (Exhibit A.1), STDERR (Exhibit A.3), and SE (using SVYBY in Exhibit A.4), respectively, and the SEMEAN is calculated in Stata by taking the square root of the variance (Exhibit A.2). However, to compute the SE of the totals, NSDUH implements different methods depending on whether the specified domain (i.e., gender in this example) is fixed or nonfixed. For the 2018 detailed tables (CBHSQ, 2019c), Table 5.1 in Chapter 5 contains a list of what are considered fixed domains. If a domain is nonfixed (e.g., not forced to match the U.S. Census Bureau population estimates), then the SE of the total comes directly out of SUDAAN, SAS, and R in the output variables SETOTAL, STDDEV, and SE (using svymean in Exhibit A.4), respectively. If the domain is fixed (e.g., forced to match the U.S. Census Bureau population estimates), then the SE of the total is calculated using an alternative SE estimation method; that is, SETOTAL (SE of fixed domain) = WSUM (weighted sample size) × SEMEAN (SE for the mean/proportion). Because gender is a fixed domain, the SE of the totals would not be taken directly from the examples in Exhibits A.1 through A.4 but rather would be computed using the alternative SE estimation method shown in

Exhibits A.5 through A.8 (note that the alternative method is the same in all three exhibits; Exhibits A.1 and A.5 use SUDAAN/SAS code, Exhibits A.2 and A.6 use Stata code, Exhibits A.3 and A.7 use SAS code, and Exhibits A.4 and A.8 use R code).

Exhibit A.5 SAS Code Based on SUDAAN Output (Calculation of Standard Error of Totals for Fixed Domains)

```
DATA ESTIMATE;
SET OUT.SUDFILE; /*input the output file from above SUDAAN
                   procedure*/
/************************************************************
 Define SETOTAL for gender because it is a fixed domain.
  In the SUDAAN procedure in Exhibit A.1, IRSEX is in the
subgroup
      Statement with 2 levels indicated. Therefore, values for
      0=total male & females, 1=males, and 2=females are
      automatically produced.
************************************************************/

IF IRSEX IN (0,1,2) THEN SETOTAL=WSUM*SEMEAN;

RUN;
```

Exhibit A.6 Stata Code (Calculation of Standard Error of Totals for Fixed Domains)

```
generate setotal2=wsum*semean
replace setotal = setotal2 if inlist(irsex,1,2)
/*Note, Stata does not automatically produce overall estimates,
i.e., irsex=0*/
```

Exhibit A.7 SAS Code Based on SAS Output (Calculation of Standard Error of Totals for Fixed Domains)

```
DATA SASEST;
SET OUT.SASFILE; /*input the output file from above SAS procedure
in Exhibit A.3 */

SETOTAL=SUMWGT*STDERR;

RUN;
```

Exhibit A.8 R Code (Calculation of Standard Error of Totals for Fixed Domains)

```
# gender in nsduh is a fixed domain. Accordingly, for count estimate
of
# past month Alcohol drinkers, corrected SE is computed here.

# compute the corrected SE by gender and year here.
# weighted sample domain N by gender year
wdomain=svyby(~one, ~year+irsex, design, svytotal )
# SE of proportion estimate of Alcohol drinker by gender and year
SEprop=svyby(~alcmon, ~year+irsex, design, svymean )
combined=cbind(subset(wdomain, select=-c(se)), subset(SEprop,
select=c(se))) #combine two stats together
     combined$SE.FixedDomain=combined$one*combined$se; combined
#compute

# Repeat for combined gender by year

wdomaintot=svyby(~one, ~year, design, svytotal )

SEproptot=svyby(~alcmon, ~year, design, svymean )

combinedtot=cbind(subset(wdomaintot, select=-c(se)), subset(SEproptot,
select=c(se))) #combine two stats together

combinedtot$SE.FixedDomain=combinedtot$one*combinedtot$se; combinedtot
#compute
```

Suppression Rule

As described in Chapter 10 of this report, each published NSDUH estimate goes through a suppression rule to detect whether the estimate is unreliable because of an unacceptably large sampling error. The suppression rules as they apply to different types of estimates are shown in Table 10.1 in Chapter 10. The examples in Exhibit A.9 (SAS code based on SUDAAN output), Exhibit A.10 (Stata code), Exhibit A.11 (SAS code), and Exhibit A.12 (R code) show the prevalence estimate rule and the rule for means not bounded by 0 and 1 (i.e., averages). The average suppression rule is commented out for these examples, but it would replace the prevalence estimate suppression rule if averages were shown in the examples in place of means bounded by 0 and 1. Note that Exhibit A.12 also calculates the design effect, which cannot be directly obtained from the SAS SURVEYMEANS procedure in Exhibit A.3.

Exhibit A.9 SAS Code Based on SUDAAN Output (Implementation of Suppression Rule)

```
      DATA ESTIMATE;
      SET OUT.SUDFILE; /*input the output file from above Exhibit A.1
SUDAAN procedure*/

      /******APPLY THE PREVALENCE ESTIMATE SUPPRESSION RULE*******/

      /* CALCULATE THE RELATIVE STANDARD ERROR */
            IF MEAN GT 0.0 THEN RSE=SEMEAN/MEAN;

      /* CALCULATE THE RELATIVE STANDARD ERROR OF NATURAL LOG P */
            IF 0.0 LT MEAN LE 0.5 THEN RSELNP=RSE/ABS(LOG(MEAN));
            ELSE IF 0.5 LT MEAN LT 1.0 THEN
            RSELNP=RSE*(MEAN/(1-MEAN))/(ABS(LOG(1-MEAN)));

      /*CALCULATE THE EFFECTIVE SAMPLE SIZE*/
            EFFNSUM=NSUM/DEFFMEAN;

      /*SUPPRESSION RULE FOR PREVALENCE ESTIMATES*/
      IF (MEAN LT 0.00005) OR (MEAN GT 0.99995) OR (RSELNP GT 0.175) OR
      (EFFNSUM < 68) OR (NSUM <100) THEN SUPRULE=1;

      /*SUPPRESSION RULE FOR MEANS NOT BOUNDED BY 0 AND 1, I.E.
      AVERAGES (COMMENTED OUT FOR THIS EXAMPLE)*/
      /*IF (RSE GT 0.5) OR (NSUM < 10) THEN SUPRULE=1;*/

      RUN;
```

Exhibit A.10 Stata Code (Implementation of Suppression Rule)

```
      /******APPLY THE PREVALENCE ESTIMATE SUPPRESSION RULE*******/

      /*CALCULATE THE RELATIVE STANDARD ERROR*/
      generate rse=.
      replace rse=semean/mean_out ///
      if mean_out > 0.0 & !missing(mean_out)

      /* CALCULATE THE RELATIVE STANDARD ERROR OF NATURAL LOG P */
      generate rselnp=.
      replace rselnp=rse/(abs(log(mean_out))) ///
      if mean_out <= 0.5 & mean_out > 0.0
      replace rselnp=rse*(mean_out/(1-mean_out)) ///
      /(abs(log(1-mean_out))) if mean_out < 1.0 & mean_out > 0.5

      /*CALCULATE THE EFFECTIVE SAMPLE SIZE*/
      generate effnsum=nsum/deffmean

      /*SUPPRESSION RULE FOR PREVALENCE ESTIMATES*/
      generate suprule1a=1 if rselnp > 0.175 & !missing(rselnp)
      generate suprule1b=1 if mean_out <.00005 & !missing(mean_out)
      generate suprule1c=1 if mean_out >.99995 & !missing(mean_out)
      generate suprule2=1 if effnsum < 68 & !missing(nsum)
```

Exhibit A.10 Stata Code (Implementation of Suppression Rule) (continued)

```
generate suprule3=1 if nsum < 100 & !missing(nsum)

generate suppress=0
replace suppress=1 if suprule1a==1 | suprule1b==1 | ///
suprule1c==1 | suprule2==1 | suprule3==1

/*SUPPRESSION RULE FOR MEANS NOT BOUNDED BY 0 AND 1, I.E.
AVERAGES
(COMMENTED OUT FOR THIS EXAMPLE)*/
/*generate suprule=1 if (nsum < 10 & !missing(nsum))///
| (rse > 0.5 & !missing(rse))*/
```

Exhibit A.11 SAS Code Based on SAS Output (Implementation of Suppression Rule)

```
/*Sort dataset by domain variables*/
PROC SORT DATA = DATANAME;
BY YEAR IRSEX;
RUN;

/*Calculate the variance under simple random sampling for gender
by year. Similar code could be run using only Year on the BY
statement to get the overall combined gender estimates*/
PROC UNIVARIATE DATA=DATANAME VARDEF=WGT;
VAR ALCMON;
WEIGHT ANALWT; /*Standard single-year, person-level analysis
weight*/
BY YEAR IRSEX; /*Gender by year*/
ODS OUTPUT MOMENTS=SASUNI;
RUN;
/*Manipulate dataset output from PROC UNIVARIATE to keep only the
domain variables and the standard error*/
DATA DEFF (RENAME = (NVALUE1 = SESRS) KEEP = YEAR IRSEX NVALUE1);
SET SASUNI;
WHERE LABEL1 = "Std Deviation";
RUN;

/*sort output dataset from Exhibit A.3 by domain variables*/
PROC SORT DATA = OUT.SASFILE;
BY YEAR IRSEX;
RUN;

/*Merge DEFF dataset with dataset output in Exhibit A.3*/
DATA SASEST_MERGE;
MERGE OUT.SASFILE DEFF;
BY YEAR IRSEX;
RUN;

DATA SASEST;
SET SASEST_MERGE;
```

Exhibit A.11 SAS Code Based on SAS Output (Implementation of Suppression Rule) (continued)

```
/*Calculate DEFF of the mean*/
DEFFMEAN = (STDERR/SESRS)**2*(N-1);

/******APPLY THE PREVALENCE ESTIMATE SUPPRESSION RULE*******/
/* CALCULATE THE RELATIVE STANDARD ERROR */
IF MEAN GT 0.0 THEN RSE=STDERR/MEAN;

/* CALCULATE THE RELATIVE STANDARD ERROR OF NATURAL LOG P */
IF 0.0 LT MEAN LE 0.5 THEN RSELNP=RSE/ABS(LOG(MEAN));
ELSE IF 0.5 LT MEAN LT 1.0 THEN RSELNP=RSE*(MEAN/(1-
MEAN))/(ABS(LOG(1-MEAN)));

/*CALCULATE THE EFFECTIVE SAMPLE SIZE*/
EFFNSUM=N/DEFFMEAN;

/*SUPPRESSION RULE FOR PREVALENCE ESTIMATES*/
IF (MEAN LT 0.00005) OR (MEAN GT 0.99995) OR (RSELNP GT 0.175) OR
(EFFNSUM < 68) OR (N <100) THEN SUPRULE=1;

/*SUPPRESSION RULE FOR MEANS NOT BOUNDED BY 0 AND 1, I.E.
AVERAGES (COMMENTED OUT FOR THIS EXAMPLE)*/
/*IF (RSE GT 0.5) OR (N < 10) THEN SUPRULE=1;*/

RUN;
```

Exhibit A.12 R Code (Implementation of Suppression Rule)

```
#Here, we focus on suppression rule for proportion estimates. See
Table 10.1 for other estimates
## proportion estimate of Alcohol use by gender and year
prop=svyby(~alcmon, ~year+irsex, design, svymean, deff = "replace");
prop
ndomain=svyby( ~ one , ~ year+irsex , design , unwtd.count ); ndomain

# sample domain N by gender and year
##combine together
prop=cbind(prop, subset(ndomain, select=c(counts))); prop
## Compute relative standard error (RSE)
prop$RSE=ifelse(prop$se > 0.0, prop$se/prop$alcmon, NA)
##Compute RSE's natural log P
prop$RSELNP=ifelse(prop$alcmon >0.0 & prop$alcmon<=0.5,
prop$RSE/abs(log(prop$alcmon)),
ifelse(prop$alcmon >0.5 & prop$alcmon<1.0,

prop$RSE*((prop$alcmon/(1-prop$alcmon))/(abs(log(1-prop$alcmon)))),
NA))

##compute effective sample size
prop$EffNsum=prop$counts/prop$DEff.alcmon
```

Exhibit A.12 R Code (Implementation of Suppression Rule) (continued)

```
#SUPPRESSION RULE FOR proportion ESTIMATES: if suprule=1 then supress;
#do not if suprule=NA
prop$suprule=ifelse(prop$alcmon < 0.00005 | prop$alcmon> 0.99995 |
prop$RSELNP > 0.175
| prop$EffNsum < 68 | prop$counts <100, 1, NA); prop

#Use for Suppression rule for means (i.e., averages, not proportion)
     #prop$suprule=ifelse((prop$RSE>0.5|prop$counts<10), 1, NA); prop
```

For tables that display totals along with multiple means from differing populations (e.g., initiation tables in Section 4 of the 2018 detailed tables [CBHSQ, 2019c]), suppression is not as straightforward as coding the rule in the SAS/SUDAAN or Stata programs. As discussed in Chapter 10, perhaps some means are suppressed and others are not suppressed. In that instance, suppression of the total estimate is based on the level of suppression present across all corresponding mean estimates. If all mean estimates associated with a total estimate are suppressed, the total estimate should also be suppressed. If at least one mean estimate is not suppressed, the total estimate is also not suppressed. The best way to ensure that this happens is to program the total estimate in the table to be suppressed if, and only if, the mean with the largest denominator is suppressed. The analyst should also check the final table to ensure that the suppression follows the rule after the program has been run.

Statistical Tests of Differences

As described in Chapter 7 of this report, significance tests were conducted on differences of prevalence estimates between the 2018 NSDUH and previous years of NSDUH back to 2002. For the 2018 detailed tables (CBHSQ, 2019c), no combined data were presented. Note that for year-to-year tests of differences, if the estimate for either year is suppressed, then the resulting *p* value is also suppressed. This is the rule used when creating the detailed tables; however, this code does not show this rule being implemented.

For the SUDAAN example (Exhibit A.13), testing of differences requires a separate PROC DESCRIPT run from the initial DESCRIPT run that produces the corresponding yearly estimates. Tests of differences can be generated using DESCRIPT's CONTRAST, PAIRWISE, or DIFFVAR statements. The SUDAAN example (Exhibit A.13) uses the DIFFVAR statement to test for differences between a pair of years (e.g., 2017 and 2018) of past month alcohol use estimates for all people aged 12 or older (IRSEX=0), all males (IRSEX=1), and all females (IRSEX=2). It also includes an example of using multiple DIFFVAR statements to test for differences between each year (i.e., 2002-2017) and the current year in this example (i.e., 2018). Similarly, for the Stata example (Exhibit A.14), a separate svy: mean command is needed. The SAS procedure SURVEYREG is used to compute the test of differences. R uses SVYTTEST from the *survey* package.

Similar to computing the SEs of the totals, calculating *p* values for tests of differences of totals differs depending on whether an estimate is considered to be from a fixed domain or a nonfixed domain. Both ways are described as follows with accompanying example code: Exhibits A.13 and A.17 show example code for nonfixed domains using SUDAAN and auxiliary SAS, Exhibits A.14 and A.18 show the same example using Stata, Exhibits A.15 and A.19 show

the example using SAS, and Exhibits A.16 and A.20 show the example using R. Exhibits A.13, A.17, A.21, A.25, and A.27 show example code for fixed domains using SUDAAN and auxiliary SAS. Exhibits A.14, A.18, A.22, A.26, and A.28 show example code for fixed domains using Stata. Exhibit A.29 shows the example using SAS, and Exhibit A.30 shows the example using R.

Exhibit A.13 SUDAAN DESCRIPT Procedure (Tests of Differences)

```
      PROC DESCRIPT DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS;
      NEST VESTR VEREP;
      WEIGHT ANALWT;
      VAR ALCMON;
      SUBGROUP YEAR IRSEX;
      LEVELS 2 2;
      TABLES IRSEX;
      DIFFVAR YEAR=(1 2) / NAME="2017 vs 2018";
      PRINT WSUM NSUM MEAN SEMEAN TOTAL SETOTAL T_MEAN P_MEAN /
        REPLACE STYLE=NCHS;
      OUTPUT WSUM MEAN SEMEAN TOTAL SETOTAL NSUM T_MEAN P_MEAN /
         REPLACE
         NSUMFMT=F8.0 WSUMFMT=F12.0 MEANFMT=F15.10 SEMEANFMT=F15.10
         TOTALFMT=F12.0 SETOTALFMT=F12.0 FILENAME="OUT.SUDTESTS";
      TITLE "TESTS OF DIFFERENCES BETWEEN 2017 AND 2018 ESTIMATES OF
      PAST MONTH ALCOHOL BY GENDER";
      RUN;
Note: For testing of multiple years vs the current year as shown in
Multiyear Detailed Tables, more years could be included in the data
(and LEVELS statement) and several DIFFVAR statements as shown below
could be used in place of the single DIFFVAR statement in the above
example:
      LEVELS 17 2;
      DIFFVAR YEAR=(1 17) /NAME="2002 vs 2018";
      DIFFVAR YEAR=(2 17) /NAME="2003 vs 2018";
      DIFFVAR YEAR=(3 17) /NAME="2004 vs 2018";
      DIFFVAR YEAR=(4 17) /NAME="2005 vs 2018";
      DIFFVAR YEAR=(5 17) /NAME="2006 vs 2018";
      DIFFVAR YEAR=(6 17) /NAME="2007 vs 2018";
      DIFFVAR YEAR=(7 17) /NAME="2008 vs 2018";
      DIFFVAR YEAR=(8 17) /NAME="2009 vs 2018";
      DIFFVAR YEAR=(9 17) /NAME="2010 vs 2018";
      DIFFVAR YEAR=(10 17) /NAME="2011 vs 2018";
      DIFFVAR YEAR=(11 17) /NAME="2012 vs 2018";
      DIFFVAR YEAR=(12 17) /NAME="2013 vs 2018";
      DIFFVAR YEAR=(13 17) /NAME="2014 vs 2018";
      DIFFVAR YEAR=(14 17) /NAME="2015 vs 2018";
      DIFFVAR YEAR=(15 17) /NAME="2016 vs 2018";
      DIFFVAR YEAR=(16 17) /NAME="2017 vs 2018";

      TITLE "TESTS OF DIFFERENCES BETWEEN EACH YEAR AND 2018 ESTIMATES
      OF PAST MONTH ALCOHOL BY GENDER";
Note: The following CLASS statement could be used in place of SUBGROUP
and LEVELS statements in the above examples:
      CLASS YEAR IRSEX;
```

When one or more contrasts are specified in SUDAAN, as in the DIFFVAR statement above, the output variable MEAN becomes the contrast mean where the number assigned to the output variable, CONTRAST, represents the tests in order of appearance in the SAS code, and SEMEAN becomes the SE of the contrast mean. The examples above also output the *t*-statistic (T_MEAN) and the corresponding *p* value (P_MEAN).

SUDAAN does not test differences in the corresponding totals explicitly. However, it will output the contrast total (TOTAL) and the SE of the contrast total (SETOTAL). With these statistics and the correct degrees of freedom (750 in this example), the *p* value (PVALT) for the test of differences between totals for nonfixed domains can be calculated as indicated in Exhibit A.17. The SAS function PROBT returns the probability from a *t*-distribution.

Exhibit A.14 Stata COMMANDS svy: mean and svy: total (Tests of Differences)

```
use using ".\\dataname.dta", clear

/*Ensure all variables are lower case*/
rename *, lower

/*ID Nesting variables (VESTR and VEREP) and weight variable
(ANALWT - standard single-year, person-level analysis weight)*/
svyset verep [pweight=analwt], strata(vestr) dof(750)
{
svy: mean alcmon, over(year irsex)
local max=2*2 /*number of years*number of gender categories. This
is the total number of subpops*/
local range=2 /*number of gender categories. This is the number
of subpops per year*/
local compmin=`max'-`range'
gen pmean=. /*P-value T-test Cont. Mean=0*/
local counter=1
forvalues i=1/1 { /*number of contrasts needed to compare year==1
vs year==2*/
      local counter2=1
      forvalues j=1/2 { /*number of gender categories*/
            local stop=`counter2'+`compmin'
            test [alcmon]_subpop_`counter' = ///
            [alcmon]_subpop_`stop', nosvyadjust
            replace pmean=r(p) if year==`i' & irsex==`j' /*p-value
      t-test cont. mean=0*/
            local counter=`counter'+1
            local counter2=`counter2'+1
      }
            }
      }

svy: total alcmon, over(year irsex)
      {
matrix M = e(b) /*The totals for each subpopulation are stored in
here*/
```

Exhibit A.14 Stata COMMANDS svy: mean and svy: total (Tests of Differences) (continued)

```
local max=2*2     /*number of years*number of gender categories.
This is the total number of subpops*/
local range=2    /*number of gender categories. This is the number
of subpops per year*/
local compmin=`max'-`range'
gen total=. /*Contrast total*/
gen setotal=.    /*Total Standard error*/
      local counter=1
      forvalues i=1/1 { /*number of contrasts needed to compare
year==1 vs year==2*/
            local counter2=1
            forvalues j=1/2 {     /*number of gender categories*/
                  local stop=`counter2'+`compmin'
                  test [alcmon]_subpop_`counter' = ///
      [alcmon]_subpop_`stop', nosvyadjust matvlc(test`counter')

                  replace setotal= sqrt((test`counter'[1,1])) ///
      if year==`i' & irsex==`j'
                  replace total=M[1,`counter']-M[1,`stop'] ///
      if year==`i' & irsex==`j' /*Calculating the difference
between the totals of the subpopulation*/
                  local counter=`counter'+1
                  local counter2=`counter2'+1
                  }
            }
      }

      *Keeping variables that match SUDAAN
      keep irsex total setotal pmean
      duplicates drop irsex total setotal pmean, force /*keep one
record per contrast*/

      drop if total ==. /* drop the rows where there is no
information */
      format pmean %-15.10f
      format total %-12.0fc
      format setotal %-12.0fc

      /* Output the dataset*/
      list irsex total setotal pmean
```

Note: For testing of multiple years vs the current year as shown in Multiyear Detailed Tables, more years could be included in the data and the number of tests conducted can be increased by changing the number of for loops as shown below. The first block of code applies to means while the second block of code applies to totals. Note, this only demonstrates how the for loops would change. The svy: statements demonstrated above would still need to be utilized.

Exhibit A.14 Stata COMMANDS svy: mean and svy: total (Tests of Differences) (continued)

```
local max=17*2 /*number of years*number of gender categories.
This is the total number of subpops*/
local range=2 /*number of gender categories. This is the number
of subpops per year*/
local compmin=`max'-`range'
gen pmean=. /*P-value T-test Cont. Mean=0*/
local counter=1
forvalues i=1/16 { /*number of contrasts needed to compare each
year to the current year*/
      local counter2=1
      forvalues j=1/2 { /*number of gender categories*/
            local stop=`counter2'+`compmin'
            test [alcmon]_subpop_`counter' = ///
            [alcmon]_subpop_`stop', nosvyadjust
            replace pmean=r(p) if year==`i' & irsex==`j' /*p-value
      t-test cont. mean=0*/
            local counter=`counter'+1
            local counter2=`counter2'+1
      }
            }
      }

local max=17*2 /*number of years*number of gender categories.
This is the total number of subpops.*/
local range=2 /*number of gender categories. This is the number
of subpops per year.*/
local compmin=`max'-`range'
gen total=. /*Contrast total*/
gen setotal=. /*Total Standard error*/
local counter=1
forvalues i=1/16 { /*number of contrasts needed to compare each
year to the current year*/
      local counter2=1
      forvalues j=1/2 { /*number of gender categories*/
            local stop=`counter2'+`compmin'
            test [alcmon]_subpop_`counter' = ///
       [alcmon]_subpop_`stop', nosvyadjust ///
       matvlc(test`counter')
            replace setotal= sqrt((test`counter'[1,1])) if ///
       year==`i' & irsex==`j'
            replace total=M[1,`counter']-M[1,`stop'] if ///
       year==`i' & irsex==`j' /*Calculating the difference between
the totals of the subpopulation*/
            local counter=`counter'+1
            local counter2=`counter2'+1
            }
      }
}
```

Exhibit A.15 SAS Code (Tests of Differences)

```
TITLE "TESTS OF DIFFERENCES BETWEEN 2017 AND 2018 ESTIMATES OF PAST
MONTH ALCOHOL BY GENDER";
      PROC SURVEYREG DATA=DATANAME;
            CLUSTER VEREP;
            STRATA VESTR;
            WEIGHT ANALWT;
            DOMAIN IRSEX;
            CLASS YEAR;
            MODEL ALCMON = YEAR /NOINT VADJUST=NONE SOLUTION COVB;
            /* option NOINT omits the intercept from the model; option
            VADJUST=NONE specifies that no variance adjustment is used;
            option SOLUTION displays the parameter estimates; option
            COVB displays the estimated covariance matrix of the
            estimated regression estimates. */
            LSMEANS YEAR / DIFF ;
            ODS OUTPUT DIFFS = OUT.SASTESTS /*output the estimates of
            difference*/
            COVB = OUT.COVB; /*output the variance covariance matrix*/
RUN;
/*Note: to compare other years, the dataset could be changed to
ALLYEAR, and add codes to restricted the years included: WHERE YEAR IN
(1, 17); */
```

Exhibit A.16 R Code (Tests of Differences)

```
#We make significance test of difference in proportion estimates of
# alcohol drinkers between 2 years, separately for each of the
following
# domains: 1) total pop; 2) male only; 3) female only
## sample count: 1=year2017, 2=year2018
count(DATANAME, year)
## sample count: 1=male, 2=female
count(DATANAME, irsex)

# 1) total pop
## proportions by year
svyby(~alcmon, ~year, design, svymean )

# sig. testing: Alcohol drinker proportion difference=year2 - year1
svyttest(alcmon~year, design)

# 2) male only
## proportions within male
svyby(~alcmon, ~year, subset(design , irsex == "male" ), svymean )

# sig. testing: Alcohol drinker proportion difference: year1 vs year2
# within male
svyttest( alcmon ~ year , subset(design , irsex == "male" ) )

# 3) female only
## proportions within female
```

Exhibit A.16 R Code (Tests of Differences) (continued)

```
svyby(~alcmon, ~year, subset(design , irsex == "female" ), svymean )
# sig. testing: Alcohol drinker proportion difference: year1 vs year2
# within female
svyttest( alcmon ~ year , subset(design , irsex == "female" ) )
```

Exhibit A.17 SAS Code Based on SUDAAN Output (Calculation of the *P* Value for the Test of Differences between Totals for Nonfixed Domains)

```
IF SETOTAL GT 0.0 THEN DO; /*SETOTAL and TOTAL come from
Exhibit A.13*/
  PVALT=2*(1-PROBT(ABS(TOTAL/SETOTAL),750));
END;
```

Exhibit A.18 Stata Code (Calculation of the *P* Value for the Test of Differences between Totals for Nonfixed Domains)

```
generate pvalt = tprob(750,abs(total /setotal)) ///
if setotal > 0 & !missing(setotal) /* two-tail*/
/*total_out and setotal come from Exhibit A.14.
*/
```

Exhibit A.19 SAS Code (Calculation of the *P* Value for the Test of Differences between Totals for Nonfixed Domains)

NOTE: In SAS, the SE of the difference between totals cannot be produced. The current SAS procedure only provides SE of difference between means.

Exhibit A.20 R Code (Calculation of the *P* Value for the Test of Differences between Estimated Number Totals for Nonfixed Domains)

```
# Sig. testing of difference in past month Alcohol drinker number
between
# two years, separately in each of the following domains: 1) total
pop; 2)
# male only; 3) female only

# year and gender are fixed domains, but here we pretend that they are
# non-fixed domains and compute p values.

# 1) among total pop
#Difference in alcohol drinker numbers between two years (2017 versus
#2018)
##estimated number of alcohol drinker
total=svytotal(~I(alcmon*year0)+I(alcmon*year1), design); total
contrast=svycontrast(total, list(diff=c(1,-1))); contrast

#calculation of p value for test of differences between totals for
#nonfixed domains
pvalueT=2*(1-pt(abs(coef(contrast)/SE(contrast)),750)); pvalueT

# 2) among male
#Difference in alcohol drinker numbers between two years among males
#(2017.male versus 2018.male)
```

Exhibit A.20 R Code (Calculation of the *P* Value for the Test of Differences between Estimated Number Totals for Nonfixed Domains) (continued)

```
#estimated number of alcohol drinker
total=svytotal(~I(alcmon*year0)+I(alcmon*year1), design=subset(design,
irsex=="male")); total
contrast=svycontrast(total, list(diff=c(1,-1))); contrast
#calculation of p value for test of differences between totals for
#nonfixed domains
pvalueT=2*(1-pt(abs(coef(contrast)/SE(contrast)),750)); pvalueT
# 3) among female
#Difference in alcohol drinker numbers between two years among females
#(2017.female versus 2018.female)
##estimated number of alcohol drinker
total=svytotal(~(alcmon*year0)+I(alcmon*year1), design=subset(design,
irsex=="female")); total
contrast=svycontrast(total, list(diff=c(1,-1))); contrast

#calculation of p value for test of differences between totals for
#nonfixed domains
pvalueT=2*(1-pt(abs(coef(contrast)/SE(contrast)),750)); pvalueT
```

In Exhibits A.1 through A.4, all people aged 12 or older and both genders are considered fixed domains. For fixed domains like these, additional steps are needed to compute similar *p* values for tests of differences. One approach uses an additional DESCRIPT procedure in SUDAAN to output the appropriate covariance matrix (Exhibit A.21), and an additional svy: mean command in Stata outputs a similar matrix (Exhibit A.22). Then, through further data manipulations, the weighted sample sizes (WSUM), variances, and the covariance of the two means (obtained from the covariance matrix) are used to generate the standard *t* test statistic. The corresponding *p* value can once again be produced using the SAS PROBT function or Stata TPROB function and calculated *t* test statistic.

For the corresponding SAS and R examples, all the steps to compute the *p* values are included in one exhibit. Exhibit A.29 (SAS) brings in the covariance matrix from Exhibit A.15, calculates the variances, then computes the *p* value using the SAS PROBT function incorporating all these components. Exhibit A.30 (R) also computes all the needed components in order to produce the *p* value using the PT function.

Exhibit A.21 SUDAAN DESCRIPT Procedure (Covariance Matrix)

```
      PROC DESCRIPT DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS DEFT4;
      NEST VESTR VEREP;
      WEIGHT ANALWT;
      VAR ALCMON;
      SUBGROUP YEAR IRSEX;
      LEVELS 2 2;
      TABLES IRSEX*YEAR;
      PRINT COVMEAN / STYLE = NCHS;
      OUTPUT / MEANCOV = DEFAULT REPLACE FILENAME="OUT.SUDCOV";
      TITLE "Variance Covariance Matrices";
RUN;
```

Exhibit A.21 SUDAAN DESCRIPT Procedure (Covariance Matrix) (continued)

```
Note: The following CLASS statement could be used in place of SUBGROUP
and LEVELS statements in the above example:
     CLASS YEAR IRSEX;
```

Exhibit A.22 Stata COMMAND svy: mean (Covariance Matrix)

```
use using ".\\dataname.dta", clear

/*Ensure all variables are lower case*/
rename *, lower

/*ID Nesting variables (VESTR and VEREP) and weight variable
(ANALWT - standard single-year, person-level analysis weight)*/

svyset verep [pweight=analwt], strata(vestr) dof(750)
svy: mean alcmon, over(year irsex)
*Save and display the Covariance Matrix
matrix M = e(V)
matrix list M
```

The covariances of the estimated means can be obtained from the output of the DESCRIPT procedure (Exhibit A.21) and svy: mean command (Exhibit A.22). The covariance matrix in SUDAAN consists of a row and column for each gender (total, male, female) and year (both years; i.e., 2017 and 2018) combination with each cell corresponding to a particular variance component (i.e., a 9 x 9 matrix). Because the rows and columns of the matrix are identical, the cells in the top half (above the diagonal) and the bottom half (below the diagonal) are identical. Table A.2 shows a shell for what the SUDAAN covariance matrix would look like for this example. The Stata matrix would look similar but with a few exceptions: Total rows and columns would not be included (i.e., year=0 and irsex=0), and the order would be reversed (i.e., year would be listed first, followed by irsex).

Table A.3 presents the Stata matrix shell.

Table A.2 SUDAAN Matrix Shell

			IRSEX=0			IRSEX=1			IRSEX=2		
			YEAR=0	YEAR=1	YEAR=2	YEAR=0	YEAR=1	YEAR=2	YEAR=0	YEAR=1	YEAR=2
IRSEX	YEAR	ROWNUM	B01	B02	B03	B04	B05	B06	B07	B08	B09
IRSEX=0	YEAR=0	1									
	YEAR=1	2									
	YEAR=2	3									
IRSEX=1	YEAR=0	4									
	YEAR=1	5									
	YEAR=2	6									
IRSEX=2	YEAR=0	7									
	YEAR=1	8									
	YEAR=2	9									

Table A.3 Stata Matrix Shell

OVER:	YEAR	IRSEX
_subpop_1:	1	1
_subpop_2:	1	2
_subpop_3:	2	1
_subpop_4:	2	2

Subpopulation	alcmon: _subpop_1	alcmon: _subpop_2	alcmon: _subpop_3	alcmon: _subpop_4
alcmon:_subpop_1				
alcmon:_subpop_2				
alcmon:_subpop_3				
alcmon:_subpop_4				

In the SUDAAN output, each cell of the variance-covariance matrix is identified by a separate variable of the form B0*x*, where *x* is a particular cell number. (Cells are numbered left to right.) The variable *ROWNUM* is an additional output variable that simply identifies the matrix row. The covariance data needed for a particular significance test can be pulled out of the matrix using SAS code. For this example, the covariance for IRSEX=0 between YEAR=1 and YEAR=2 would be B03 from ROWNUM2 or B02 from ROWNUM3. These two values would be the same in this case. The needed covariances are kept in the SAS code shown in Exhibit A.23.

The three SAS datasets created by the following examples, one containing the covariances (Exhibit A.23) and two containing the variances (Exhibit A.25), are then merged with the output dataset from the DESCRIPT procedure that generated the tests of differences (Exhibit A.13). With the proper statistics contained in one dataset, the corresponding *p* value for the tests of differences between fixed domain totals can be produced using the SAS PROBT function and calculated *t* test statistic (Exhibit A.27). Interwoven with these three SAS codes based on SUDAAN output examples are Exhibits A.24, A.26, and A.28, which show Stata code performing the same functions. For the corresponding SAS and R code (Exhibits A.29 and A.30, respectively), all the steps to compute the *p* values are included in one exhibit.

Exhibit A.23 SAS Code Based on SUDAAN Output (Identification of Covariance Components)

```
DATA COV(KEEP=IRSEX COV1);
SET OUT.SUDCOV;
IF ROWNUM=2 THEN DO; IRSEX=0; COV1=B03; END;
ELSE IF ROWNUM=8 THEN DO; IRSEX=2; COV1=B09; END;
ELSE IF ROWNUM=5 THEN DO; IRSEX=1; COV1=B06; END;

IF ROWNUM IN (2,5,8) THEN OUTPUT;

RUN;

PROC SORT DATA=COV;
BY IRSEX;
RUN;
```

Exhibit A.24 Stata Code (Identification of Covariance Components)

```
local max=2*2    /*number of years*number of gender categories.
This is the total number of subpops*/
local range=2    /*number of gender categories. This is the number
of subpops per year*/
local compmin=`max'-`range'

gen cov1=1
local counter=1
forvalues i=1/1 { /*number of contrasts needed to compare year=1
vs year=2*/
      local counter2=1
      forvalues j=1/2 {     /*number of gender categories*/
            local stop=`counter2'+`compmin'
            replace cov1=M[`j', `stop'] if irsex==`j'
            local counter=`counter'+1
            local counter2=`counter2'+1
            }
      }

duplicates drop irsex cov1, force
      list irsex cov1
      keep irsex cov1
/* Save data to network*/
save ".\\cov.dta" , replace /*Need to save dataset since Stata
can only work with one at a time*/
```

The variances of the means are calculated in separate data steps shown in Exhibits A.25 and A.26. The variance is simply the square of the SE of the mean. The SEs of the means were output in the original procedure that generated the estimates (DESCRIPT for the SUDAAN/SAS example and svy: mean for the Stata example; see Exhibits A.1 and A.2).

Exhibit A.25 SAS Code Based on SUDAAN Output (Calculation of Variances)

```
DATA EST1(KEEP=WSUM1 VAR1 YEAR IRSEX);
SET OUT.SUDFILE;
WHERE YEAR=1;
WSUM1=WSUM;
VAR1=SEMEAN**2; /*THE variance is the SEMEAN squared*/
RUN;

DATA EST2(KEEP=WSUM2 VAR2 YEAR IRSEX);
SET OUT.SUDFILE;
WHERE YEAR=2;
WSUM2=WSUM;
VAR2 = SEMEAN**2;
RUN;
```

Exhibit A.26 Stata Code (Calculation of Variances)

```
/*Run code from Exhibit A.2 or save the output from that exhibit
into a dataset then read in that dataset here then run the
remaining code.*/
/*Note: The remaining code for this exhibit will need to be run as
a block to avoid errors.*/
preserve /*keep dataset in memory*/


keep if year ==1
gen wsum1 = wsum
gen var1 = semean^2
keep wsum1 var1 year irsex

duplicates drop year irsex, force /*keep one record per
subpopulation of interest*/

save ".\\est1.dta" , replace /*Need to save dataset since Stata
could only work with one at a time*/

restore, preserve /*restore dataset back to normal and edit for
second dataset*/


keep if year==2
gen wsum2 = wsum
gen var2 = semean^2
keep wsum2 var2 year irsex

duplicates drop year irsex, force /*keep one record per
subpopulation of interest*/

save ".\\est2.dta" , replace /*Need to save dataset since Stata
could only work with one dataset at a time*/

restore, preserve
```

Exhibit A.27 SAS Code Based on SUDAAN Output (Calculation of the *P* Value for the Test of Differences between Totals for Fixed Domains)

```
DATA P_VALUE;
MERGE EST1 EST2 OUT.SUDTESTS COV;
BY IRSEX;

PVALT=2*(1-PROBT(ABS(TOTAL/SQRT(WSUM1**2*VAR1+WSUM2**2*VAR2-
      2*WSUM1*WSUM2*COV1)),750));
RUN;
```

Exhibit A.28 Stata Code (Calculation of the *P* Value for the Test of Differences between Totals for Fixed Domains)

```
/*Run code from Exhibits A.14, A.24 and A.26 then run the
remaining code to calculate the p values*/

keep irsex total

*merge by irsex for dataset est1 est2 cov
merge m:m irsex using ".\\est1.dta", generate(_merge1)
merge m:m irsex using ".\\est2.dta", generate(_merge2)
merge m:m irsex using ".\\cov.dta", generate(_merge3)
generate pvalt = tprob(750,abs(total ///
/sqrt(wsum1^2*var1+wsum2^2*var2-2*wsum1*wsum2*cov1))) /*
 two-tail*/

drop _merge1 _merge2 _merge3
list irsex year wsum1 var1 wsum2 var2 cov1 pvalt
```

Exhibit A.29 SAS Code (Covariance Matrix, Calculations of Variances, and Calculation of the *P* Value for the Test of Differences between Totals for Fixed Domains)

```
/*Option used to allow trailing blanks in variable names*/
OPTIONS VALIDVARNAME=ANY;

TITLE "CALCULATE THE P-VALUE FOR THE TEST OF DIFFERENCE BETWEEN FIXED
DOMAINS";

/*Bring in covariance dataset from Exhibit A.15*/
TITLE "COVARIANCE MATRIX";
DATA OUT.SASCOV (KEEP=IRSEX COV);
SET OUT.COVB;
      COV='YEAR 1'N;
      IF DOMAIN = 'IMPUTATION-REVISED SEX OF RESPONDENT=1' THEN
IRSEX=1;
      ELSE IF DOMAIN = 'IMPUTATION-REVISED SEX OF RESPONDENT=2' THEN
IRSEX=2;
      ELSE IRSEX=0; IF PARAMETER = 'Year 2' THEN OUTPUT;
RUN;

/*Bring in dataset from Exhibit A.3 and Calculate the Variances*/
TITLE "TESTS OF DIFFERENCES BETWEEN TOTALS";
DATA EST1 (KEEP=N1 WSUM1 VAR1 TOTAL1 VAR_TOTAL1 IRSEX);
SET OUT.SASFILE;
      WHERE YEAR=1;
      N1=N;
      WSUM1=SUMWGT;
      VAR1=STDERR**2;
      TOTAL1=SUM;
      VAR_TOTAL1=STDDEV**2;
      IF IRSEX=. THEN IRSEX=0;
RUN;
```

Exhibit A.29 SAS Code (Covariance Matrix, Calculations of Variances, and Calculation of the *P* Value for the Test of Differences between Totals for Fixed Domains) (continued)

```
DATA EST2 (KEEP=N2 WSUM2 VAR2 TOTAL2 VAR_TOTAL2 IRSEX);
SET OUT.SASFILE;
      WHERE YEAR=2;
      N2=N;
      WSUM2=SUMWGT;
      VAR2=STDERR**2;
      TOTAL2=SUM;
      VAR_TOTAL2=STDDEV**2;
      IF IRSEX=. THEN IRSEX=0;
RUN;

/*Bring in estimate of differences dataset from Exhibit A.15*/
DATA OUT.SASTESTS;
SET OUT.SASTESTS;
      IF DOMAIN = 'IMPUTATION-REVISED SEX OF RESPONDENT=1' THEN
IRSEX=1;
      ELSE IF DOMAIN = 'IMPUTATION-REVISED SEX OF RESPONDENT=2' THEN
      IRSEX=2;
      ELSE IRSEX=0;
RUN;

/*Create P-value*/
DATA OUT.P_VALUE;
MERGE EST1 EST2 OUT.SASTESTS OUT.SASCOV;
      BY IRSEX;
      TOTAL=TOTAL1-TOTAL2;
      PVALT=2*(1-PROBT(ABS(TOTAL/SQRT(WSUM1**2*VAR1+WSUM2**2*VAR2-
      2*WSUM1*WSUM2*COV)),750));
RUN;
```

Exhibit A.30 R Code (Covariance Matrix, Calculations of Variances, and Calculation of the *P* Value for the Test of Differences between Totals for Fixed Domains)

```
# We do the same sig. testing of difference as we did in Exhibit A.16.
# The difference is that here we do so correctly because gender and
year
# are fixed domains.
# Here, we make sig. testing of two estimates of alcohol drinkers
(year
# 2017 vs year 2018) by computing the correct p value
# we do so separately in each of the three population groups: 1) total
pop; 2) male ; 3) female

#pulling the relevant covariance of alcohol drinker proportion
estimates #for two years
#the covariance matrix of proportion estimates between 2 years
## 1) total pop
prop1=svyglm(alcmon~yearfactor, design); vcov1=vcov(prop1)
## 2) male only
```

Exhibit A.30 R Code (Covariance Matrix, Calculations of Variances, and Calculation of the *P* Value for the Test of Differences between Totals for Fixed Domains) (continued)

```
prop2=svyglm(alcmon~yearfactor, subset(design, irsex=="male"));
vcov2=vcov(prop2)
# 3) female only
prop3=svyglm(alcmon~yearfactor, subset(design, irsex=="female"));
vcov3=vcov(prop3)
cov1=vcov1[1,1]+vcov1[2,1] # covariance (1) total pop
cov2=vcov2[1,1]+vcov2[2,1] # covariance (2) male
cov3=vcov3[1,1]+vcov3[2,1] # covariance (3) female

# calculate the SE of alcohol drinker proportion estimates by year
## 1) total Pop
se1=svyby(~alcmon, ~year, design, svymean );SE(se1)
## 2) male
se2=svyby(~alcmon, ~year, subset(design , irsex == "male" ), svymean
); SE(se2)
## 3) female
se3=svyby(~alcmon, ~year, subset(design , irsex == "female" ), svymean
); SE(se3)

# difference in estimated alcohol drinker number totals
##estimated number of alcohol drinker: 1) total
total1=svytotal(~I(alcmon*year0)+I(alcmon*year1), design); total1
## difference 1) among total pop
contrast1=svycontrast(total1, list(diff=c(1,-1))); contrast1
##estimated number of alcohol drinker: 2) male
total2=svytotal(~I(alcmon*year0)+I(alcmon*year1),
design=subset(design, irsex=="male")); total2
# difference in male (2017.male versus 2018.male)
contrast2=svycontrast(total2, list(diff=c(1,-1))); contrast2

#estimated number of alcohol drinker: 3) female
total3=svytotal(~I(alcmon*year0)+I(alcmon*year1),
design=subset(design, irsex=="female")); total3
# difference in female (2016.female versus 2017.female)
contrast3=svycontrast(total3, list(diff=c(1,-1))); contrast3

# weighted sample N by year
wdomain1=svyby(~one, ~year, design, svytotal ); coef(wdomain1) # 1)
total pop
wdomain2=svyby(~one, ~year, subset(design , irsex == "male" ),
svytotal ); coef(wdomain2) # 2) male
wdomain3=svyby(~one, ~year, subset(design , irsex == "female" ),
svytotal ); coef(wdomain3) # 3) female

#Calculate p value for three comparisons (all, male, female)

#calculation of p value for test of differences between estimated
alcohol #drinker user totals for Fixed domains
```

Exhibit A.30 R Code (Covariance Matrix, Calculations of Variances, and Calculation of the *P* Value for the Test of Differences between Totals for Fixed Domains) (continued)

```
# 1) total pop
pvalueT1=2*(1-
pt(abs(contrast1[1]/sqrt(coef(wdomain1)[1]^2*SE(se1)[1]^2+coef(wdomain
1)[2]^2*SE(se1)[2]^2
           -2*coef(wdomain1)[1]*coef(wdomain1)[2]*cov1)),750));
      pvalueT1

# 2) male
pvalueT2=2*(1-
pt(abs(contrast2[1]/sqrt(coef(wdomain2)[1]^2*SE(se2)[1]^2+coef(wdomain
2)[2]^2*SE(se2)[2]^2
           -2*coef(wdomain2)[1]*coef(wdomain2)[2]*cov2)),750));
      pvalueT2
# 3) female
pvalueT3=2*(1-
pt(abs(contrast3[1]/sqrt(coef(wdomain3)[1]^2*SE(se3)[1]^2+coef(wdomain
3)[2]^2*SE(se3)[2]^2
           -2*coef(wdomain3)[1]*coef(wdomain3)[2]*cov3)),750));
      pvalueT3
```

Recoding and Missing Values

In the example in Exhibit A.31 (using auxiliary SAS and SUDAAN), Exhibit A.32 (using Stata), Exhibit A.33 (using SAS), and Exhibit A.34 (using R), the mean age of first use of marijuana will be calculated in two ways in each exhibit. Respondents who have never used marijuana are assigned IRMJAGE=991, and if this level is included in the analysis, then the mean age calculated will be too high. Thus, two methods are shown on how to omit this level in calculating mean age of first use of marijuana.

Exhibit A.31 SAS Code (Recoding a Variable) and SUDAAN DESCRIPT Procedure (Estimate Generation with (1) Missing Values and (2) Using Subpopulation)

```
/* Method 1, recoding unused values to missing*/

      PROC SORT DATA=DATANAME;
      BY VESTR VEREP;
      RUN;

      DATA DATANAME;
      SET DATANAME;
      IF IRMJAGE=991 THEN IRMJAGE_R=.;
      ELSE IRMJAGE_R=IRMJAGE;
      RUN;
```

Exhibit A.31 SAS Code (Recoding a Variable) and SUDAAN DESCRIPT Procedure (Estimate Generation with (1) Missing Values and (2) Using Subpopulation) (continued)

```
PROC DESCRIPT DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS DEFT4;
NEST VESTR VEREP;
WEIGHT ANALWT;  /*Standard single-year, person-level analysis
weight*/
VAR IRMJAGE_R;  /*Marijuana Age of First Use recoded analysis
variable*/
SUBGROUP IRSEX;
/*Gender variable, where male=1 & female=2*/
LEVELS 2;
TABLES IRSEX; /*Gender*/
PRINT MEAN SEMEAN / REPLACE STYLE=NCHS;
TITLE "ESTIMATES OF AGE OF FIRST USE OF MARIJUANA BY GENDER";
RUN;

/* Method 2, using subpopulation to omit the unused values*/

PROC DESCRIPT DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS DEFT4;
NEST VESTR VEREP;
WEIGHT ANALWT;  /*Standard single-year, person-level analysis
weight*/
SUBPOPN MRJFLAG=1; /*Subsetting to omit those respondents who had
never used marijuana, i.e., omitting respondents where
IRMJAGE=991*/
VAR IRMJAGE;  /*Marijuana Age of First Use analysis variable*/
SUBGROUP IRSEX;
     /*Gender variable, where male=1 & female=2*/
LEVELS 2;
TABLES IRSEX; /*Gender*/
PRINT MEAN SEMEAN / REPLACE STYLE=NCHS;
TITLE "ESTIMATES OF AGE OF FIRST USE OF MARIJUANA BY GENDER";
RUN;
```

Exhibit A.32 Stata Code (Recoding a Variable, Estimate Generation with (1) Missing Values and (2) Using Subpopulation)

```
/*Read in data*/
use using ".\\dataname.dta", clear
/*Ensure all variables are lower case*/
rename *, lower
generate irmjage_r = irmjage
replace irmjage_r =. if irmjage == 991
/*Method 1, recoding unused values to missing*/
svyset verep [pweight=analwt], strata(vestr) dof(750)
svy: mean irmjage_r, over(irsex)
/*marijuana age of first use analysis variable, gender variable*/

/*Method 2, using subpopulation to omit the unused values*/
svyset verep [pweight=analwt], strata(vestr) dof(750)
svy, subpop(mrjflag): mean irmjage, over(irsex)
```

Exhibit A.33 SAS Code (Recoding a Variable, Estimate Generation with (1) Missing Values and (2) Using Subpopulation)

```
/*Method 1, recoding unused values to missing*/
TITLE "PRODUCE ESTIMATES WHERE THE VARIABLE HAS MISSING VALUES";
DATA DATANAME;
SET DATANAME;
      IF IRMJAGE=991 THEN IRMJAGE_R=.;
      ELSE IRMJAGE_R=IRMJAGE;
RUN;

PROC SURVEYMEANS DATA=DATANAME;
      CLUSTER VEREP;
      STRATA VESTR;
      WEIGHT ANALWT;
      VAR IRMJAGE_R;
RUN;

PROC SURVEYMEANS DATA=DATANAME;
      CLUSTER VEREP;
      STRATA VESTR;
      WEIGHT ANALWT;
      DOMAIN IRSEX; /*Estimates by Gender*/
      VAR IRMJAGE_R;
RUN;

/*Method 2, using subpopulation to omit the unused values*/
PROC SURVEYMEANS DATA=DATANAME;
      WHERE MRJFLAG=1;
      CLUSTER VEREP;
      STRATA VESTR;
      WEIGHT ANALWT;
      VAR IRMJAGE;
RUN;

PROC SURVEYMEANS DATA=DATANAME;
      WHERE MRJFLAG=1;
      CLUSTER VEREP;
      STRATA VESTR;
      WEIGHT ANALWT;
      DOMAIN IRSEX; /*Estimates by Gender*/
      VAR IRMJAGE;
RUN;
```

Exhibit A.34 R Code (Estimate Generation with (1) Missing Values and (2) Using Subpopulation)

```
# Marijuana Age of First Use recoded analysis variable is used
# Data management and create survey design
##recode as missingDATANAME$irmjage_r=with(keep, ifelse(irmjage==991,
NA, irmjage))

design.A34 <-
  svydesign(
    id = ~ verep ,
    strata = ~ vestr ,
    data = DATANAME ,
    weights = ~ analwt ,
    nest = TRUE
  )

design.A34 <-
  update(design.A34,
       one = 1 ,

       yearfactor =
         factor(
           year ,
           levels = 1:2 ,
           labels = c( "2017" , "2018" ) ) ,
        irsex =
          factor(
          irsex ,
          levels = 1:2 ,
          labels = c( "male" , "female"))
  )
#Method1: recoding those who never used Marijuana as missing
#Mean Marijuana Age of First Use and its SE
mean=svyby( ~ irmjage_r , ~ irsex , design.A34 , svymean , na.rm =
TRUE ); mean
#Method2: using subpopulation to omit those who never used Marijuana
mean=svyby( ~ irmjage , ~ irsex , subset(design.A34, mrjflag==1),
svymean, na.rm = TRUE ); mean
```

Confidence Intervals

As discussed in Chapter 8 of this report, CIs can be calculated using means (MEAN) and SEs (SEMEAN) from PROC DESCRIPT in SUDAAN, svy: mean in Stata, the SURVEYMEANS procedure in SAS, and svymean in R. After the means and SEs are obtained (Exhibits A.1 through A.4), the code in Exhibits A.35 through A.38 can be used to create the 95 percent CIs for means and totals.

Exhibit A.35 SAS Code Based on SUDAAN Output (Calculating a 95 Percent Confidence Interval)

```
DATA CI;
SET OUT.SUDFILE; /*output data from Exhibit A.1*/
T_QNTILE=TINV(0.975,750); /*define t-statistic*/
NUMBER=SEMEAN/(MEAN*(1-MEAN));
L=LOG(MEAN/(1-MEAN));

A=L-T_QNTILE*NUMBER;
B=L+T_QNTILE*NUMBER;

PLOWER=1/(1+EXP(-A));
PUPPER=1/(1+EXP(-B));
/*PLOWER AND PUPPER ARE THE 95% CIS ASSOCIATED WITH MEAN FROM
SUDAAN*/
TLOWER=WSUM*PLOWER;
TUPPER=WSUM*PUPPER;
/*TLOWER AND TUPPER ARE THE 95% CIS ASSOCIATED WITH TOTAL FROM
SUDAAN*/
RUN;
```

Exhibit A.36 Stata Code (Calculating a 95 Percent Confidence Interval for a Mean)

```
/*Run code from Exhibit A.2 or save output dataset from
Exhibit A.2 and use that as input to this code.*/
generate t_qntile = invt(750,0.975)
generate number = semean/(mean_out*(1-mean_out))
generate l=log(mean_out/(1-mean_out))
generate a = l-t_qntile*number
generate b = l+t_qntile*number
generate plower = 1/(1+exp(-a))
generate pupper = 1/(1+exp(-b))

/*plower and pupper are the 95% CIs associated with mean_out from
Stata*/

generate tlower = wsum*plower
generate tupper = wsum*pupper

/*tlower and tupper are the 95% CIs associated with total_out
from Stata*/
duplicates drop year irsex, force /*keep one record per
subpopulation of interest*/

keep year irsex nsum wsum mean_out semean total_out setotal
///t_qntile number l a b plower pupper tlower tupper
```

Exhibit A.37 SAS Code (Calculates a CI for Alcohol Drinker Prevalence and Estimated Totals Produced in Exhibit A.3)

```
TITLE "CALCULATE CONFIDENCE INTERVALS";
DATA OUT.CI;
SET OUT.SASFILE; /*output data from Exhibit A.3*/
      T_QNTILE=TINV(0.975,750); /*define t-statistic*/
      NUMBER=STDERR/(MEAN*(1-MEAN));
      L=LOG(MEAN/(1-MEAN));
      A=L-T_QNTILE*NUMBER;
      B=L+T_QNTILE*NUMBER;
      PLOWER=1/(1+EXP(-A));
      PUPPER=1/(1+EXP(-B));
      /*PLOWER AND PUPPER ARE THE 95% CIS ASSOCIATED WITH MEAN*/
      TLOWER=SUMWGT*PLOWER;
      TUPPER=SUMWGT*PUPPER;
      /*TLOWER AND TUPPER ARE THE 95% CIS ASSOCIATED WITH TOTAL*/
RUN;
```

Exhibit A.38 R Code (Calculates a CI for Alcohol Drinker Prevalence and Estimated Totals Produced in Exhibit A.4)

```
# define t-statistic
T_QNTILE=qt(c(.975), df=750); T_QNTILE

# Of various domains in Exhibit A.4, we only focus on estimates by
year: 2017 #and 2018.
##alcohol drinker proportion estimates by year
prop=svyby(~alcmon, ~year, design, svymean); prop
## weighted sample N by year
wdomain=svyby(~one, ~year, design, svytotal); wdomain

# For 2017 pop
NUMBER.17=SE(prop)[1]/(coef(prop)[1]*(1-coef(prop)[1])); NUMBER.17
L.17=log(coef(prop)[1]/(1-coef(prop)[1])); L.17
A.17=L.17-T_QNTILE*NUMBER.17; A.17
B.17=L.17+T_QNTILE*NUMBER.17; B.17

# For 2018 pop
NUMBER.18=SE(prop)[2]/(coef(prop)[2]*(1-coef(prop)[2])); NUMBER.18
L.18=log(coef(prop)[2]/(1-coef(prop)[2])); L.18
A.18=L.18-T_QNTILE*NUMBER.18; A.18
B.18=L.18+T_QNTILE*NUMBER.18; B.18

# PLOWER AND PUPPER ARE THE 95% CIS ASSOCIATED WITH prevalence
PLOWER.17=1/(1+exp(-A.17)); PLOWER.17 # for 2017 pop
PUPPER.17=1/(1+exp(-B.17)); PUPPER.17 # for 2017 pop
PLOWER.18=1/(1+exp(-A.18)); PLOWER.18 # for 2018 pop
PUPPER.18=1/(1+exp(-B.18)); PUPPER.18 # for 2018 pop
```

Exhibit A.38 R Code (Calculates a CI for Alcohol Drinker Prevalence and Estimated Totals Produced in Exhibit A.4) (continued)

```
# TLOWER AND TUPPER ARE THE 95% CIS ASSOCIATED WITH estimated total N
TLOWER.17=coef(wdomain)[1]*PLOWER.17; TLOWER.17 # for 2017 pop
TUPPER.17=coef(wdomain)[1]*PUPPER.17; TUPPER.17 # for 2017 pop
TLOWER.18=coef(wdomain)[2]*PLOWER.18; TLOWER.18 # for 2018 pop
TUPPER.18=coef(wdomain)[2]*PUPPER.18; TUPPER.18 # for 2018 pop
```

Calculating Percentages for Categories

Exhibits A.39 through A.42 demonstrate how to compute estimates corresponding to levels of a categorical variable. This example uses the number of days used marijuana in the past month among past month marijuana users. The variable that will be analyzed (MRJMDAYS) is a categorical variable with days grouped into four levels (1=1-2 days, 2=3-5 days, 3=6-19 days, 4=20+ days). Because SUDAAN now needs to estimate percentages and SEs for each level of the variable instead of computing only one estimate for the variable overall, the CATLEVEL statement is introduced, and the PERCENT and SEPERCENT keywords replace the MEAN and SEMEAN keywords. Note that the suppression rule for percentages is the same as the suppression rule for means shown in Exhibit A.9, except PERCENT and SEPERCENT have to be divided by 100 (and thus are equivalent to MEAN and SEMEAN in the formulas). The same would apply to the means output in SAS and percentages output in R that come out as percentages and would need to be divided by 100 before applying the suppression rule. In Stata, the output will be proportions that can be directly used in the suppression rule formulas. However, if for reporting purposes, percentages need to be shown, then these proportions would need to be multiplied by 100.

Exhibit A.39 SUDAAN DESCRIPT Procedure (Frequency of Use; i.e., Number of Days Used Substance in the Past Month among Past Month Users)

```
PROC DESCRIPT DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS DEFT4;
NEST VESTR VEREP;
WEIGHT ANALWT;  /*Standard single-year, person-level analysis
weight*/
VAR MRJMDAYS MRJMDAYS MRJMDAYS MRJMDAYS;  /*Marijuana Use frequency
in the past month variable: 1=1-2 days, 2=3-5 days, 3=6-19 days,
4=20+ days, 5=did not use in the past month*/
CATLEVEL 1 2 3 4; /*levels of MRJMDAYS to be shown in table*/
SUBGROUP MRJMON;
/*Past month marijuana use variable, where used in past month=1 &
   did not use in past month=0*/
LEVELS 1;
TABLES MRJMON; /*Tables will show percentages among marijuana
users*/
PRINT WSUM NSUM PERCENT SEPERCENT TOTAL SETOTAL / REPLACE
STYLE=NCHS;
OUTPUT WSUM PERCENT SEPERCENT TOTAL SETOTAL NSUM / REPLACE
FILENAME="OUT.SUDFILE_FREQ";
TITLE "FREQUENCY OF MARIJUANA USE BY PAST MONTH MARIJUANA USERS";
RUN;
```

Exhibit A.40 Stata Code (Frequency of Use; i.e., Number of Days Used Substance in the Past Month among Past Month Users)

```
use using ".\\dataname.dta", clear
/*Ensure all variables are lower case*/
rename *, lower

svyset verep [pw=analwt], strata(vestr) dof(750)
svy: proportion mrjmdays, subpop(mrjmon)
/*This code will produce output showing proportions for marijuana
use frequency in the past month, to get percentages, these proportions
would need to be multiplied by 100*/
```

Exhibit A.41 SAS (Frequency of Use; i.e., Number of Days Used Substance in the Past Month among Past Month Users)

```
TITLE "CALCULATE PERCENTAGES AND ASSOCIATED SEs";
PROC SURVEYFREQ DATA=DATANAME;
     WHERE MRJMON=1;
     CLUSTER VEREP;
     STRATA VESTR;
     WEIGHT ANALWT;
     TABLE MRJMDAYS;
RUN;
```

Exhibit A.42 R (Frequency of Use; i.e., Number of Days Used Substance in the Past Month among Past Month Users)

```
# FREQUENCY OF MARIJUANA USE BY PAST MONTH MARIJUANA USERS: 5 groups
including #non-users
# Sample total N of past month marijuana users
svyby( ~ one , ~ mrjmon, design=subset(design, mrjmon==1), unwtd.count
)
# weighted sample N in total
svytotal(~mrjmon, design=subset(design, mrjmon==1), na.rm = TRUE)

# Estimated N of past month marijuana users by 5 categories of
marijuana use #days
svytotal( ~ mrjmdays , design=subset(design, mrjmon==1), na.rm = TRUE)

# Estimated % and SE by 5 categories of marijuana use days
b=svymean( ~ mrjmdays, design=subset(design, mrjmon==1), na.rm = TRUE)
coef(b)*100 # percentage
SE(b)*100 # percentage SE
```

Testing between Overlapping Domains

In addition to testing between-year differences shown in Exhibits A.13 through A.16, Exhibits A.43 through A.46 demonstrate testing between two overlapping domains. Specifically, these exhibits show how to use a stacked dataset to test whether past month cigarette use among the full population aged 18 or older is different from cigarette use among people aged 18 or older who are employed full time.

This code will apply when one domain is completely contained in another or when there is only partial overlap. The example below uses two domains, where one domain is completely contained in the other (i.e., comparing full-time employed adults with all adults—the employed group is completely contained by the all adults group). Note that the correlations between the two estimates are accounted for in this test (i.e., correlation between past month cigarette use among people aged 18 or older and past month cigarette use among people aged 18 or older employed full time).

Exhibit A.43 SAS Code (Stacking a Dataset) and SUDAAN DESCRIPT Procedure (Test of Difference when Two Groups Overlap Using Stacked Data)

```
DATA STACKED;
SET DATANAME(IN=A) DATANAME(IN=B); /*reading in data twice*/
IF A THEN DO;
  INDIC=1;
  IF IRWRKSTAT18 IN (1,2,3,4) THEN EMPLOY=1;
  /* IRWRKSTAT18 is a four-level employment variable for adults,
  where level 1 is those employed full time, 2 is those employed
  part time, 3 are those unemployed, and 4 are all other adults.
  Respondents aged 12 to 17 are coded as level 99*/
  ELSE EMPLOY=0;
END;
ELSE IF B THEN DO;
  INDIC=2;
  IF IRWRKSTAT18=1 THEN EMPLOY=1;
  ELSE EMPLOY=0;
END;
RUN;

/*create an indicator variable for the stacked data, this will be
used in the diffvar statement in PROC DESCRIPT
When indic=1, employ=1 represents the full population
When indic=2, employ=1 represents those employed full time*/

PROC SORT DATA=STACKED;
BY VESTR VEREP;
RUN;

PROC DESCRIPT DATA=STACKED DDF=750 DESIGN=WR FILETYPE=SAS;
NEST VESTR VEREP;
WEIGHT ANALWT;
VAR CIGMON;
SUBGROUP INDIC;
LEVELS 2
DIFFVAR INDIC=(1 2); /*Since subsetting in the next line to
employ=1, this is testing all persons 18+ vs. employed persons
18+*/
SUBPOPN CATAG18=1 AND EMPLOY=1;
PRINT WSUM NSUM MEAN SEMEAN TOTAL SETOTAL T_MEAN P_MEAN /
  REPLACE STYLE=NCHS;
```

Exhibit A.43 SAS Code (Stacking a Dataset) and SUDAAN DESCRIPT Procedure (Test of Difference when Two Groups Overlap Using Stacked Data) (continued)

```
OUTPUT WSUM MEAN SEMEAN TOTAL SETOTAL NSUM T_MEAN P_MEAN /
   REPLACE
   NSUMFMT=F8.0 WSUMFMT=F12.0 MEANFMT=F15.10 SEMEANFMT=F15.10
   TOTALFMT=F12.0 SETOTALFMT=F12.0 FILENAME="OUT.SUDTESTS";
TITLE "TESTS OF DIFFERENCES BETWEEN ALL PERSONS 18 OR OLDER AND
EMPLOYED PERSONS 18 OR OLDER";
RUN;
```

Exhibit A.44 Stata Code (Test of Difference when Two Groups Overlap Using Stacked Data)

```
/*Creating the first dataset*/
/*Read in data */
use using ".\\dataname.dta", clear
/*Ensure all variables are lower case*/
rename *, lower

gen indic = 1
gen employ = 0
replace employ = 1 if inlist(irwrkstat18,1,2,3,4)
/*Save the dataset*/
save ".\\a44_a.dta" , replace /*Need to save dataset since Stata
can only work with one at a time*/

/*Creating the second dataset*/
/*Read in data a second time*/
use using ".\\dataname.dta", clear
/*Ensure all variables are lower case*/
rename *, lower
gen indic = 2
gen employ = 0
replace employ = 1 if inlist(irwrkstat18,1)
*Save the dataset
save ".\\a44_b.dta" , replace /*Need to save dataset since Stata
could only work with one at a time*/

/*Need to stack the dataset together */
use using ".\\a44_a.dta", clear
append using ".\\a44_b.dta"

/*Create the subpopulation variable*/
generate subpop = 1 if catag18 == 1 & employ == 1
svyset verep [pweight=analwt], strata(vestr) dof(750)
svy, subpop(subpop): mean cigmon, over(indic)
test [cigmon]1 = [cigmon]2
/*Since subsetting to employ=1, this is testing all persons 18+
vs. employed persons 18+ for past month cigarette use*/
/* employ is defined earlier in this exhibit and catag18=1 for
persons 18 or older and 0 otherwise  */
```

Exhibit A.45 SAS Code (Statistical Tests of Differences between Two Groups when the Two Groups Overlap)

```
TITLE "PERFORM TEST OF DIFFERENCE BTWEEN TWO OVERLAP GROUPS";
DATA STACKED;
SET DATANAME(IN=A) DATANAME(IN=B); /*reading in data twice*/
      IF A THEN DO;
      INDIC=1;
      IF IRWRKSTAT18 IN (1,2,3,4) THEN EMPLOY=1;
      /*IRWRKSTAT18 is a four-level employment variable for adults,
      where level 1 is those employed full time, 2 is those employed
      part time, 3 are those unemployed, and 4 are all other adults.
      Respondents aged 12 to 17 are coded as level 99*/
      ELSE EMPLOY=0;
      END;
      ELSE IF B THEN DO;
      INDIC=2;
      IF IRWRKSTAT18=1 THEN EMPLOY=1;
      ELSE EMPLOY=0;
END;

/*create an indicator variable for the stacked data, this will be used
in PROC SURVEYREG
When indic=1, employ=1 represents the full population
When indic=2, employ=1 represents those employed full time*/
RUN;

PROC SORT DATA=STACKED;
      BY VESTR VEREP;
RUN;

PROC SURVEYREG DATA=STACKED;
      WHERE CATAG18=1 & EMPLOY=1;
      CLUSTER VEREP;
      STRATA VESTR;
      WEIGHT ANALWT;
      CLASS INDIC;
      MODEL CIGMON=INDIC/NOINT VADJUST=NONE SOLUTION COVB;
      LSMEANS INDIC/DIFF;
RUN;
```

Exhibit A.46 R Code (Statistical Tests of Differences between Two Groups when the Two Groups Overlap)

```
# Significance testing of difference in cigarette user prevalence
between #full pop ages 18+ and full employed ages 18+ in year 2018
# Note that the two groups overlap and we only focus on year 2018

#Data management and create survey design
#1st data (indic=1) has employ=1 when IRWRKSTAT18 IN (1,2,3,4)
d18.1=DATANAME
d18.1$indic=1
d18.1$employ=ifelse(d18.1$irwrkstat18 %in% c(1,2,3,4), 1, 0)
#2nd data (indic=2) has employ=1 when IRWRKSTAT18 ==1
d18.2= DATANAME
d18.2$indic=2
d18.2$employ=ifelse(d18.2$irwrkstat18==1, 1, 0)
#append the two
d18.1.2=rbind(d18.1, d18.2)

# svydesign
design.A46 <-
  svydesign(
    id = ~ verep ,
    strata = ~ vestr ,
    data = d18.1.2 ,
    weights = ~ analwt ,
    nest = TRUE
  )

# Significance testing of difference in cigarette user prevalence
between
#full pop ages 18+ and full employed ages 18+ in year 2018
svyttest(cigmon~indic, subset(design.A46, catag18==1 & employ==1))
```

Testing Independence of Two Variables when One Variable Has Three or More Levels

When comparing population subgroups defined by three or more levels of a categorical variable, log-linear chi-square tests of independence of the subgroup and the prevalence variables are conducted first to control the error level for multiple comparisons (i.e., if the goal is to compare cigarette use among several levels of employment, first test whether cigarette use is associated with employment). Exhibits A.47 through A.50 show the code for calculating the Wald *F* test to determine whether cigarette use is associated with employment status. If Shah's Wald *F* test (transformed from the standard Wald chi-square) indicated overall significant differences, the significance of each particular pairwise comparison of interest can be tested using the SUDAAN procedure DESCRIPT (as shown in Exhibit A.43), Stata (Exhibit A.44), SAS (Exhibit A.45), or R (Exhibit A.46). The additional pairwise testing can determine which levels of employment status show significant differences in cigarette use compared with other levels of employment as shown using SUDAAN (Exhibit A.51), Stata (Exhibit A.52), SAS (Exhibit A.53), or R (Exhibit A.54).

Exhibit A.47 SUDAAN CROSSTAB Procedure (Test for Independence Based on a Log-Linear Model)

```
PROC CROSSTAB DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS DEFT4;
NEST VESTR VEREP;
WEIGHT ANALWT;
CLASS CIGMON;
SUBGROUP IRWRKSTAT18; /*four level employment status variable*/
LEVELS  4;
TABLES IRWRKSTAT18*CIGMON;
TEST LLCHISQ / WALDF;  /*log linear hypothesis test, wald F test
statistic, if test statistic is significant, then reject null
hypothesis of no interaction*/
SETENV DECWIDTH=4 COLWIDTH=15;
PRINT NSUM WSUM TOTPER ROWPER COLPER STESTVAL SPVAL SDF /
  REPLACE STYLE=NCHS;
OUTPUT STESTVAL SPVAL SDF / REPLACE FILENAME="TEST_CHI";
RUN;
```

Exhibit A.48 Stata Code (Test for Independence Based on a Log-Linear Model)

```
use using ".\\dataname.dta", clear
/*Ensure all variables are lower case*/
rename *, lower

/*Need to subset to just 4 levels of empstat4*/
generate subpop = 1 if inlist(irwrkstat18,1,2,3,4)
/*four level employment status variable*/

svyset verep [pw=analwt], strata(vestr) dof(750)
svy, subpop(subpop): tab cigmon irwrkstat18, llwald noadjust

/*This will give you both the adjusted and non-adjusted Wald F,
the non-adjusted test statistic will match SUDAAN*/
```

Exhibit A.49 SAS Code (Tests of the Independence of the Prevalence Variable and Subgroup Variable)

```
TITLE "PERFORM TEST OF INDEPENDENCE OF THE PREVALENCE VARIABLE AND
SUBGROUP VARIABLE";
PROC SURVEYFREQ DATA=DATANAME;
      WHERE IRWRKSTAT18 IN (1,2,3,4);
      CLUSTER VEREP;
      STRATA VESTR;
      WEIGHT ANALWT;
      TABLE IRWRKSTAT18*CIGMON / COL ROW CHISQ WLLCHISQ; /*option COL
      displays column percentages; option ROW displays row percentages;
      option CHISQ requests Rao-Scott chi-square test; option WLLCHISQ
      requests Wald log-linear chi-square test. */
      ODS OUTPUT WLLCHISQ=OUT.SAS_CHI;
RUN;
```

Exhibit A.50 R Code (Tests of the Independence of the Prevalence Variable and Subgroup Variable)

```
# Chi-square test of independence: evaluate the association between
two #categorical variables

#prepare data
design.A50 <-
  svydesign(
    id = ~ verep ,
    strata = ~ vestr ,
    data = DATANAME,
    weights = ~ analwt ,
    nest = TRUE
  )

design.A50=update(design.A50,
                  one = 1,
                  irwrkstat18=factor(irwrkstat18, levels=1:4, labels =
                c("full-time", "part-time","unemployed","all other
                adults"))
)

# sample N of cigarette use
## cigarettes user / non-user
svyby( ~ one, ~ cigmon , subset(design.A50, catag18==1), unwtd.count )
## N by employment status
svyby( ~ one, ~ irwrkstat18 , subset(design.A50, catag18==1),
unwtd.count )
## cigarettes user by employment status
svyby(~ one, ~irwrkstat18+cigmon, subset(design.A50, catag18==1),
unwtd.count)

# weighted sample N
## cigarettes user by employment status
svytable(~irwrkstat18+cigmon, subset(design.A50, catag18==1),
round=TRUE)
## % cigarettes user by employment status
svytable(~irwrkstat18+cigmon, subset(design.A50, catag18==1),) %>%
prop.table(1)

# wald F test for independence between cigarettes use and employment
status #with 4 levels
a=svyloglin(~irwrkstat18+cigmon, subset(design.A50, catag18==1) )
b=update(a,~.^2); regTermTest(b, ~irwrkstat18:cigmon)
```

Exhibit A.51 SUDAAN DESCRIPT Procedure (Pairwise Testing)

```
PROC DESCRIPT DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS;
NEST VESTR VEREP;
WEIGHT ANALWT;
VAR CIGMON;
SUBGROUP IRWRKSTAT18;
LEVELS 4;
PAIRWISE IRWRKSTAT18 / NAME="Tests of differences for all
levels";
PRINT WSUM NSUM MEAN SEMEAN TOTAL SETOTAL T_MEAN P_MEAN /
  REPLACE STYLE=NCHS;
OUTPUT WSUM MEAN SEMEAN TOTAL SETOTAL NSUM T_MEAN P_MEAN /
   REPLACE
   NSUMFMT=F8.0 WSUMFMT=F12.0 MEANFMT=F15.10 SEMEANFMT=F15.10
   TOTALFMT=F12.0 SETOTALFMT=F12.0 FILENAME="OUT.SUDTESTS";
TITLE "TESTS OF DIFFERENCES IN PAST MONTH CIGARETTE USE AMONG ALL
LEVELS OF EMPLOYMENT STATUS";
RUN;
```

Exhibit A.52 Stata Code (Pairwise Testing)

```
use using ".\\dataname.dta", clear
/*Ensure all variables are lower case*/
rename *, lower

/*Need to subset to just 4 levels of empstat4*/
generate subpop = 1 if inlist(irwrkstat18,1,2,3,4)
/*four level employment status variable*/

svyset verep [pw=analwt], strata(vestr) dof(750)

/*Estimated means of past month cigarette use by employment
status*/
svy: mean cigmon, over(irwrkstat18)
matrix Me = e(b)

local max=4 /*number of irwrkstat18 categories*/
matrix output = J(6,7,.) /*empty matrix to store results - the
number of rows should match the number of contrasts needed*/

local counter1 = `max' - 1
local counter2 = `max' - 1
local contrast = 0

forvalues i=1/`counter1' {
 local stop = `max' - `i' + 1
 forvalues j=1/`counter2' {
  local contrast = `contrast' + 1
  test [cigmon]`j' = [cigmon]`stop', nosvyadjust ///
   matvlc(mtest`contrast')
  matrix output[`contrast', 1] = `j'
  matrix output[`contrast', 2] = `stop'
```

Exhibit A.52 Stata Code (Pairwise Testing) (continued)

```
        matrix output[`contrast',7]=r(p)
        matrix output[`contrast',4]=sqrt((mtest`contrast'[1,1]))
        matrix output[`contrast',3]=Me[1,`j']-Me[1,`stop']
       }
       local counter2 = `counter2' - 1
      }
      /*Estimated Totals*/
      svy: total cigmon, over(irwrkstat18)

      matrix M = e(b) /*Store total estimates in matrix M*/
      local max=4 /*number of categories*/

      local counter1 = `max' - 1
      local counter2 = `max' - 1
      local contrast = 0

      forvalues i=1/`counter1' {
       local stop = `max' - `i' + 1
       forvalues j=1/`counter2' {
        local contrast = `contrast' + 1
        test [cigmon]`j' = [cigmon]`stop', nosvyadjust ///
         matvlc(test`contrast')
        matrix output[`contrast',6]=sqrt((test`contrast'[1,1]))
        matrix output[`contrast',5]=M[1,`j']-M[1,`stop']
       }
       local counter2 = `counter2' - 1
      }
      matrix colnames output = level1 level2 mean semean total_out ///
       setotal mean_pval
      matrix list output
```

Exhibit A.53 SAS Code (Pairwise Testing)

```
TITLE "PERFORM PAIRWISE TESTS FOR EACH SUBGROUP VARIABLE";
PROC SURVEYREG DATA=DATANAME;
      WHERE IRWRKSTAT18 IN (1,2,3,4);
      CLUSTER VEREP;
      STRATA VESTR;
      WEIGHT ANALWT;
      CLASS IRWRKSTAT18 ;
      MODEL CIGMON=IRWRKSTAT18/NOINT VADJUST=NONE SOLUTION COVB;
      LSMEANS IRWRKSTAT18/DIFF;
RUN;
```

Exhibit A.54 R Code (Pairwise Testing)

```
# Create study design using previously created design from
Exhibit A.50

design.A54=design.A50
design.A54 <-
  update(design.A54,
         employed1=ifelse(irwrkstat18=='full-time', 1, 0),
         employed2=ifelse(irwrkstat18=='part-time', 1, 0),
         employed3=ifelse(irwrkstat18=='unemployed', 1, 0),
         employed4=ifelse(irwrkstat18=='all other adults', 1, 0),
         group1=factor(ifelse(irwrkstat18 %in% c("full-time","part-
time"), 1, 0), levels=0:1, labels=c("No", "Yes")),
         group2=factor(ifelse(irwrkstat18 %in% c("full-
time","unemployed"), 1, 0), levels=0:1, labels=c("No", "Yes")),
         group3=factor(ifelse(irwrkstat18 %in% c("full-time","all
other adults"), 1, 0), levels=0:1, labels=c("No", "Yes")),
         group4=factor(ifelse(irwrkstat18 %in% c("part-
time","unemployed"), 1, 0), levels=0:1, labels=c("No", "Yes")),
         group5=factor(ifelse(irwrkstat18 %in% c("part-time","all
other adults"), 1, 0), levels=0:1, labels=c("No", "Yes")),
         group6=factor(ifelse(irwrkstat18 %in% c("unemployed","all
other adults"), 1, 0), levels=0:1, labels=c("No", "Yes"))
)

# sample N involved in each of comparisons for Tukey test
## comparison: full-time vs part-time
svyby( ~ one, ~ group1, design.A54, unwtd.count )
svyby( ~ one, ~ group2, design.A54, unwtd.count )
svyby( ~ one, ~ group3, design.A54, unwtd.count )
svyby( ~ one, ~ group4, design.A54, unwtd.count )
svyby( ~ one, ~ group5, design.A54, unwtd.count )
svyby( ~ one, ~ group6, design.A54, unwtd.count )

# weighted sample N involved in each of comparisons for Tukey test
## comparison: full-time vs part-time
svyby( ~ one, ~ group1, design.A54, svytotal )
svyby( ~ one, ~ group2, design.A54, svytotal )
svyby( ~ one, ~ group3, design.A54, svytotal )
svyby( ~ one, ~ group4, design.A54, svytotal )
svyby( ~ one, ~ group5, design.A54, svytotal )
svyby( ~ one, ~ group6, design.A54, svytotal )

# pairwise test
a=svyglm(cigmon ~irwrkstat18, subset(design.A54, catag18==1))
pw = summary(glht(a, mcp(irwrkstat18="Tukey")))
summary(pw, test = adjusted("none")) #Sudaan output
summary(pw, test = adjusted("bonf")) #bonferroni adjustment
```

Testing of Linear Trends

Users may also find it useful to test the linear trend for all data points, across all years of interest. The linear trend test can inform users about whether prevalence use has decreased, increased, or remained steady over the entire span of the years of interest. This type of test can be done using SUDAAN (as shown in Exhibits A.55 and A.59), Stata (as shown in Exhibits A.56 and A.60), SAS (as shown in Exhibits A.57 and A.61), or R (as shown in Exhibits A.58 and A.62). This linear trend test can be performed using a *t* test (Exhibits A.55 through A.58) or modeling (Exhibits A.59 through A.62), depending on the analysis.

Contrast Method

The *t* test method for testing linear trends is more simplistic and better suited for large-scale table production similar to that used in NSDUH's detailed tables if the primary purpose is to test whether any observed differences across years are significant without consideration of other covariates. This method is also consistent with the method used in the detailed tables to test means between years and between demographic levels as shown in Exhibits A.13 through A.16. In SUDAAN, the *t* test method would be implemented using the CONTRAST statement in the DESCRIPT procedure as shown in Exhibit A.55. The corresponding Stata code using test statements is shown in Exhibit A.56, the SAS code is shown in Exhibit A.57, and the corresponding R code is shown in Exhibit A.58.

All approaches are based on orthogonal polynomial coefficients. The code in Exhibits A.55 and A.56 includes two placeholders that need to be specified by the user. For each year of data that a user wants to include in the test, an additional contrast is required to account for that year. Certain variables are available for only a subgroup of NSDUH years, and sometimes the analysis of interest involves only a subgroup of years. For this reason, Table A.4 is provided to help users specify the needed information for linear trend tests involving from 3 to 17 years of data. Recall that 2 years of data would be the same as the comparison shown in Exhibits A.13 through A.16. Thus, Exhibits A.55 through A.58 are for tests across a combination of 3 or more years of data.

Table A.4 Contrast Statements for Exhibits A.55 and A.56

Number of Years (X)	Contrast Statement (Y)
17	(-8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8)
16	(-15 -13 -11 -9 -7 -5 -3 -1 1 3 5 7 9 11 13 15)
15	(-7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7)
14	(-13 -11 -9 -7 -5 -3 -1 1 3 5 7 9 11 13)
13	(-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6)
12	(-11 -9 -7 -5 -3 -1 1 3 5 7 9 11)
11	(-5 -4 -3 -2 -1 0 1 2 3 4 5)
10	(-9 -7 -5 -3 -1 1 3 5 7 9)
9	(-4 -3 -2 -1 0 1 2 3 4)
8	(-7 -5 -3 -1 1 3 5 7)
7	(-3 -2 -1 0 1 2 3)
6	(-5 -3 -1 1 3 5)
5	(-2 -1 0 1 2)
4	(-3 -1 1 3)
3	(-1 0 1)

NOTE: Replace the placeholders (X) and (Y) in Exhibits A.55 and A.56 per the information in this table. Replace (X) with the numbers of years included in the linear trend test and (Y) with the corresponding contrast statement.

Exhibit A.55 SUDAAN DESCRIPT Procedure (Test of Linear Trends with DESCRIPT)

```
PROC DESCRIPT DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS;
NEST VESTR VEREP;
WEIGHT ANALWT;
VAR ALCMON;
SUBGROUP YEAR IRSEX;
LEVELS X 2; /*define X as the # of years*/
TABLES IRSEX;
CONTRAST YEAR = Y / NAME="LINEAR TREND TEST"; /*define Y as the
coefficients according to the number of years see Table A.4*/
PRINT WSUM NSUM MEAN SEMEAN TOTAL SETOTAL T_MEAN P_MEAN /
  REPLACE STYLE=NCHS;
OUTPUT WSUM MEAN SEMEAN TOTAL SETOTAL NSUM T_MEAN P_MEAN /
    REPLACE
    NSUMFMT=F8.0 WSUMFMT=F12.0 MEANFMT=F15.10 SEMEANFMT=F15.10
    TOTALFMT=F12.0 SETOTALFMT=F12.0 FILENAME="OUT.SUDTESTS";
TITLE "TEST OF LINEAR TREND IN PAST MONTH ALCOHOL USE BY GENDER";
RUN;
```

Exhibit A.56 Stata Code (Test of Linear Trends with TEST Statements)

```
use ".\\dataname.dta", clear
/*Ensure all variables are lower case*/
rename *, lower

svyset verep [pw=analwt], strata(vestr) dof(750)

svy: mean alcmon, over(year irsex)
matrix Me = e(b)

matrix coeff = (Y) /*define Y as the coefficients according to
the # of years see Table A.4, note the coefficients have to be
separate by commas*/
local max=X*2 /*total number of subpops - # of years(X)*# levels
of irsex(2)*/
local counter1 = 2 /*number of categories, i.e. number of levels
of irsex*/

generate pmean=.
generate mean=.
generate semean=.
forvalues i=1/`counter1' {  /*number of categories, i.e. number
of levels of irsex*/
 local stop = `max' / `counter1'
 local test
 local mean
 forvalues j=1/`stop' { /*stop should be equal to the # of
coefficients defined in coeff*/
  local sub = `i' + `counter1'*(`j'-1)
  local co = coeff[1,`j']
  local test = "`test' (`co')*[alcmon]_subpop_`sub'"
  local mean = "`mean' `co'*Me[1,`sub']"
  if (`j' < `stop') {
    local test = "`test' + "
    local mean = "`mean' + "
     }
 }
 test`test' = 0, nosvyadjust matvlc(mtest`counter')
 replace pmean=r(p) if irsex==`i'
 replace semean = sqrt((mtest`counter'[1,1])) if irsex==`i'
 replace mean = `mean' if irsex==`i'
}

/*Estimated Totals*/

svy: total alcmon, over(year irsex)
matrix M = e(b)

generate total=.
generate setotal=.
local counter=1
```

Exhibit A.56 Stata Code (Test of Linear Trends with TEST Statements) (continued)

```
forvalues i=1/`counter1' {  /*number of categories, i.e. number
of levels of irsex*/
 local stop = `max' / `counter1'
 local test
 local total
 forvalues j=1/`stop' { /*stop should be equal to the # of
coefficients defined in coeff*/
  local sub = `i' + `counter1'*(`j'-1)
  local co = coeff[1,`j']
  local test = "`test' (`co')*[alcmon]_subpop_`sub'"
  local total = "`total' `co'*M[1,`sub']"
      if (`j' < `stop') {
       local test = "`test' + "
      local total = "`total' + "
      }
 }
 test 'test` = 0, nosvyadjust matvlc(test`counter')
 replace setotal= sqrt((test`counter'[1,1])) if irsex==`i'
 replace total=`total' if irsex==`i' /*Calculating the difference
between the totals of the subpopulation*/
local counter = `counter'+1
}

/*Keeping variables that matches SUDAAN*/
keep irsex mean semean total setotal pmean
duplicates drop irsex mean semean total setotal pmean, force
/*keep one record per contrast*/

drop if total ==. /* drop the rows where there is no information
*/
format pmean %-15.10f
format total %-12.0fc
format setotal %-12.0fc

/* Output the dataset*/
list irsex mean semean total setotal pmean
```

Exhibit A.57 SAS Code (Linear Trend Test of Significance across Years Using Test Statements)

```
Note: The example input dataset includes 2016-2018 NSDUH data. The
DIFF=CONTROL('3') requests tests testing the difference between each
level against the control group (i.e. 3). Year must be defined as
values of 1 to x, where x is the total number of years.

TITLE "LINEAR TREND TEST OF SIGNIFICANCE ACROSS YEARS USING TEST
STATEMENT";
PROC SURVEYREG DATA=DATANAME; *Using 3 years of data 2016-2018;
     DOMAIN IRSEX;
     CLUSTER VEREP;
     STRATA VESTR;
     WEIGHT ANALWT;
     CLASS YEAR ;
     MODEL ALCMON=YEAR/NOINT VADJUST=NONE SOLUTION COVB;
     LSMEANS YEAR/DIFF=CONTROL('3');
     CONTRAST 'TEST' YEAR -1 0 1;/*test the customized hypothesis*/
RUN;
```

Exhibit A.58 R Code (Linear Trend Test of Significance across Years Using Test Statements)

```
# trend for 3 years: total, among males and females
# note that we use 3-year data: years 2016-2018
# create survey design
DATANAME$yearind<-ifelse(DATANAME$year==3,1,2)
design.A58 <-
  svydesign(
    id = ~ verep ,
    strata = ~ vestr ,
    data = DATANAME ,
    weights = ~ analwt ,
    nest = TRUE
  )

#add new columns
design.A58 <-
  update(design.A58,
         one = 1 ,
         year =
            factor(
              year ,
              levels = 1:3 ,
              labels = c( "2016", "2017" , "2018" ) ) ,
         irsex =
            factor(
              irsex ,
              levels = 1:2 ,
              labels = c( "male" , "female")),
              yearcombined=ifelse(year %in% c('2017', '2018'), 1, 0),
              year0=ifelse(year=='2016', 1, 0),
              year1=ifelse(year=='2017', 1, 0),
```

Exhibit A.58 R Code (Linear Trend Test of Significance across Years Using Test Statements) (continued)

```
                year2=ifelse(year=='2018', 1, 0),
                sexmale=ifelse(irsex=='male', 1, 0),
                sexfemale=ifelse(irsex=='female', 1, 0)
  )

# crosstab results of N and % of alcohol drinkers by year
svytable(~year+alcmon, design=design.A58)
svytable(~year+alcmon, Ntotal=100, design=design.A58) %>%
prop.table(1)
svytable(~year+alcmon, Ntotal=100, design=subset(design.A58,
irsex=='male')) %>% prop.table(1)
svytable(~year+alcmon, Ntotal=100, design=subset(design.A58,
irsex=='female')) %>% prop.table(1)

# contrast sig test: 2016 versus 2018
## 1) total pop
##coefficient of 2018 is contrast with 2016
overall <- svyglm(alcmon~year, family=quasibinomial,
design=design.A58); summary(overall)
summary(overall)$coefficients[3,4] # p value
summary(overall)$coefficients[3,3] # t-stat for sig testing
# 2) male
##coefficient of 2018 is contrast with 2016
male <- svyglm(alcmon~year, family=quasibinomial, subset(design.A58,
irsex=="male")); summary(male)
summary(male)$coefficients[3,4] # p value
summary(male)$coefficients[3,3] # t-stat for sig testing
# 3) female
##coefficient of 2018 is contrast with 2016
female <- svyglm(alcmon~year, family=quasibinomial, subset(design.A58,
irsex=="female")); summary(female)
summary(female)$coefficients[3,4] # p value
summary(female)$coefficients[3,3] # t-stat for sig testing
```

Modeling Method

The model-based method is more complex and flexible. This method, which was used in the analyses for the 2014 redesign impact assessment report (RIAR) (CBHSQ, 2015c) and the 2015 RIAR (CBHSQ, 2017c), can measure a change in a variable over time while controlling for covariates. The modeling method can be used for more specific tests, such as controlling for the linear year trend across years to determine a break in trend for the current year. In the examples below, the variable YEAR should be defined as a continuous variable (i.e., 1 to X, with X being the number of years included in the test), and the variable YEARIND should be defined as a categorical variable (i.e., 1 if in the current year of interest or 2 if not in the current year of interest). The SUDAAN modeling method shown in Exhibit A.59 uses the procedure RLOGIST for logistic regression, and the Stata modeling example shown in Exhibit A.60 uses the svy: logit command for logistic regression. The SAS modeling example shown in Exhibit A.61 uses PROC SURVEYLOGISTIC, and the R modeling example shown in Exhibit A.62 uses svytable.

The models shown below were used to determine change, but a simpler model could be run to test overall trends across years similar to Exhibits A.59 through A.62 by removing the YEARIND variable from the code below. Note that the simplified modeling method may give a slightly different result than the DESCRIPT method under similar settings.

Exhibit A.59 SUDAAN RLOGIST Procedure (Modeling Test of Linear Trends)

```
Note: The example input dataset includes 2002-2018 NSDUH data, so YEAR
= 1 to 17 and YEARIND = 1 if in 2018 and YEARIND = 2 if not in 2018.

/*Overall model, no subpopulations*/

PROC RLOGIST DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS;
NEST VESTR VEREP;
WEIGHT ANALWT;
REFLEVEL YEARIND=2; /*Not in Current Year is Reference Level*/
SUBGROUP YEARIND;
LEVELS 2;
MODEL ALCMON=YEARIND YEAR; /*Model controlling for linear trend of
year to determine change in the current year*/
SETENV DECWIDTH=6 COLWIDTH=18;
PRINT BETA="BETA" SEBETA="STDERR" DEFT="DESIGN EFFECT"
T_BETA="T:BETA=0" P_BETA="P-VALUE"/ RISK=ALL TESTS=DEFAULT
T_BETAFMT=F8.2 WALDCHIFMT=f6.2 ORFMT=f10.2 LOWORFMT=f10.2
UPORFMT=f10.2 DFFMT=f7.0;
OUTPUT BETA SEBETA T_BETA P_BETA / REPLACE
FILENAME="OUT.MODEL_OUTPUT";
TITLE "MAIN MODEL OF ALCMON - OVERALL";
RUN;

/*model below is subset for Gender where IRSEX=1 is Males. Similar
model can be run for IRSEX=2 for Females*/

PROC RLOGIST DATA=DATANAME DDF=750 DESIGN=WR FILETYPE=SAS;
NEST VESTR VEREP;
WEIGHT ANALWT;
REFLEVEL YEARIND=2; /*Not in Current Year is Reference Level*/
SUBGROUP YEARIND;
LEVELS 2;
MODEL ALCMON=YEARIND YEAR; /*Model controlling for linear trend of
year to determine change in the current year*/
SUBPOPN IRSEX=1; /*Subset for Males*/
SETENV DECWIDTH=6 COLWIDTH=18;
PRINT BETA="BETA" SEBETA="STDERR" DEFT="DESIGN EFFECT"
T BETA="T:BETA=0" P BETA="P-VALUE"/ RISK=ALL TESTS=DEFAULT
T BETAFMT=F8.2 WALDCHIFMT=f6.2 ORFMT=f10.2 LOWORFMT=f10.2
UPORFMT=f10.2 DFFMT=f7.0;
OUTPUT BETA SEBETA T BETA P BETA / REPLACE
FILENAME="OUT.MODEL OUTPUT";
TITLE "MAIN MODEL OF ALCMON - MALES";
RUN;
```

Exhibit A.60 Stata Code (Modeling Test of Linear Trends)

```
Note: The example input dataset includes 2002-2018 NSDUH data, so YEAR
= 1 to 17 and YEARIND = 1 if in 2018 and YEARIND = 2 if not in 2018.

use using ".\\dataname.dta", clear

svyset verep [pw=analwt], strata(vestr) dof(750)

/*Overall model controlling for linear trend of year to determine
change in the current year.*/
svy: logit alcmon ib2.yearind year

/*Create a subsetting variable, irsex_1 that will be 1 for males
(IRSEX=1) and zero otherwise. A similar variable can be created to
subset for females (IRSEX=2)*/
generate irsex_1 = 0
replace irsex_1 = 1 if irsex == 1

/*Model subsetting by gender and controlling for linear trend of year
to determine change in the current year. A similar model can be run
for females(IRSEX=2).*/
svy, subpop (irsex_1): logit alcmon ib2.yearind year
```

Exhibit A.61 SAS Code (Linear Trend Test of Significance across Years Using Modeling)

```
Note: The example input dataset includes 2016-2018 NSDUH data, so YEAR
= 1 to 3 and YEARIND = 1 if in 2018 and YEARIND = 2 if not in 2018.

TITLE "LINEAR TREND TEST OF SIGNIFICANCE ACROSS YEARS USING MODELING";
DATA ALLYEAR;
SET DATANAME;
     IF YEAR=3 THEN YEARIND=1;
     ELSE YEARIND=2;
RUN;

/*OVERALL*/
PROC SURVEYLOGISTIC DATA=ALLYEAR; *Using 3 years of data 2016-2018;
     CLUSTER VEREP;
     STRATA VESTR;
     WEIGHT ANALWT;
     CLASS YEARIND /PARAM=REF;
     MODEL ALCMON(EVENT='1')=YEARIND YEAR/COVB;
     CONTRAST 'OVERALL MODEL' INTERCEPT 1, YEARIND 1, YEAR 1;
     CONTRAST 'OVERALL MODEL MINUS INTERCEPT' YEARIND 1, YEAR 1;
     CONTRAST 'YEARIND' YEARIND 1;
     CONTRAST 'YEAR' YEAR 1;
RUN;

/*MALE*/
PROC SURVEYLOGISTIC DATA=ALLYEAR;
     WHERE IRSEX=1; *Subset for males;
     CLUSTER VEREP;
```

Exhibit A.61 SAS Code (Linear Trend Test of Significance across Years Using Modeling) (continued)

```
      STRATA VESTR;
      WEIGHT ANALWT;
      CLASS YEARIND /PARAM=REF;
      MODEL ALCMON(EVENT='1')=YEARIND YEAR/COVB;
      CONTRAST 'OVERALL MODEL' INTERCEPT 1, YEARIND 1, YEAR 1;
      CONTRAST 'OVERALL MODEL MINUS INTERCEPT' YEARIND 1, YEAR 1;
      CONTRAST 'YEARIND' YEARIND 1;
      CONTRAST 'YEAR' YEAR 1;
RUN;
```

Exhibit A.62 R Code (Linear Trend Test of Significance across Years Using Modeling)

```
# Note that we use 3-year data: from 2016 to 2018
# The example input dataset includes 2016-2018 NSDUH data, so YEAR = 1
# to 3 and YEARIND = 1 if in 2018 and YEARIND = 2 if not in 2018.

design.A62 <-
  svydesign(
    id = ~ verep ,
    strata = ~ vestr ,
    data = DATANAME ,
    weights = ~ analwt ,
    nest = TRUE)

#add new columns
design.A62 <-
  update(design.A62,
         one = 1 ,
         irsex =
           factor(
           irsex ,
           levels = 1:2 ,
           labels = c( "male" , "female")),
         yearfactor =
           factor(
             year ,
             levels = 1: 3 ,
             labels = c("2016" , "2017", "2018" ) ),
         yearindicator=
           factor(
             ifelse(yearind==2, 0, 1), levels=0:1, labels=c("No",
"Yes"))
)

# N and % of alcohol drinkers by year
svytable(~yearfactor+alcmon, design=design.A62)
svytable(~yearfactor+alcmon, Ntotal=100, design=design.A62) %>%
prop.table(1)
svytable(~yearfactor+alcmon, Ntotal=100, design=subset(design.A62,
irsex=='male')) %>% prop.table(1)
```

Exhibit A.62 R Code (Linear Trend Test of Significance across Years Using Modeling) (continued)

```
svytable(~yearfactor+alcmon, Ntotal=100, design=subset(design.A62,
irsex=='female')) %>% prop.table(1)

# 1) total population
##parameter estimates
overall <- svyglm(alcmon~factor(yearindicator)+year,
family=quasibinomial, design=design.A62);summary(overall)
exp(coef(overall)) %>% round(2) #Odds Ratio
exp(confint(overall)) %>% round(2) #Confidence interval

# 2) male population
##parameter estimates
male <- svyglm(alcmon~factor(yearindicator)+year,
family=quasibinomial, design=subset(design.A36, irsex=="male"));
summary(male)
exp(coef(male)) %>% round(2) #Odds Ratio
exp(confint(male)) %>% round(2) #Confidence interval

# contrast wald test for predictor significance
# 1) total population
##wald significance test of the predictor 'yearindicator'
regTermTest(overall,'factor(yearindicator)')
##wald significance test of the predictor 'year'
regTermTest(overall,'year')
# 2) male population
##wald significance test of the predictor 'yearindicator'
regTermTest(male,'factor(yearindicator)')
##wald significance test of the predictor 'year'
regTermTest(male,'year')
```

SAMHSA
Substance Abuse and Mental Health
Services Administration

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

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