2009 NATIONAL SURVEY ON DRUG USE AND HEALTH

STATISTICAL INFERENCE REPORT

Prepared for the 2009 Methodological Resource Book

Contract No. 283-2004-00022
RTI Project No. 0209009.581.002
Deliverable No. 39

Authors:

Jeremy Aldworth
James R. Chromy
Teresa R. Davis
Misty S. Foster
David Heller
Erica Hirsch
Phillip Kott
Lisa E. Packer
Kathryn Spagnola

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

December 2010

2009 NATIONAL SURVEY ON DRUG USE AND HEALTH

STATISTICAL INFERENCE REPORT

Prepared for the 2009 Methodological Resource Book

Contract No. 283-2004-00022
RTI Project No. 0209009.581.002
Deliverable No. 39

Authors:

Jeremy Aldworth
James R. Chromy
Teresa R. Davis
Misty S. Foster
David Heller
Erica Hirsch
Phillip Kott
Lisa E. Packer
Kathryn Spagnola

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

December 2010

Acknowledgments

This publication was developed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Office of Applied Studies (OAS), by RTI International (a trade name of Research Triangle Institute), Research Triangle Park, North Carolina, under Contract No. 283-2004-00022. Significant contributors at RTI include Jeremy Aldworth, James R. Chromy, Teresa R. Davis, Misty S. Foster, David Heller, Erica Hirsch, Phillip Kott, Lisa E. Packer, and Kathryn Spagnola. Claudia Clark copyedited the report, and Thomas G. Virag is the RTI Project Director.

DISCLAIMER

SAMHSA provides links to other Internet sites as a service to its users and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.

Table of Contents

Section

1. Introduction 1

2. Background 3

3. Prevalence Rates 5
   3.1 Mental Illness 5
   3.2 Adult Major Depressive Episode (MDE) 7

4. Missingness 9

5. Sampling Error 11

6. Degrees of Freedom 15

7. Statistical Significance of Differences 17

8. Confidence Intervals 19

9. Incidence Estimates 21

10. Suppression of Estimates with Low Precision 25

References 29

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 2009 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 2009. Measurements for this target population were the responses to the survey questions provided by persons participating in the 2009 survey.

Statistical inferences concerning characteristics of interest for this population and various subpopulations are presented in the form of estimates derived from the sample data collected. Examples of the inferences made from the 2009 NSDUH data are presented in the 2009 detailed tables (Office of Applied Studies [OAS], 2010a) and the 2009 summary of national findings report (OAS, 2010d; 2010e) and include estimates of the number of persons who were substance users during the past month, past year, and their lifetime, as well as the associated percentages (prevalence rates) of substance use for these reference periods. Inferences also were made for such categories as substance initiation; risk and protective factors; substance dependence, dependence or abuse, and treatment. Estimates of measures related to mental health problems are presented in the 2009 mental health detailed tables (OAS, 2010b) and the 2009 mental health findings report (OAS, 2010c).

The focus of this report is to describe the statistical inference procedures used to produce design-based estimates as presented in the 2009 detailed tables, 2009 mental health detailed tables, 2009 national findings report, and 2009 mental health findings report.[2] The statistical procedures and information found in this report can also be generally applied to analyses based on the public use file. This report is organized as follows: Section 2 provides background information concerning the 2009 NSDUH; Section 3 discusses the prevalence rates and how they were calculated; Section 4 briefly discusses how missing item responses of variables that are not imputed may lead to biased estimates; Section 5 discusses sampling errors and how they were calculated; Section 6 describes the degrees of freedom that were used when comparing estimates; and Section 7 discusses how the statistical significance of differences between estimates was determined. Section 8 discusses confidence interval estimation, and Section 9 describes how past year incidence of drug use was computed. Finally, Section 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 using SAS® and SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2008).

---

[1] Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[2] Inconsistencies may be found by users of the 2009 public use file in the variable names referenced in this report, the information presented in Table 1, and other specific numbers presented in this report (i.e., degrees of freedom). The specific information referenced in this report is based on the restricted-use dataset that was used to create the 2009 detailed tables (OAS, 2010a), 2009 mental health detailed tables (OAS, 2010b), 2009 national findings report (OAS, 2010d; 2010e), and 2009 mental health findings report (OAS, 2010c).

2. Background

The 2009 NSDUH is the fifth survey in a coordinated 5-year sample design providing estimates for all 50 States plus the District of Columbia for the years 2005 through 2009. The survey is conducted using computer-assisted interviewing methods for the screening and interviewing of selected respondents. The respondent universe is the civilian, noninstitutionalized population aged 12 years old or older residing within the United States and the District of Columbia. Persons excluded from the universe include active-duty military personnel, persons with no fixed household address (e.g., homeless and/or transient persons not in shelters), and residents of institutional group quarters, such as jails and hospitals.

Although there is no planned overlap with the 1999–2004 samples, a coordinated design for 2005 through 2009 facilitates 50 percent overlap in second-stage units (area segments) within each successive 2-year period from 2005 through 2009. Because the 2005–2009 design enables estimates to be developed by State in all 50 States plus the District of Columbia, States may be viewed as the first level of stratification as well as a reporting variable.

For the 50-State design, 8 States were designated as large sample States (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas). In 2009, sample sizes in these States ranged from 3,557 to 3,707 respondents. For the remaining 42 States and the District of Columbia, sample sizes ranged from 886 to 984. State estimates combining multiple years of data and using either small area estimation (SAE)[3] or direct estimation have been tabulated.

States were first stratified into a total of 900 State sampling (SS) regions (48 regions in each large sample State and 12 regions in each small sample State). These regions were contiguous geographic areas designed to yield on average the same number of interviews.[4] Unlike the 1999–2004 surveys in which the first-stage sampling units were clusters of census blocks called area segments, the first stage of selection for the 2005–2009 surveys was census tracts.[5] This stage was included to contain sample segments within a single census tract to the extent possible.[6]

A total of 48 census tracts per SS region were selected, and within these sampled census tracts, adjacent census blocks were combined to form the second-stage sampling units or area segments. Eight sample segments per SS region were fielded during the 2009 survey year. These sampled segments were allocated equally into four separate samples, one for each 3-month

---

[3] SAE is a hierarchical Bayes modeling technique used to produce State-level estimates for a selected number of measures. For more details, see the *State Estimates of Substance Use from the 2007–2008 National Surveys on Drug Use and Health* (Hughes, Muhuri, Sathe, & Spagnola, 2010).

[4] Areas were defined using 2000 census geography. Dwelling units (DUs) and population counts were obtained from the 2000 census data supplemented with revised population counts from Claritas. Claritas, Inc., is a market research firm headquartered in San Diego, California.

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

[6] Some census tracts had to be aggregated in order to meet the minimum DU requirement of 150 DUs in urban areas and 100 DUs in rural areas.

period (calendar quarter) during the year, so that the survey was essentially continuous in the field.

The overall design remained the same beginning with the 2002 NSDUH and continuing through the 2009 NSDUH. Survey respondents were given a $30 incentive payment for participation, which increased response rates, thereby requiring fewer selected households than in previous surveys. Also, a pair-sampling strategy was implemented that increased the number of pairs selected in dwelling units (DUs) with older persons on the roster (Chromy & Penne, 2002).

As was done in the 2008 NSDUH, a Mental Health Surveillance Study (MHSS) was embedded in the 2009 NSDUH. Each respondent in a subsample of about 500 adults (in 2008, the subsample was about 1,500) who had completed the NSDUH interview was administered the Structured Clinical Interview for DSM-IV-TR Axis I Disorders Non-Patient Edition (SCID) (First, Spitzer, Gibbon, & Williams, 2002). SCID was adapted for this study and was administered via paper and pencil over the telephone approximately 2 to 4 weeks after the NSDUH interview. In 2008, a split-sample MHSS was conducted to develop models using the SCID data that would use the K6 nonspecific psychological distress scale and two competing functional impairment scales in order to generate prevalence estimates of serious mental illness (SMI) among adults aged 18 or older for the entire sample. Based on the results from the 2008 MHSS, a modified World Health Organization Disability Assessment Scale (WHODAS) (Rehm et al., 1999) was adopted for the 2009 survey. The purpose of the 2009 MHSS was to monitor the efficacy of the selected screening measure. For more information about the MHSS sample design, see the Sample Design Report in the *2009 NSDUH Methodological Resource Book* (Morton, Martin, Chromy, Foster, & Hirsch, 2010).

The final respondent sample of 68,700 persons for the 2009 NSDUH provides a sufficient sample to create domain estimates for a broad range of ages and other demographic categories. Individual observations are weighted in a manner such that the weighted sample is representative of the civilian, noninstitutionalized population aged 12 or older for both the general U.S. population as well as for each of the individual States. For more information on the sampling weight calibration in the 2009 NSDUH, see the Person-Level Sampling Weight Calibration report in the *2009 NSDUH Methodological Resource Book* (Chen et al., 2011).

3. Prevalence Rates

The national prevalence rates were computed using a multiprocedure package called SUDAAN. The final, nonresponse-adjusted, and poststratified analysis weights were used in SUDAAN to compute unbiased design-based drug use estimates. Appendix A contains an example that demonstrates how to compute the prevalence rates using SUDAAN procedures as defined below.

Prevalence rates are the proportions of the population who exhibit characteristics of interest (such as substance use). Let $\hat{p}_d$ represent the prevalence rate 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 $\hat{Y}_d$ = estimated number of persons exhibiting the characteristic of interest in domain *d,* and $\hat{N}_d$ = estimated population total for domain *d*.

$\hat{N}_d$ is estimated as $\sum w_i \delta_i$, where w_i represents the analysis weight and δ_i represents an indicator variable, which is defined as

$$\delta_i\,(d) = \begin{cases} 1 \text{ if the i}^{\text{th}} \text{ sample unit is in subgroup } d, \\ 0 \text{ otherwise.} \end{cases}$$

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 were produced from annual averages based on combined data from 2 or more survey years, and they are clearly labeled in the detailed tables. The data were combined for the 2006–2007, 2008–2009, or 2006–2009 surveys to obtain annual averages, and then the prevalence rates were computed in SUDAAN as described above. The annual averages were derived by concatenating the data for the respective years and then dividing the analysis weights by a factor that varied depending on the number of years of concatenated data. The weight was divided by a factor of 2 for 2 years of concatenated data and a factor of 4 for 4 years of concatenated data.

3.1 Mental Illness

To address the Substance Abuse and Mental Health Services Administration's need for estimates of SMI and any mental illness, the adult mental health module of the 2009 NSDUH questionnaire included the K6 nonspecific psychological distress scale and the WHODAS functional impairment scale that can be included in models estimating SMI and other mental illness measures. For the clinical subsample, functional impairment ratings were assigned by

clinical interviewers using the Global Assessment of Functioning (GAF) scale.[7] The GAF score and the presence of any of the mental disorders assessed in the SCID were used to determine the SMI "gold standard" and any mental illness status. Within the subsample, statistical models were developed between "gold standard" mental illness measures and those based on the K6 and WHODAS questions. Model-based estimates of SMI and any mental illness were then extrapolated to the full adult NSDUH data to produce national estimates for the entire adult civilian, noninstitutionalized population. For more information on the MHSS design and analysis, see Appendix B in the 2009 mental health findings report (OAS, 2010c).

The same statistical model and parameter estimates (WHODAS version) that were used to produce estimates of SMI for 2008 were used for the 2009 NSDUH data. The model (WHODAS version) used in both 2008 and 2009 to estimate SMI is described as follows: With SMI status based on having a SCID diagnosis plus a GAF score less than or equal to 50, the response variable Y was defined so that $Y = 1$ when an SMI diagnosis is positive; otherwise, $Y = 0$. If $\mathbf{X}$ is a vector of explanatory variables, then the response probability $\pi = \Pr(Y = 1 \mid \mathbf{X})$ can be estimated using the following logistic regression models for the WHODAS, respectively:

$$\text{logit}(\pi_w) \equiv \log[\pi_w / (1 - \pi_w)] = -4.7500 + 0.2098 X_k + 0.3839 X_w$$

where the X_k and X_w terms refer to the K6 and WHODAS terms, respectively, and are defined as follows:

- X_k = *Alternative Past Year K6 Score*: Past year K6 scores less than 8 recoded as 0; past year K6 score 8 to 24 recoded as 1 to 17.
- X_w = *Alternative WHODAS Score*: WHODAS item scores less than 2 recoded as 0; WHODAS item scores 2 to 3 recoded as 1, then summed for a score ranging from 0 to 8.

The model was used to calculate a predicted probability of SMI. Cut points for these predicted probabilities were then determined to create 0/1 SMI and mental illness variables. Receiver operating characteristic (ROC) analyses were used to determine the cut point that resulted in the weighted number of false positive and false negative counts being (approximately) equal and thus ensuring unbiased estimates. For this model, the optimal cut point for SMI was determined to be 0.26971946. See Aldworth et al. (2009) for further details.

The calculation of the standard error of the predicted SMI variable was conditional on the cut point (i.e., this fixes the cut point). Thus, the SMI and other mental illness estimates and standard errors were calculated similarly to other 0/1 variables (for more details on calculating standard errors, see Section 5). Note that this method of calculating the standard error will likely

[7] The GAF is a numeric scale (0 through 100) used to subjectively rate the social, occupational, and psychological functioning of adults and is described in the DSM-IV-TR (American Psychiatric Association, 2008, p. 32). Lower scores represent higher levels of functional impairment. Descriptions of impairment are provided at 10-point intervals such as 1 to 10, 11 to 20, and so on up to 91 to 100. For example, a GAF score between 51 and 60 is described as having moderate symptoms of impairment, while a score higher than 60 represents several categories of impairment ranging from none to slight, and a score lower than 51 represents several categories ranging from serious to extreme.

result in an underestimate because it does not take into account the variability of the cut point. In addition, the calculation of the standard error does not take into account the variability of the parameter estimates of the model used for predicting SMI due to the sampling variation of the MHSS subsample (i.e., a different subsample would likely have resulted in a model with slightly different parameter estimates). Investigations are currently being conducted to address the likely underestimation of the standard error of the predicted SMI and other mental illness variables.

When computing 2009 prevalence rates for SMI or any other mental illness category (any, moderate, serious or moderate, or mild mental illness), the standard analysis weight (ANALWT) should be used. However, for 2008 prevalence rates, ANALWT should only be used for SMI, and a split-sample weight (MHSAMPWT) needs to be used for the other mental illness categories since it was recommended that these 2008 estimates only be based on the WHODAS half sample. See B.4.3 in Appendix B of the 2009 mental health findings report (OAS, 2010c) for more details.

3.2 Adult Major Depressive Episode (MDE)

The 2009 adult major depressive episode (MDE) estimates are based on all adults aged 18 or older. This differs from 2008 where adult MDE estimates were based only on the half sample of adult respondents who received the WHODAS questions in the preceding mental health questionnaire module, because of apparent reporting differences (context effects) in the other half sample of adult respondents who received the Sheehan Disability Scale (SDS) questions in 2008 (Dean & LeBaron, 2009). However, an adjusted version of the MDE measure was subsequently created for the SDS half sample that was comparable with the MDE values from the WHODAS half sample, and this allowed for the entire 2008 sample to be used for MDE-related estimation. Due to the timing of the creation of these adjusted variables, this adjusted version was not used in the 2009 detailed tables (OAS, 2010a) and the 2009 mental health detailed tables (OAS, 2010b), but they can be used in the analyses of any 2008 MDE data. In 2008 and 2009, adult MDE estimates were not compared with MDE estimates prior to 2008 when WHODAS questions were not asked prior to the depression module. Additionally, adjusted MDE variables for 2005 through 2007 have been developed that can be used to make the aforementioned comparisons (Aldworth, Kott, Yu, Mosquin, & Barnett-Walker, in press). Adult MDE prevalence rates for 2009 should be computed using the standard weight (ANALWT), but MDE estimates based only on the WHODAS sample use the split-sample weight (MHSAMPWT). All adjusted estimates should use ANALWT.

4. Missingness

In the 2009 NSDUH, many variables, including core drug and demographic variables, had missing item response values imputed. See the 2009 NSDUH imputation report (Ault et al., 2011) for further details. However, the missing item responses of many other variables were not imputed, and these missing responses may lead to biased estimates in the 2009 detailed tables (OAS, 2010a) and 2009 mental health detailed tables (OAS, 2010b). In addition, another source of potential uncertainty about some estimates may occur due to the way unknown item responses (e.g., blank, "don't know," "refused") were actually 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.

Recall from Section 3 that prevalence rates are defined as the proportions of the population who exhibit characteristics of interest. Let $\hat{p}_d$ represent the estimated prevalence rate 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 $\hat{Y}_d$ = estimated number of persons 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 exist 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 persons 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.

In the 2009 detailed tables and the 2009 mental health detailed tables, potential bias in the $\hat{N}_d$, $\hat{Y}_d$, or $\hat{p}_d$ estimates was not treated, although footnotes included on the tables provide detailed information about which estimates were based on or excluded missing values.

This problem may be illustrated by the following example, which corresponds to information presented in Tables 2.9A and 2.9B of the 2009 mental health detailed tables.

Mental health Table 2.9A presents estimates of the past year use of several types of illicit drugs among persons aged 12 to 17 for 2008 and 2009. These analysis variables are grouped into a two-level domain variable that is categorized according to whether a respondent had a past year MDE. In 2009, mental health Table 3.2A shows the population estimate of persons aged 12 to 17 as approximately 24,609,000. However, the subdomain population estimates summed to approximately 24,102,000, resulting in an estimate of $\hat{N}_m$ = 507,000 (approximately 2.1 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 2009, the population estimate of persons aged 12 to 17 who used illicit drugs in the past year was approximately 4,809,000. However, the subdomain population estimates summed to 4,691,000, resulting in an estimate of $\hat{Y}_m$ = 119,000 (approximately 2.5 percent of the total population).

Mental health Table 2.9B presents prevalence estimates of the past year use of several types of illicit drugs among persons aged 12 to 17 for 2008 and 2009. Because $\hat{N}_m$ is positive and $\hat{Y}_m$ is positive for the analysis variable, "Illicit Drugs," the prevalence estimates for this variable may be biased by some unknown amount across the two domains. The 2009 prevalence estimates reported in Table 2.9B for youths who had or did not have past year MDE are 35.7 and 18.0 percent, respectively. It can be shown that the approximate range of possible bias values for each of these estimates is as follows: between -5.93 and 3.69 percent and between 0.44 and -0.31 percent, respectively.

5. Sampling Error

As were the prevalence rates, all of the variance estimates for prevalences (including those for prevalence based on annual averages from combined data) were 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.

Due to the complex nature of the sampling design for the NSDUH (specifically the use of stratified-clustering sampling), key nesting variables were created for use in SUDAAN to capture explicit stratification and to identify clustering. Starting with the 2005 NSDUH, there was a change made in the way the key nesting variables were defined. Each SS region appears in a different variance estimation stratum every quarter. This method had the effect of assigning the regions to strata in a pseudo-random fashion while ensuring that each stratum consists of four SS regions from four different States.

Two replicates per year were 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 standard errors for estimates based on combined data from adjacent survey years.

Although the standard errors (SEs) of estimates of means and proportions can be calculated appropriately in SUDAAN using a Taylor series linearization approach, SEs of estimates of totals may be underestimated in situations where the domain size is poststratified to data from the U.S. Census Bureau. Because of this underestimation, alternatives for estimating SEs of totals were implemented in all of the 2009 detailed tables (OAS, 2010a) and 2009 mental health detailed tables (OAS, 2010b), where appropriate.

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

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

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 persons in domain *d* (both users and nonusers). The SUDAAN software package is used to calculate direct estimates of $\hat{Y}_d$ and $\hat{N}_d$ and also 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, an appropriate estimate of the SE for the total number of substance users is

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

This approach is theoretically correct 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 (Chen et al., 2011). In these cases, $\hat{N}_d$ is not subject to a sampling error induced by the NSDUH design.

For estimated domain totals, $\hat{Y}_d$, where $\hat{N}_d$ is not fixed (i.e., where domain size estimates are not forced to match the U.S. Census Bureau population estimates), this formulation 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 cases in this study.

For various subsets of estimates, the above approach yielded an underestimate of the variance of a total because $\hat{N}_d$ was subject to considerable variation. 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 formula above 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. This method was applied to selected tables in the 2004 NSDUH, and it was implemented across all tables starting with the 2005 NSDUH and continuing through the 2009 NSDUH. 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 formula above, and all other estimates were calculated directly in SUDAAN, regardless of other estimates within the same table. The set of domains considered controlled (i.e., those with a fixed $\hat{N}_d$) was restricted to main effects and two-way interactions in order to maintain continuity between years. Domains consisting of three-way interactions may be controlled in 1 year but not necessarily in preceding or subsequent years. The use of 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 example SAS and SUDAAN code that demonstrate how to compute standard errors of proportions as well as both types of standard errors of totals (controlled or uncontrolled).

Table 1 contains a list of domains with a fixed $\hat{N}_d$. This table includes both the main effects and two-way interactions and may be used to identify the method of SE calculation

employed for estimates of totals in the 2009 detailed tables and 2009 mental health detailed tables. For example, Table 1.23 of the 2009 detailed tables presents estimates of illicit drug use among persons 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) were treated as controlled in this table, and the formula above was used 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. It is important to note that estimates presented in the 2009 detailed tables and 2009 mental health 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. Although not reported on in the 2009 detailed tables or 2009 mental health detailed tables, additional geographic interactions are also treated as domains with fixed $\hat{N}_d$ for other NSDUH analyses. Similar to geographic region and division, a State is considered a controlled domain, and two-way interactions with State and gender, Hispanic origin, quarter, and age group (12–17, 18–25, and 26 or older) are all treated as domains with fixed $\hat{N}_d$.

Table 1. Demographic and Geographic Domains Forced to Match Their Respective U.S. Census Bureau Population Estimates through the Weight Calibration Process (since 2005)

Main Effects	Two-Way Interactions
Age Group	
12-17	
18-25	**Age Group * Gender**
26-34	(e.g., Males Aged 12 to 17)
35-49	
50-64	
65 or Older	**Age Group * Hispanic Origin**
All Combinations of Groups Listed Above[1]	(e.g., Hispanics or Latinos Aged 18 to 25)
Gender	
Male	
Female	**Age Group * Race**
Hispanic Origin	(e.g., Whites Aged 26 or Older)
Hispanic or Latino	
Not Hispanic or Latino	
Race	**Age Group * Geographic Region**
White	(e.g., Persons Aged 12 to 25 in the Northeast)
Black or African American	
Geographic Region	
Northeast	**Age Group * Geographic Division**
Midwest	(e.g., Persons Aged 65 or Older in New England)
South	
West	
Geographic Division	**Gender * Hispanic Origin**
New England	(e.g., Not Hispanic or Latino Males)
Middle Atlantic	
East North Central	
West North Central	**Hispanic Origin * Race**
South Atlantic	(e.g., Not Hispanic or Latino Whites)
East South Central	
West South Central	
Mountain	
Pacific	

NOTE: An asterisk "*" represents an interaction between two variables.

[1] Combinations of the age groups (including but not limited to 12 or older, 18 or older, 26 or older, 35 or older, and 50 or older) also were forced to match their respective U.S. Census Bureau population estimates through the weight calibration process.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2009.

6. Degrees of Freedom

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 degrees of freedom 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.

Starting in 2005, there was a change in definition to the variance estimation strata for the NSDUH. This change in definition, which was applied to the 2005–2009 NSDUHs, has the effect of increasing the number of degrees of freedom for State-level estimates while preserving the number of degrees of freedom for national estimates (900). The degrees of freedom are calculated as the number of primary sampling units (variance replicates) minus the number of strata for the data being analyzed. Because the 5-year NSDUH design provides for estimates by State in all 50 States plus the District of Columbia, States may be viewed as the first level of stratification. When producing NSDUH estimates on the national level, including estimates based on annual averages from combined data, there are 900 degrees of freedom. If an analysis only involves certain States, the degrees of freedom change depending on whether the State is a large sample or small sample State. The large sample States (i.e., California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas) each have 192 degrees of freedom because each large State is in 192 strata. All of the other States (i.e., the small sample States, which include the District of Columbia) have 48 degrees of freedom because each small State is in 48 different strata. Note that the 2009 detailed tables (OAS, 2010a) and 2009 mental health detailed tables (OAS, 2010b) use 900 degrees of freedom for all estimates, including those for geographic regions and divisions. Appendix A contains an example demonstrating how to define the degrees of freedom within the SUDAAN procedure to compute design-based estimates.

For an analysis of a group of States, the degrees of freedom would be less than or equal to the sum of the degrees of freedom for each individual State due to overlap of strata. The specific number of degrees of freedom can be computed by counting the unique values of VESTR (variance estimation [pseudo] stratum) for the particular geographic area of interest. For these type of specific State analyses (or other subpopulations of interest), the degrees of freedom can be specifically indicated in SUDAAN; otherwise, the degrees of freedom are computed using the entire dataset. Similar methods can be used to compute appropriate degrees of freedom for any geographic region comprised of counties or States as well. The technique of counting the number of unique values of VESTR can also be used for analyses combining survey data across years.

7. Statistical Significance of Differences

Once the degrees of freedom have been determined, various methods used to compare prevalence estimates may be employed. This section describes some of these methods. Customarily, the observed difference between estimates is evaluated in terms of its statistical significance. Statistical significance is based on the *p* value of the test statistic and refers to the probability that a difference as large as that observed would occur due to random variability in the estimates if there were no difference in the prevalence rates being compared. The significance of observed differences is generally reported at the .05 and .01 levels when the *p* value is defined as less than or equal to the designated significance level.

Significance tests were conducted on differences between prevalence estimates from the 2009 NSDUH and previous years of NSDUH back to 2002. Due to survey design changes implemented in 2002, data from the 2002–2009 NSDUHs should not be compared with data from earlier survey years. Significance tests also were conducted on differences of prevalence estimates between combined 2006–2007 survey data and combined 2008–2009 survey data. Within-year tests were conducted on differences between prevalence estimates for various populations (or subgroups) of interest using data from the 2009 survey.

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

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

where *df* = the appropriate degrees of freedom, $\hat{p}_1$ = first prevalence estimate, $\hat{p}_2$ = second prevalence estimate, $\mathrm{var}(\hat{p}_1)$ = variance of first prevalence estimate, $\mathrm{var}(\hat{p}_2)$ = variance of second prevalence estimate, and $\mathrm{cov}(\hat{p}_1, \hat{p}_2)$ = covariance between $\hat{p}_1$ and $\hat{p}_2$. Note that the first and second prevalence estimates may take the form of prevalence estimates from two different survey years (e.g., 2008 and 2009, respectively), prevalence estimates from sets of combined survey data (e.g., 2006–2007 annual averages and 2008–2009 annual averages, respectively), or prevalence estimates for populations of interest within a single survey year.

Under the null hypothesis, *t* is distributed as a random variable from the *t*-distribution. Therefore, calculated values of *t*, along with the appropriate degrees of freedom, 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* will, in general, not be equal to zero. SUDAAN 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 outside 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 outside SUDAAN. SAS and SUDAAN examples showing the computational methods for generating *p* values of estimates of *t* and estimated totals can be found in Appendix A.

As the degrees of freedom approach infinity, the *t* distribution approaches the standard normal (*Z*) distribution. That is, because most of the statistical tests performed have 900 degrees of freedom, the *t* tests performed produce approximately the same numerical results as if a *Z* test had been performed.

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 were conducted first to control the error level for multiple comparisons. 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, 2008).

If SUDAAN is not available to compute the significance testing, using published estimates can provide similar testing results. When comparing prevalence rates shown with SEs, independent *t* tests for the difference of proportions can be performed and usually will provide the same results as tests performed in SUDAAN. 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. 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 rates shown with 95 percent CIs, it is important to note that two overlapping 95 percent CIs do not imply that their rates are statistically equivalent at the 5 percent level of significance. For additional information, see Schenker and Gentleman (2001) and Payton, Greenstone, and Schenker (2003).

8. Confidence Intervals

In some NSDUH publications, sampling error has been quantified using 95 percent CIs. Because NSDUH estimates are frequently small percentages, the CIs are based on logit transformations. Logit transformations yield asymmetric interval boundaries between 0 and 1 that are more balanced with respect to the probability that the true value falls below or above the interval boundaries than is the case for standard symmetric CIs for small proportions.

To illustrate the method, let the proportion P_d represent the true prevalence rate 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 "1n" 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

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

where $\mathrm{var}(\hat{p}_d)$ 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 constant chosen to yield a level of confidence based on the degrees of freedom. In order to produce 95 percent confidence limits for national estimates, the value of *K* would be 1.96 based on 900 degrees of freedom (similarly, for large States, *K* would be 1.97 based on 192 degrees of freedom, and for small States, *K* would be 1.98 based on 48 degrees of freedom).

Applying the inverse logit transformation to *A* and *B* above yields a confidence interval 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 $\hat{N}_d$ is not fixed, the CI approximation assumes that the sampling variation in $\hat{N}_d$ is negligible relative to the sampling variation in $\hat{p}_d$.

9. Incidence Estimates

In epidemiological studies, incidence is defined as the number of new cases of a disease occurring within a specific period of time. Similarly, in substance use studies, incidence refers to the first use of a particular substance.

Starting with the 2004 NSDUH data, the evaluation of trends in the initiation of drug use was presented by estimates of past year drug use incidence or initiation (i.e., the number of users whose first use was within the 12 months prior to their interview date). This incidence measure, termed "past year initiation," is determined by self-reported past year use, age at first use, year and month of most recent new use, and the interview date.

Since 1999, the NSDUH questionnaire allowed for the collection of year and month of first use for recent initiates (i.e., persons who used a particular substance for the first time in a given survey year). Month, day, and year of birth also were obtained directly or imputed for item nonrespondents as part of the data processing. In addition, the questionnaire call record provided the date of the interview. By imputing a day of first use within the year and month of first use, a specific date of first use, $t_{fu,d,i}$, can be used for estimation purposes.

Past year initiation among persons using a substance in the past year can be viewed as an indicator variable defined as follows:

$$I_{(Past\ Year\ Initiate)}(i) = \begin{cases} 1 & \text{if } \left(DOI_i MOI_i YOI_i - t_{fu,d,i}\right) \leq 365 \\ 0 & \text{otherwise} \end{cases},$$

where DOI_i, MOI_i, and YOI_i denote the day, month, and year of the interview for person *i*, respectively, and $t_{fu,d,i}$ denotes the date of first use associated to person *i*.

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 incidence estimates is the relationship between a main substance category and subcategories of substances (e.g., illicit drugs would be a main category and inhalants and marijuana would be examples of subcategories in relation to illicit drugs). 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 a particular drug, then he or she is also a past month user of illicit drugs in general). However, this is not the case with regard to incidence statistics. Because an individual can only be an initiate of a particular substance category (main or sub) a single time, a respondent with lifetime use of multiple substances may not, by

necessity, be included as an initiate of a main category, even if he or she were an initiate for a particular subcategory because his or her first initiation of other substances could have occurred earlier.

In addition to estimates of the number of persons initiating use of a substance in the past year, estimates of the mean age of past year first-time users of these substances were computed. Unless specified otherwise, estimates of the mean age at initiation in the past 12 months have been restricted to persons 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 constraint 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 civilian, noninstitutionalized population in the United States. This constraint was applied only to estimates of mean age at first use and does not affect estimates of incidence.

Because NSDUH is a survey of persons aged 12 years old or older at the time of the interview, younger individuals in the sample dwelling units are not eligible for selection into the NSDUH sample. Some of these younger persons 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 earlier years, data can be obtained retrospectively based on the age at and date of first use. As an example, persons who were 12 years old on the date of their interview in the 2009 survey may have reported initiating use of cigarettes between 1 and 2 years ago; these persons would have been past year initiates reported in the 2008 survey had persons who were 11 years old on the date of the 2008 interview been allowed to participate in the survey. Similarly, estimates of past year use by younger persons (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 an impression of the potential undercoverage in the current year, reports of substance use initiation reported in 2009 by persons aged 12 or older were estimated for the years in which these persons would have been 1 to 11 years younger. These estimates do not necessarily reflect behavior by persons who were 1 to 11 years younger in 2009. Instead, the data for the 11-year-olds reflect initiation in the year prior to the 2009 survey, the data for the 10-year-olds reflect behavior between the 12th and 23rd month prior to the 2009 survey, and so on. A very rough 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 2009 to the same estimates for the prior applicable survey year. To illustrate the calculation, consider past year use of alcohol. In the 2009 survey, 100,376 persons who were 12 years old in 2008 were estimated to have initiated use of alcohol between 1 and 2 years earlier. These persons would have been past year initiates in the 2008 survey conducted on the same dates had the 2008 survey covered younger persons. The estimated number of lifetime users currently aged 12 to 17 was 9,382,813 for 2009 and 9,540,037 for 2008, indicating fewer overall initiates of alcohol use among persons aged 17 or younger in 2009. Thus, an adjusted estimate of initiation of alcohol use by persons who were 11 years old in 2009 is given by

$$(Estimated\ Past\ Year\ Initiates\ Age\ 11)_{2008} * \frac{(Estimated\ Lifetime\ Users\ Age\ 12\ to\ 17)_{2009}}{(Estimated\ Lifetime\ Users\ Age\ 12\ to\ 17)_{2008}}$$

Numerically, this yielded an adjusted estimate of 98,722 persons who were 11 years old on a 2008 survey date and initiated use of alcohol in the past year:

$$100{,}376 * \frac{9{,}382{,}813}{9{,}540{,}037} = 98{,}722.$$

A similar procedure was used to adjust the estimated number of past year initiates among persons who would have been 10 years old on the date of the interview in 2007 and for younger persons in earlier years. The overall adjusted estimate for past year initiates of alcohol use by persons aged 11 or younger on the date of the interview was 230,373, or about 5.1 percent of the estimate based on past year initiation by persons aged 12 or older only (230,373 ÷ 4,560,449 = 0.0505).

Based on similar analyses, the estimated undercoverage of past year initiates aged 11 or younger was about 4.8 percent for cigarettes, about 1.2 percent for marijuana, and about 18.7 percent for inhalants. These 2009 results are comparable with undercoverage estimates presented in prior reports using data from the 2005–2008 surveys.

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 current survey and the numbers of persons aged 11 or younger in the past year obtained in the aforementioned analysis for estimating undercoverage of past year initiates. A comparison of the reported mean age at first use estimate and the corresponding adjusted estimate resulted in a change from 16.9 to 16.5 (or a decrease of 2.4 percent) for alcohol, from 17.5 to 17.0 (or a decrease of 2.6 percent) for cigarettes, from 17.0 to 16.9 (or a decrease of 0.4 percent) for marijuana, and from 16.9 to 15.6 (or a decrease of 7.7 percent) for inhalants. The percentage decreases reported above are comparable with results generated in prior survey years.

10. Suppression of Estimates with Low Precision

Direct survey estimates that were considered to be unreliable due to unacceptably large sampling errors were not reported, but rather were noted by an asterisk (*). The criterion used for suppressing all direct survey estimates was based on the relative standard error (RSE), which is defined as the ratio of the SE over the estimate.

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

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

or

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

Based on a first-order Taylor series approximation of RSE $[-\ln(\hat{p})]$ and RSE $[-\ln(1-\hat{p})]$, the following suppression rule was used for computational purposes:

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

or

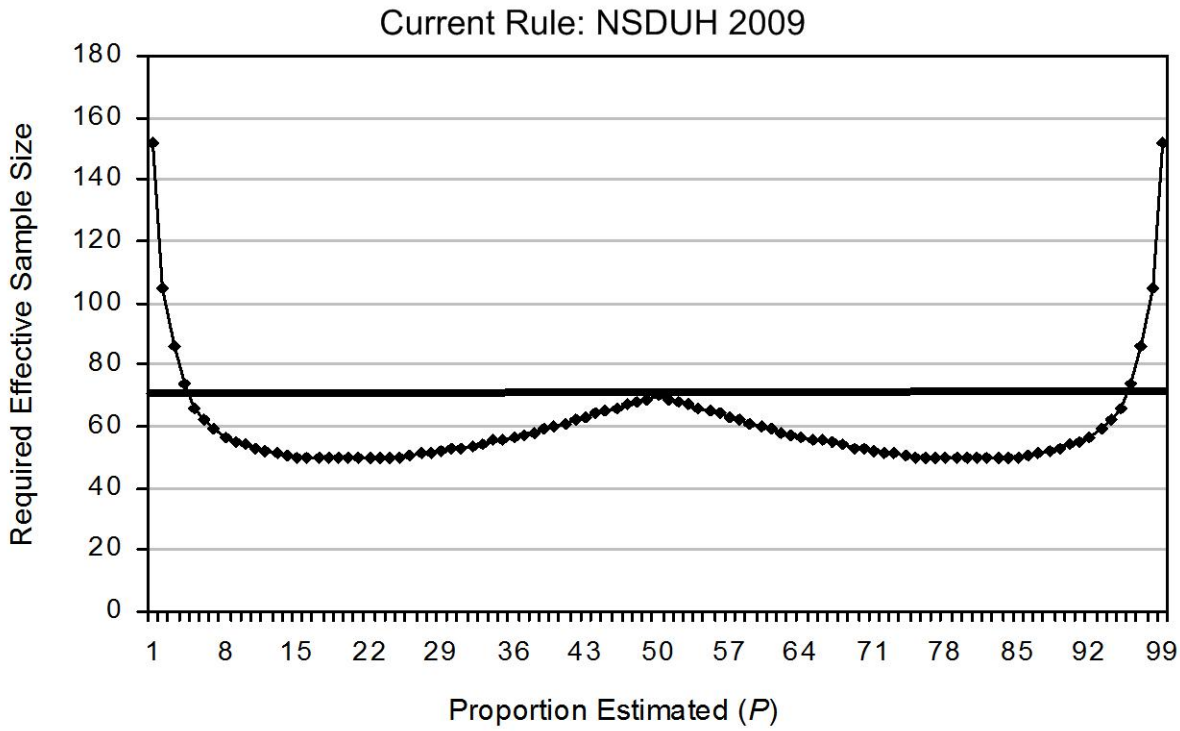
$$\frac{\text{SE}(\hat{p})/(1-\hat{p})}{-\ln(1-\hat{p})} > .175 \text{ when } \hat{p} > .5.$$

The separate formulas for $\hat{p} \le .5$ 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 Figure 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 $\hat{p} = .5$. So, to simplify requirements and maintain a conservative suppression rule, estimates of $\hat{p}$ between .05 and .95, which had effective sample sizes below 68, were suppressed.

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 in SAS. Since the 2005 NSDUH analysis, the direct SUDAAN design effect was used to provide a more precise and accurate reflection of the design effect (due to 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 only occur at approximately the tenth decimal place or later; however, previously published estimates that were on the borderline of being suppressed or unsuppressed due to the effective sample size suppression rule may potentially change from suppressed to unsuppressed, or vice versa.

Figure 1. Required Effective Sample as a Function of the Proportion Estimated



A minimum nominal sample size suppression criterion (n = 100) that protects against unreliable estimates caused by small design effects and small nominal sample sizes was employed. Prevalence estimates also were suppressed if they were close to 0 or 100 percent (i.e., if $\hat{p} < .00005$ or if $\hat{p} \geq .99995$).

Estimates of other totals (e.g., number of initiates), along with means and rates not bounded between 0 and 1 (e.g., mean age at first use) were suppressed if the RSEs of the estimates were larger than .5.

Additionally, estimates of the mean age at first use were suppressed if the sample sizes were smaller than 10 respondents; also, the estimated number of initiates was suppressed if they rounded to 0.

The suppression criteria for various NSDUH estimates are summarized in Table 2, and sample SAS code demonstrating how to implement these rules can be found in Appendix A.

Table 2. Summary of 2009 NSDUH Suppression Rules

Estimate	Suppress if:
Prevalence Rate, $\hat{p}$, with Nominal Sample Size, *n*, and Design Effect, *deff*	(1) The estimated prevalence rate, $\hat{p}$, is < 0.00005 or ≥ 0.99995, or (2) $\frac{\mathrm{SE}(\hat{p}) \ / \ \hat{p}}{-\ln(\hat{p})} > 0.175$ when $\hat{p} \le 0.5$, or $\frac{\mathrm{SE}(\hat{p}) \ / \ (1 - \hat{p})}{-\ln(1 - \hat{p})} > 0.175$ when $\hat{p} > 0.5$, or (3) Effective $n < 68$, where Effective $n = \frac{n}{deff}$ or (4) $n < 100$. Note: The rounding portion of this suppression rule for prevalence rates will produce some estimates that round at one decimal place to 0.0 or 100.0 percent but are not suppressed from the tables.
Estimated Number (Numerator of $\hat{p}$)	The estimated prevalence rate, $\hat{p}$, 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).
Mean Age at First Use, $\bar{x}$, with Nominal Sample Size, *n*	(1) $\mathrm{RSE}(\bar{x}) > 0.5$, or (2) $n < 10$.

SE = standard error; RSE = relative standard error; deff = design effect.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2009.

References

Aldworth, J., Barnett-Walker, K., Chromy, J., Karg, R., Morton, K., Novak, S., & Spagnola, K. (2009). Measuring serious mental illness with the NSDUH: Results of 2008 12-month analysis. In *2008 National Survey on Drug Use and Health: Methodological resource book* (Section 16, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Mental Health Surveillance Study Deliverable 5, RTI/0209009.423.006.008). Research Triangle Park, NC: RTI International.

Aldworth, J., Kott, P., Yu, F., Mosquin, P., & Barnett-Walker, K. (in press). Analysis of effects of 2008 NSDUH questionnaire changes: Adjusted adult MDE and SPD estimates for 2005-2008 and SMI estimates for 2005-2007 (prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, RTI/0209009.486.024). Research Triangle Park, NC: RTI International.

American Psychiatric Association. (2008). *Diagnostic and statistical manual of mental disorders, 4th ed., text revision (DSM-IV-TR)* [Web Page]. Retrieved November 20, 2010, from http://www.psych.org/mainmenu/research/dsmiv/dsmivtr

Ault, K., Barnett-Walker, K., Carpenter, L., Cummiskey, C., Frechtel, P., Laufenberg, J., Martin, P., Moore, A., & Scott, V. (2011). Imputation report [2009]. In *2009 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009.577.007). Research Triangle Park, NC: RTI International.

Chen, P., Cribb, D., Dai, L., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (2011). Person-level sampling weight calibration [2009]. In *2009 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Phase I, Deliverable No. 39, RTI/0209009.574.002). 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.

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, Office of Applied Studies, 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: Biometrics Research, New York State Psychiatric Institute.

Hughes, A., Muhuri, P., Sathe, N., & Spagnola, K. (2010). *State estimates of substance use from the 2007–2008 National Surveys on Drug Use and Health* (HHS Publication No. SMA 10-4472, NSDUH Series H-37). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies. [Available at http://www.oas.samhsa.gov/states.cfm]

Morton, K. B., Martin, P. C., Chromy, J. R., Foster, M., & Hirsch, E. L. (2010). Sample design report [2009]. In *2009 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Phase V, Deliverable No. 8, RTI/0209009.530.004). Research Triangle Park, NC: RTI International.

Office of Applied Studies. (2010a). *Results from the 2009 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Applied Studies. (2010b). *Results from the 2009 National Survey on Drug Use and Health: Mental health detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Applied Studies. (2010c). *Results from the 2009 National Survey on Drug Use and Health: Mental health findings* (HHS Publication No. SMA 10-4609, NSDUH Series H-39). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Applied Studies. (2010d). *Results from the 2009 National Survey on Drug Use and Health: Volume I. Summary of national findings* (HHS Publication No. SMA 10-4586Findings, NSDUH Series H-38A). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Applied Studies. (2010e). *Results from the 2009 National Survey on Drug Use and Health: Volume II. Technical appendices and selected prevalence tables* (HHS Publication No. SMA 10-4586Appendices, NSDUH Series H-38B). Rockville, MD: Substance Abuse and Mental Health Services Administration.

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.

Rehm, J., Üstün, T. B., Saxena, S., Nelson, C. B., Chatterji, S., Ivis, F., & Adlaf, E. (1999). On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. *International Journal of Methods in Psychiatric Research, 8*(2), 110-123.

RTI International. (2008). *SUDAAN® language manual, Release 10.0*. Research Triangle Park, NC: RTI International.

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, Office of Applied Studies, under Contract No. 283-03-9028, RTI/0208726.187.022). Research Triangle Park, NC: RTI International.

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.

Appendix A: Documentation for Conducting Various Statistical Procedures: SAS and SUDAAN Examples

This appendix contains example SAS® and SUDAAN® code that produces estimates of past month alcohol use by gender (males and females) using the statistical procedures documented within this report and implemented in the 2009 detailed tables (Office of Applied Studies [OAS], 2010a) and the 2009 mental health detailed tables (OAS, 2010b). The first SUDAAN example (Exhibit A.1) below produces estimates as described in Section 3 of the report, and the second (Exhibit A.4) and third (Exhibit A.6) SUDAAN examples perform statistical tests of differences as described in Section 7 of the report for the generated estimates. The first SAS example (Exhibit A.2) calculates the standard error (SE) of the total for controlled domains as discussed in Section 5, and the second SAS example (Exhibit A.3) implements the suppression rule as shown in Section 10. Note the SE of the total for uncontrolled domains is calculated within SUDAAN. The third SAS example (Exhibit A.5) produces the *p* value for the test of differences between totals for uncontrolled domains, whereas the last three SAS examples (Exhibits A.7, A.8, and A.9) combined produce the *p* value for the test of differences between totals for controlled domains.

Before running the SUDAAN procedures shown below, 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 and unweighted sample sizes, means, totals, standard errors of means and totals, as well as *p* values for testing of the means and totals.

The following options are specified within the SUDAAN examples below to correctly produce estimates using the National Survey on Drug Use and Health (NSDUH) data.

DESIGN=WR (with replacement)

Due to the 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.

Nesting Variables (VESTR and VEREP)

The nesting variables are used to capture explicit stratification and to identify clustering with the NSDUH data, which are needed in order 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 standard errors for estimates based on combined data from adjacent survey years.

Degrees of Freedom (DDF)

As described in Section 6 of this report, the degrees of freedom are 900 for national estimates, 192 for large States (California, Florida, Illinois, Michigan, New York, Pennsylvania, Ohio, Texas), and 48 for all other States. 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 comparing estimates in two domains with different degrees of freedom, err on the conservative side and use the smaller degrees of freedom.

Design Effect – DEFT4

This option within SUDAAN provides the correct measure of variance inflation due to stratification (or blocking), clustering, and unequal weighting in NSDUH estimation.

The following SAS and SUDAAN examples apply the specific NSDUH options described above to compute estimates, apply the suppression rule, and perform significance testing by using the data produced by the example in Exhibit A.1.

Generation of Estimates

Exhibit A.1 demonstrates how to compute various types of estimates, including the prevalence estimate (MEAN), standard error of the mean (SEMEAN), weighted sample size (WSUM), unweighted sample size (NSUM), weighted total (TOTAL), and the standard error of the totals (SETOTAL). Whether or not the standard error of the totals (SETOTAL) is taken directly from SUDAAN depends on whether or not the domain in question (i.e., gender in this example) is among those forced to match their respective U.S. Census Bureau population estimates through the weight calibration process. See the section below on standard errors for additional information.

Exhibit A.1 DESCRIPT Procedure (Estimate Generation)

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

PROC DESCRIPT DATA=DATANAME DDF=900 DESIGN=WR FILETYPE=SAS DEFT4;
NEST VESTR VEREP;
WEIGHT ANALWT;   /*Standard single-year, person-level analysis
weight*/
VAR ALCMON;   /*Past month alcohol analysis variable*/
SUBGROUP YEAR IRSEX;
      /*Year variable, where 2008=1 & 2009=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";
```

Exhibit A.1 DESCRIPT Procedure (Estimate Generation) (continued)

```
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;
```

Standard Errors

As discussed in Section 4, the SE for the mean (or proportion) comes directly out of SUDAAN defined in the output variable SEMEAN (Exhibit A.1). However, to compute the SE of the totals, there are different methods depending on whether the domains in question (i.e., gender in this example) are controlled or uncontrolled for during the weighting process. If a domain is uncontrolled for (i.e., it is not one of the domains described in Table 1 in Section 5), then the standard error of the total comes directly out of SUDAAN in the output variable SETOTAL. If the domain is controlled for (i.e., it is one of the domains described in Table 1), then the standard error of the total is calculated outside of SUDAAN as SETOTAL (SE of controlled domain) = WSUM (weighted sample size) * SEMEAN (SE for the mean/proportion). Because gender is controlled for, the standard error of the totals would not be taken directly from the Exhibit A.1 example but rather would be computed using this formula as shown in Exhibit A.2.

Exhibit A.2 SAS Code (Implementation of Standard Error of Total Calculation for Controlled Domains)

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

/*************************************************************
  Define SETOTAL for gender because it is a controlled domain
   In the SUDAAN procedure above IRSEX is in the class
      statement, 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;
```

Suppression Rule

As described in Section 10, each published NSDUH estimate goes through a suppression rule to detect if the estimate is unreliable due to an unacceptably large sampling error. The

suppression rules as they apply to different types of estimates are shown in Table 2 in Section 10. The example in Exhibit A.3 applies the prevalence rate rule.

Exhibit A.3 SAS Code (Implementation of Suppression Rule)

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

/******APPLY THE PREVALENCE RATE SUPRESSION 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;

IF (MEAN LT .00005) OR (MEAN GE 0.99995) OR (RSELNP GT 0.175) OR
(EFFNSUM < 68) OR (NSUM <100) THEN SUPRULE=1;

RUN;
```

Statistical Tests of Differences

As described in Section 7, significance tests were conducted on differences of prevalence estimates between the 2009 NSDUH and previous years of NSDUH back to 2002, as well as differences of prevalence estimates between combined 2006–2007 survey data and combined 2008–2009 survey data. Note that for year-to-year tests of differences, if the estimate for either year is suppressed, the resulting *p* value is also suppressed.

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 in Exhibit A.4 uses the DIFFVAR statement to test for differences between the 2008 and 2009 past month alcohol use estimates for all persons aged 12 or older (IRSEX=0), all males (IRSEX=1), and all females (IRSEX=2).

Similar to computing the standard errors of the totals, calculating *p* values for tests of differences of totals differs depending on whether an estimate is considered to be from a controlled domain or an uncontrolled domain. Both ways are described below with accompanying example code. Exhibits A.4 and A.5 show example code for uncontrolled domains, and Exhibits A.4, A.6, A.7, A.8, and A.9 show example code for controlled domains.

Exhibit A.4 DESCRIPT Procedure (Tests of Differences)

```
PROC DESCRIPT DATA=DATANAME DDF=900 DESIGN=WR FILETYPE=SAS;
NEST VESTR VEREP;
WEIGHT ANALWT;
VAR ALCMON;
SUBGROUP YEAR IRSEX;
LEVELS 2 2;
TABLES IRSEX;
DIFFVAR YEAR=(1 2); /*Tests of differences between 2008(year=1)
                      and 2009 (year=2)*/
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 OF 2008 AND 2009 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;
```

When one or more contrasts are specified in SUDAAN, as in the DIFFVAR statement above, the output variable MEAN becomes the contrast mean, and SEMEAN becomes the SE of the contrast mean. The example above also outputs 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, the *p* value (PVALT) for the test of differences between totals for uncontrolled domains can be calculated as indicated in Exhibit A.5. The SAS function PROBT returns the probability from a t distribution.

Exhibit A.5 SAS Code (Calculation of the *P* Value for the Test of Differences between Totals for Uncontrolled Domains)

```
IF SETOTAL GT 0.0 THEN DO;
   PVALT=2*(1-PROBT(ABS(TOTAL/SETOTAL),900));
END;
```

In the Exhibit A.1 example, all persons (aged 12 or older) and gender are annually controlled totals. For controlled domains like these, additional steps are needed to compute similar *p* values for tests of differences. One approach uses an additional DESCRIPT to output the appropriate covariance matrix (Exhibit A.6). Then, through further SAS 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 and calculated t-test statistic.

Exhibit A.6 DESCRIPT Procedure (Covariance Matrix)

```
PROC DESCRIPT DATA=DATANAME DDF=900 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;
```

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

```
CLASS YEAR IRSEX;
```

The covariances of the estimated means can be obtained from the output of the DESCRIPT procedure above (Exhibit A.6). The covariance matrix consists of a row and column for each gender (total, male, female) and year (both years, 2008, and 2009) combination with each cell corresponding to a particular variance component (i.e., a 9 x 9 matrix). Since 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. Below is a shell of what the covariance matrix would look like for this example.

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

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 either 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.7.

The three SAS datasets created by the examples below, one containing the covariances (Exhibit A.7) and two containing the variances (Exhibit A.8), are then merged with the output dataset from the DESCRIPT procedure that generated the tests of differences (Exhibit A.4). With the proper statistics contained in one dataset, the corresponding *p* value for the tests of differences between controlled totals can be produced using the SAS PROBT function and calculated t-test statistic (Exhibit A.9).

Exhibit A.7 SAS Code (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;
```

The variances of the means are calculated in separate data steps shown in Exhibit A.8. The variance is simply the SE of the mean squared. The SE of the means were output in the original DESCRIPT procedure that generated the estimates.

Exhibit A.8 SAS Code (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;

RUN;
```

Exhibit A.9 SAS Code (Calculation of the *P* Value for the Test of Differences between Totals for Controlled 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)),900));

RUN;
```