2014 NATIONAL SURVEY ON DRUG USE AND HEALTH

MENTAL HEALTH ESTIMATES COMPUTED DIRECTLY FROM THE CLINICAL SAMPLE OF THE MENTAL HEALTH SURVEILLANCE STUDY AND MEASURES OF THEIR STANDARD ERRORS

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

April 2016

This page intentionally left blank

2014 NATIONAL SURVEY ON DRUG USE AND HEALTH:

MENTAL HEALTH ESTIMATES COMPUTED DIRECTLY FROM THE CLINICAL SAMPLE OF THE MENTAL HEALTH SURVEILLANCE STUDY AND MEASURES OF THEIR STANDARD ERRORS

Contract No. HHSS283201300001C
RTI Project No. 0213757.004.107.008

RTI Authors:

Phillip S. Kott
Dan Liao

RTI Project Director:

David Hunter

SAMHSA Project Officer:

Peter Tice

SAMHSA Authors:

Jonaki Bose
Sarra Hedden
Matthew Williams
Art Hughes

For questions about this document, please e-mail Peter.Tice@samhsa.hhs.gov.

Prepared for Substance Abuse and Mental Health Services Administration, Rockville, Maryland

Prepared by RTI International, Research Triangle Park, North Carolina

April 2016

Recommended Citation: Center for Behavioral Health Statistics and Quality. (2016). *2014 National Survey on Drug Use and Health: Mental health estimates computed directly from the clinical sample of the Mental Health Surveillance Study and measures of their standard errors*. Substance Abuse and Mental Health Services Administration, Rockville, MD.

Acknowledgments

This methodological document was prepared for the Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality (CBHSQ), by RTI International (a registered trademark and a trade name of Research Triangle Institute). The authors thank Rebecca Ahrnsbrak from CHBSQ for her critical review and feedback on this document.

Table of Contents

Chapter **Page**

Appendix

This page intentionally left blank

List of Tables

Table **Page**

This page intentionally left blank

1. Introduction

The National Survey on Drug Use and Health (NSDUH), conducted by the Substance Abuse and Mental Health Services Administration (SAMHSA), is one of the primary sources of data for population-based prevalence estimates of substance use and mental health indicators in the United States. The NSDUH interview includes several self-administered indicators of mental health, such as assessments of lifetime and past year major depressive episode (MDE), past month and past year psychological distress and functional impairment, as well as past year suicidality. From 2008 to 2012, a subsample of NSDUH adult respondents were selected to participate in the Mental Health Surveillance Study (MHSS), which was a telephone interview that included clinical assessments of the presence of selected mental disorders. MHSS clinicians administered semistructured diagnostic interviews to this subsample to assess the presence of selected mental disorders (Aldworth et al., 2010).

The purpose of the MHSS clinical component was to develop a statistical model to apply to the full NSDUH sample that would generate serious mental illness (SMI) prevalence estimates among adults (aged 18 years or older) at national and state levels and to monitor the prevalence of SMI over time.

In addition to producing a model for the NSDUH to yield model-based estimates of SMI among adults (Center for Behavioral Health Statistics and Quality [CBHSQ], 2015a), the 2008 to 2012 MHSS clinical data can be used to generate nationally representative prevalence estimates of past year mental disorders among the adult civilian, noninstitutionalized population in 2008 to 2012, across a wide spectrum of diagnostic categories, including mood disorders (major depressive disorder [MDD], bipolar I disorder, and/or dysthymic disorder), anxiety disorders (posttraumatic stress disorder [PTSD], panic disorder with and without agoraphobia, agoraphobia without history of panic disorder, social phobia, specific phobia, obsessive compulsive disorder [OCD], and/or generalized anxiety disorder [GAD]), eating disorders (anorexia nervosa and/or bulimia nervosa), substance use disorders (alcohol abuse, alcohol dependence, illicit drug abuse, and/or illicit drug dependence), intermittent explosive disorder, adjustment disorder, as well as psychotic symptoms (delusions and/or hallucinations). Karg et al. (2014) presents the past 12-month prevalence estimates of specific mental disorders using the MHSS clinical data.

This document focuses on how the prevalence estimates and their standard errors were derived from the 2008 to 2012 MHSS clinical sample. In particular, it describes how the prevalence estimates covering the 2008 to 2012 time period were computed using sampling weights that had undergone a number of calibration adjustments with an emphasis on the last adjustment: poststratification—the annual calibration of the clinical sample to the NSDUH control totals. It then discusses several alternative methods for measuring the standard errors of those estimates. Consistent with how standard errors of NSDUH estimates are computed (CBHSQ, 2015a), all these methods use Taylor-series linearization variance estimators. The purpose of this document is not only to provide information on how the standard errors were computed in our existing MHSS reports, but also provide users of these data information on how they can compute standard errors to determine precision levels and conduct statistical inference.

The focus of this report will be on the statistical rather than measurement issues. That is, the statistical analyses discussed in this document assume the diagnostics made by the mental health professional during the clinical interview using the Structured Clinical Interview for the DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP) (First, Spitzer, Gibbon, & Williams, 2002) are accurate.

In addition, these analyses treat the yearly clinical samples as pure random samples—of both adults and time periods—with probabilities of selection accurately captured by the sample weights before the final poststratification. Therefore, these selection probabilities incorporate adjustments for unit nonresponse (unit response is treated like self-selection) and the exclusion from the clinical sample of adults who responded to the NSDUH main interview in Spanish.

A prevalence estimate is an estimated mean for a population. When measuring the standard error of a complex function of estimated population means, such as a regression coefficient, final weights can be treated as sampling weights because calibration adjustments have only marginal impact on these standard errors (except the impact resulting from changes in the weights themselves, which is captured by using the final weights). Moreover, software that fully incorporates the impact of calibration weighting using linearization techniques is not presently available.

More details on the probability sampling and weighting process can be found in Liao et al. (2014). Briefly, annual clinical sample weights were the product of five factors. The respondent's NSDUH main interview analysis weight was adjusted with

- a coverage adjustment to compensate for NSDUH main survey respondents who completed the survey in Spanish;
- the inverse of the probability the respondent was also selected for the clinical sample (the selection probability into the clinical sample was an independent function of an adult's NSDUH main interview responses, which varied across the years);
- a refusal adjustment to compensate for NSDUH respondents selected for clinical evaluation who did not wish to be recontacted;
- a second nonresponse adjustment to compensate mostly for those who agreed to be recontacted but were unavailable for the clinical evaluation (this included a few who agreed to be recontacted for the evaluations, but refused to respond when recontacted); and
- a poststratification adjustment to increase the efficiency of direct estimates from the clinical sample.

Strictly speaking the last adjustment is a calibration to totals computed from the NSDUH main interview respondents. We follow the terminology of SUDAAN 11 (RTI International, 2012) and call this adjustment "poststratification," even though the totals computed from the NSDUH main-interview responses were not for mutually exclusive groups.

This document focuses on that last weighting adjustment, but other features of the MHSS require some discussion. First, the adult NSDUH main sample in 2008 was randomly divided into two halves. One half sample, denoted the 2008A sample, was administered functional

impairment questions based on an abbreviated version of the World Health Organization Disability Assessment Schedule (WHODAS; Rehm et al., 1999). The other half sample, the 2008B sample, was administered questions based on the Sheehan Disability Scale (SDS; Leon, Olfson, Portera, Farber, & Sheehan, 1997). Both halves received psychological distress questions based on the Kessler 6 scale (K6; Kessler et al., 2003). From 2009 onward, only the WHODAS and K6 questions were used on the NSDUH main survey.

Weights were constructed separately each year, treating the 2008A and 2008B clinical samples as if they represented distinct years. In 2008 and 2009, these single-year samples were used to develop and verify statistical models that predicted SMI; however, because of to the small sample sizes (759 in 2008A, 741 in 2008B, 520 respondents from 2009, 516 in 2010, 1,495 in 2011, and 1,622 in 2012) the entire 5-year sample was used to produce the final statistical models. Similarly, the small sample sizes prevented annual estimation of the direct estimates as well.

Consequently, the clinical samples were also combined across the years to generate prevalence estimates of mental disorders. Because the sample size, sampling allocation, and weight adjustments for the clinical sample differed from year to year, gains in statistical efficiency could be realized by scaling the weights instead of letting each year contribute equally to the estimates.

These scaling factors were determined by focusing on the standard errors of prevalence estimates for SMI, any mental illness (AMI), and the occurrence of MDE in the previous year. A discussion of the assumptions underlying the determination and use of these factors and their implications on the estimation of prevalences and the annual numbers of adults with specific mental disorders over the 2008 to 2012 time period is contained in Chapter 2.

Chapter 3 describes how the clinical sample was calibrated to the NSDUH main sample each year in a nearly pseudo-optimal fashion (Kott, 2011). Chapter 4 shows how the WTADJX routine in SUDAAN 11 (RTI, 2012) was used to estimate yearly standard errors for totals and prevalences. As noted earlier, this chapter treats the weights before the final calibration as pure probability sampling weights based on the idea that this will, if anything, tend to overestimate standard errors (Kott & Day, 2014).

Chapter 5 describes how the standard error measures for prevalence estimates were calculated and discusses the implications of the displayed results. Because the weights for each year (2008A, 2008B, 2009, 2010, 2011, and 2012) were scaled when estimating the prevalences, the same scaling factors were used in computing the standard error measures.

The 900 original NSDUH variance strata (CBHSQ, 2015a) were collapsed into 100 MHSS variance strata so that MHSS standard error measures could be computed for Karg et al. (2014). An alternative set of collapsed strata had been employed in determining clinical weights and in modeling SMI (see Liao et al., 2014). Chapter 6 compares the standard errors computed using the two different sets of variance strata.

Chapter 7 provides further discussion of the statistical results in this document. It should be mentioned that by using WTADJX to measure standard errors, it is not possible, with the

software presently available, to conduct a Wald/F test when comparing prevalence estimates across three or more groups. That is why Bonferroni-adjusted *t* tests were used when comparing prevalence estimates across age groups in Karg et al. (2014).

2. The Scaling Factors

2.1 Estimating Prevalences

The yearly prevalence estimates for serious mental illness (SMI) were scaled to come close to minimizing the variance for the adult SMI prevalence estimate in 2008 to 2012. For the results from scaling the weights across years to be most relevant for prevalence estimates, we need either to (1) assume the underlying mental-health prevalence being estimated is constant across the years from 2008 to 2012, or (2) treat the target of estimation as the weighted mean of the annual prevalences, where the weight applied to each year is its scaling factor times its relative population size.

Mathematically, the true average prevalence from 2008 to 2012 can be expressed as

$$\overline{Y} = \frac{N_{2008}\overline{Y}_{2008} + N_{2009}\overline{Y}_{2009} + N_{2010}\overline{Y}_{2010} + N_{2011}\overline{Y}_{2011} + N_{2012}\overline{Y}_{2012}}{N_{2008} + N_{2009} + N_{2010} + N_{2011} + N_{2012}},$$

where N_t and $\overline{Y}_t$ are, respectively, the adult population size and the prevalence in year t. The assumption-free target of the scaled estimates is instead:

$$\overline{Y}_{scaled} = \frac{(.12)N_{2008}\overline{Y}_{2008} + (.04)N_{2009}\overline{Y}_{2009} + (.14)N_{2010}\overline{Y}_{2010} + (.35)N_{2011}\overline{Y}_{2011} + (.35)N_{2012}\overline{Y}_{2012}}{(.12)N_{2008} + (.04)N_{2009} + (.14)N_{2010} + (.35)N_{2011} + (.35)N_{2012}}. \quad (1)$$

We investigated the reasonableness of the former assumption that the 43 underlying mental health prevalence estimates were constant from 2008 to 2012 by computing the 5 yearly estimates for each variable (combining the 2008A and 2008B samples), then the standard errors of the $10 = \binom{5}{2}$ paired comparisons (e.g., the 2008 estimate for past year explosive disorder minus the 2010 estimate) using the fully corrected internal version of the standard error measure.

We deemed a difference (e.g., between the 2008 and 2010 estimates of a prevalence) to be statistically significant if the smallest of the 10 *p*-values per variable was less than .01. There was less than a 10 percent chance of this happening under the null hypothesis of an unchanging prevalence across the 5 years. Note that .01 is a Bonferroni adjustment applied to .1 (i.e., .01 =.1/10, with 10 being the number of paired comparisons per variable).

Three of the differences were statistically significant, which is about what should be expected with 43 variables (i.e., less than 4.3). There were 430 (43 x 10) paired comparisons in all. If we had alternatively used a Bonferroni adjustment for the lowest *p*-value of the 430 (.00030), the difference—and thus no difference—would be significant at the .1 level.

This means the clinical data were consistent with the null hypothesis of each prevalence remaining constant from 2008 to 2012. Note, however, that yearly sample sizes were small, so

our failure to reject the null hypothesis may have more to do with a lack of power than the underlying truth of the null hypothesis.[1]

2.2 Estimating Totals

It is more problematic to use the scaled weights when estimating the average yearly number of adults with a mental health disorder from 2008 to 2012 overall or within some demographic group (e.g. Hispanics) than when estimating yearly prevalences. This method, simply summing the scaled weights of relevant clinical interview respondents (where membership in the demographic group of interest defines relevance when needed), was used in Karg et al. (2014).

Scaling the weights actually estimates the numerator of equation (1); that is, the number of adults in the group having the disorder weighted by .12 in 2008, the number in 2009 by .04, and so forth. A more natural estimation target would weight each year equally. These targets are clearly different because the population grew between 2008 and 2012.

A possible alternative method for estimating an average yearly number of relevant adults with a mental disorder would be to compute the product of

- the relevant prevalence estimate calculated with the scaled weights, and
- the average yearly relevant population total computed from the main National Survey on Drug Use and Health (NSDUH) sample (i.e., scaling the main NSDUH weights by 1/5).

This alternative approach not only has a more natural estimation target, it should also result in smaller standard errors because it uses population estimates from the main NSDUH sample, which is considerably larger than the clinical sample. The main drawback of this "product" method is that it is more cumbersome to produce, requiring the computation of two statistics (one from the clinical sample and one from the main NSDUH sample) for each estimate. Another is that the standard error measures for the product estimates would be ad hoc.[2] By contrast, computing a standard error measure for an estimated average yearly total calculated with the scaled weights is straightforward.

[1] To illustrate the power, or lack of it, in our original Bonferroni-adjusted test, look at the yearly estimated prevalences of past year alcohol dependence or abuse from 2008 to 2012. The lowest *p*-value among the 10 pairwise comparisons is .012, which is greater than .010 and so not statistically significant at the Bonferroni-adjusted 0.1 level. (Note: the lowest pairwise comparison *p*-value for only 3 of the 43 variables is below .012.) The yearly prevalence estimates that were not significantly different for past year alcohol dependence or abuse as determined by this test ranged from 3.67 percent to 8.57 percent.

[2] The standard error measure could be converted for the estimated prevalence component into a coefficient of variation and then multiplied by the estimated average yearly population component. This assumes that the random nature of the latter estimate makes a negligible contribution to the variance of the product.

3. Nearly Pseudo-Optimal Calibration

The Mental Health Surveillance Survey (MHSS) clinical samples were calibrated separately in each year. This section describes how calibration was done in a particular year (with the 2008A and 2008B samples treated as if they were sampled from different years).

Let S denote the National Survey on Drug Use and Health (NSDUH) main adult respondent sample, w_k, denote the weights attached to main survey respondent k, and q_k denote the respondent's clinical sample weight after all adjustments for coverage and nonresponse but before the final calibration to the NSDUH main sample. By convention, $q_k = 0$ when adult k is a respondent to the NSDUH main interview but is either not sampled for the clinical interview or did not respond if sampled for some reason.

Let $a_k = q_k/w_k$. Given a vector of calibration variables $\mathbf{z}_k$ to be defined shortly and a scalar $T = .04(\sum_S w_k)$, the final adjustment factor for clinical interview respondent had this form:

$$f_k = \frac{\exp\left(\frac{U_k}{U_k - 1} a_k \mathbf{z}_k{}^T \mathbf{g}\right)}{1 + \left[\exp\left(\frac{U_k}{U_k - 1} a_k \mathbf{z}_k{}^T \mathbf{g}\right) - 1\right] \Big/ U_k}, \tag{2}$$

where $\mathbf{g}$ was chosen by successive linearizations (Newton's method) to satisfy the calibration equation:

$$\sum_S w_k \mathbf{z}_k = \sum_S q_k f_k \mathbf{z}_k, \tag{3}$$

and $U_k = T/q_k$ assures that no f_k is greater than U_k, which means that no final weight $\omega_k = q_k f_k$ exceeds T (i.e., 4 percent of the total of the weights). In fact, we first trimmed a few q_k to T before applying f_k: one in 2008A, one in 2008B, one in 2010, and three in 2009. The explanation for this and other choices inherent in equation (2) are contained in the following paragraphs.

The w_k in the NSDUH main respondent sample have been calibrated so that their sum equals the adult population size. By first trimming (an asymptotically ignorable number of weights) and then restricting the final clinical weights to be no greater than T, we are assuring that no single observation dominates a prevalence estimate, which is an implicit assumption of the asymptotics underlying probability sampling theory. It turned out that for a $\mathbf{g}$ to be found satisfying equations (2) and (3), U_k in (2) needed to be replaced by $1.25(T/q_k)$ for the 2008A clinical sample.

The vector $\mathbf{z}_k$ consisted of the following components, chosen to reduce the standard errors of the prevalence estimator for serious and any mental illness (AMI):

- indicators for six categories of gender (male and female) by age (18 to 25, 26 to 34, 35 or older) categories,

- indicators for four race/ethnicity categories (Hispanic, non-Hispanic white, non-Hispanic black, other),
- an indicator for past year suicidal thoughts,
- indicators from the NSDUH main interview for a past year and lifetime major depressive episode,
- interaction terms between an alternative Kessler 6 (K6) score and the three age categories, and
- interaction terms between an alternative World Health Organization Disability Assessment Schedule (WHODAS) score (or an alternative Sheehan Disability Scale [SDS] score for the 2008B sample) and the three age categories.

See Liao et al. (2014, Chapter 2) for details on the alternative K6, WHODAS, and SDS scores.

The a_k in equation (2) renders the adjustment factors nearly pseudo-optimal (Kott, 2011). If each $1/a_k$ were equal to the Poisson (i.e., independent across elements) probability that adult k is a respondent in the clinical sample given she or he is a respondent to the NSDUH main interview, then asymptotically optimal adjustment factors satisfying the calibration equation (3) would have the form: $f_k^{PO} = 1 + (a_k - 1)\mathbf{z}_k{}^T\mathbf{g}$. These factors can be negative and are unbounded.

A set of bounded, nonnegative adjustment factors asymptotically identical to the f_k^{PO} are

$$f_k^{NPO} = \frac{\exp\left(\frac{U_k}{U_k - 1}[a_k - 1]\mathbf{z}_k{}^T\mathbf{g}\right)}{1 + \left\{\exp\left(\frac{U_k}{U_k - 1}[a_k - 1]\mathbf{z}_k{}^T\mathbf{g}\right) - 1\right\} \Big/ U_k},$$

since $\mathbf{g} = \mathbf{O}_P(1/\sqrt{n})$ under mild conditions we assume to hold. Since all $a_k >> 1$, $f_k^{PO} \approx f_k^{NPO} \approx f_k$. (Replacing f_k by f_k^{NPO} would reduce standard error estimates in Table 1 by an average of less than 0.003 percent [this computation is discussed in Chapter 5]).

The adjustment factors produced by equation (2) can never be negative. As it happens, no final weight was less than 1. If it were necessary, we could have assured that all $\omega_k \geq 1$, by replacing equation (2) with

$$f_k{}^* = L_k + \frac{\exp\left(\frac{U_k - L_k}{(1 - L_k)(U_k - 1)} a_k \mathbf{z}_k{}^T\mathbf{g}\right)}{1/(1 - L_k) + \left[\exp\left(\frac{U_k - L_k}{(1 - L_k)(U_k - 1)} a_k \mathbf{z}_k{}^T\mathbf{g}\right) - 1\right] \Big/ (U_k - L_k)},$$

where $L_k = 1/q_k$. Because q_k is never less than 203 in the MHSS clinical sample, this is not necessary.

Another way to look at the weight adjustment function in equation (2) is to draw a distinction between the vector of *calibration variables*, $\mathbf{z}_k$, in equation (3) and the vector of *model variables* in equation (2), $\mathbf{x}_k = a_k\mathbf{z}_k$. In this formulation f_k is a function of $\mathbf{x}_k{}^T\mathbf{g}$. The f needs to be subscripted by k so that the U_k can vary.

This page intentionally left blank

4. Standard Error Estimation with WTADJX

We can express a calibration weighted total $t = \sum_S \omega_k y_k$, where ω_k is the calibration weight for adult k, as $t = \sum_S w_k \mathbf{z}_k{}^T \mathbf{b} + \sum_S \omega_k e_k$, where, for technical reasons explained in Kott and Liao (2015), the quasi-randomization regression coefficient is

$$\mathbf{b} = (\sum_S q_k f_k [(U_k - f_k)/(U_k - 1)] a_k \mathbf{z}_k \mathbf{z}_k{}^T)^{-1} \sum_S q_k f_k [(U_k - f_k)/(U_k - 1)] a_k \mathbf{z}_k y_k ,[3]$$

while $e_k = y_k - \mathbf{z}_k{}^T \mathbf{b}$.

This decomposition is effectively what WTADJX does. Each $\mathbf{x}_k = a_k \mathbf{z}_k$ in $\mathbf{b}$ can be viewed as a vector of model variables, while $\mathbf{z}_k{}^T$ in both $\mathbf{b}$ and e_k can be viewed as a (transposed) vector of calibration variables.

For analytical purposes, the NSDUH main adult sample has a stratified multistage design with ignorably small first-stage selection probabilities and the clinical sample is Poisson. As a result, the standard error of t can be estimated using the "with-replacement" linearization variance estimator by noting $t = \sum_S w_k h_k$, where $h_k = \mathbf{z}_k{}^T \mathbf{b} + (\omega_k / w_k) e_k$. The standard error of an estimated mean can be computed in an analogous manner since $\sum_S w_k = \sum_S \omega_k$ by our calibration equations (the sex/age categories exhaust the population).

Getting WTADJX with ADJUST = POST to compute these standard errors takes some innovation. First, let $S^{(1)}$ denote that sample of S for which $q_k > 0$. Then, create the dataset $S^{(2)} = S + S^{(1)}$. This new dataset contains two versions of the adults originally in $S^{(1)}$, which are treated as distinct elements of $S^{(2)}$ from the same variance primary sampling unit (PSU).

For the weight variable in WTADJX (WEIGHT), use q_k for elements originally from $S^{(1)}$ and w_k for elements from S. WTADJX allows different model and calibration variables. For the calibration variables (CALVARS), use the components of $\mathbf{z}_k$ from Chapter 3, multiplying each by 1 for elements from $S^{(1)}$ and by -1 for elements from S. The final sum of weights is then specified as 0 for all calibration variables (POSTWGT). For the model variables (in the MODEL statement: MODEL _ONE_ = [model variables]) multiply the components of $\mathbf{z}_k$ by 0 for elements from S and by a_k for elements from $S^{(1)}$, so that only the elements from $S^{(1)}$ are used in computing $\mathbf{b}$. As a result, only the weights for elements in $S^{(1)}$ are adjusted.

The y_k (which appeared in the VAR statement) for elements from S are implicitly set to missing. This means that these elements are treated as if they were outside the domain of interest but still have an impact on variance estimation. In particular, in the computation of the variance

[3]Since f_k is close to 1 (because $\mathbf{g}$ converges to 0 as the sample grows larger), nothing would be lost asymptotically in this application by replacing $f_k [(U_k - f_k)/(U_k - 1)]$ in $\mathbf{b}$ with 1 or f_k. The WTADJX software, however, does not make this simplification.

of $t = \sum_S \omega_k y_k$, an element k from $S^{(1)}$ contributes $y_k - \mathbf{z}_k{}^T\mathbf{b}$, while an element from S contributes $0 - (-\mathbf{z}_k{}^T)\mathbf{b} = \mathbf{z}_k{}^T\mathbf{b}$.

Finally, the WTMAX statement is used to truncate the q_k to T, while the UPPERBD statement interjects the U_k into the weight adjustments.

5. Exploring Alternative Standard Error Measures

As discussed in Chapters 1 and 2, when the clinical samples were combined across years, weights were scaled using the factors 0.06 for 2008A and 2008B, 0.04 for 2009, 0.14 for 2010, 0.35 for 2011, and 0.35 for 2012. In computing the standard errors for estimates computed with these scaled weights using WTADJX, the weights before the poststratification step (i.e., before the trimming) that appeared in the WEIGHT statement had to be scaled by the same factors.

The variance strata and variance primary sampling units (PSUs) used in WTADJX remained the same: MHVESTR and MHVEREP, which were designed for analysis of clinical data combined across years. All the variables in the MODEL and CALVARS step were crossed by a categorical indicator for year (CALV1, which ranges from 1 to 6[4]) because the clinical samples were calibrated yearly.

Table 1 displays estimated prevalences for the disorders in Karg et al. (2014) along with alternative measures of their standard errors. Chapters 3 and 4 explain how the "fully corrected internal" column of standard error measures in Table 1 (and used in Karg et al., 2014) was computed.

The "not corrected" column was computed using the DESCRIPT procedure in SUDAAN 11 with the scaled final clinical sample weights in the WEIGHT statement. There is no MODEL or CALVARS statement in DESCRIPT, and only data from the clinical sample were needed for the calculations.

An operational problem with the fully corrected internal method is that it requires access to data from all adult respondents to the entire National Survey on Drug Use and Health (NSDUH) main interview. It also requires categorical year indicators, the clinical weights *before* poststratification, and the T and U_k values in equation (2).

Clinical Mental Health Surveillance Survey (MHSS) datasets being made available to qualified researchers will not contain identifiers for the year of the interview. This is one reason why the final two standard error measures in Table 1 were calculated. Both assume that there was a single calibration across all years with the NSDUH main sample calibration targets either calculated within the WTADJX procedure ("internal") or not ("external"). The latter further simplifies standard error computation because it does not require that data from the full adult NSDUH main respondent sample be included in the calculations.

The hope was that one of these standard error measures could be used instead of the fully corrected internal one, despite the following changes in how the measure was determined:

[4] This variable is currently not available on the NSDUH main-sample dataset but can be created for users upon request.

- The final clinical weight appeared in the WEIGHT statement.
- Neither the model nor the calibration variables were cross-classified with the categorical year indicator (e.g., a single Hispanic indicator variable served as a calibration variable in place of separate yearly Hispanic indicators).
- Only clinical sample data were used in the program.

The last change meant that calibration targets were supplied by an external source. Scaled versions of the NSDUH main survey weights were used in computing the calibration variable targets in the POSTWGT statement.

Observe that since the computation of these alternative standard error measures started WTADJX with the final clinical weights, the program did not change the weights at all. In addition, model variables (for the MODEL statement) were created by multiplying the calibration variables not by a_k , as in Chapter 3, but by $a_k q_k/\omega_k$, where q_k is the clinical weight of adult k before poststratification and before trimming values greater than T to T. The result was to produce a new quasi-randomization regression coefficient in equation (4) more similar to the fully corrected one.

The standard error measure described above acknowledges some of the calibration but not the separate yearly targets; it is corrected, but not by year. Furthermore, it uses a source "external" to WTADJX for the calibration variable targets. An alternative "internal" measure was computed for comparison purposes. It too started with the final clinical weights and used the same model variables, but the target calibration totals were computed within the WTADJX program as they were with the fully corrected standard error measure. This captured any additional error caused by the calibration targets themselves being estimated from a sample.

Finally, a fully corrected external standard error measure was computed like the corrected external measure that was not correct by year, but with different calibration and model variables cross-classified with the categorical year indicator. In particular, it was computed using only the MHSS clinical sample, beginning with the final clinical sample weight, and not requiring knowledge of the T and U_k values. This method is much simpler to implement than that for the fully corrected internal standard error measure.

Appendix A provides SAS-callable SUDAAN code that could be used to reproduce the results in this document. In addition to estimated prevalences, the code also produced estimates and standard error measures for average yearly totals.[5]

[5] The target of this estimation weights the adults in 2008 by .12, adults in 2009 by .04, adults in 2010 by .14, and adults in 2011 or 2012 by .35. See ***Appendix A*** for a further discussion of this estimation target.

Table 1. Alternative Standard Error Measures for Mental Health Prevalence Estimates: NSDUH Adult Clinical Interview Data File, 2008 to 2012

Variable	Prevalence Estimate	Standard Error Measures				
		Not Corrected	Fully Corrected		Corrected, But Not by Year	
			Internal	External	Internal	External
Lifetime MDD	19.79	0.95	0.82	0.82	0.87	0.88
Lifetime MDE Disorder	20.68	0.97	0.84	0.84	0.87	0.89
Lifetime Manic Disorder	0.71	0.16	0.15	0.15	0.16	0.16
Lifetime Bipolar Disorder	0.69	0.16	0.15	0.15	0.16	0.16
Lifetime MDE or Manic Disorder	20.78	0.97	0.84	0.84	0.88	0.89
Past Year MDD	5.99	0.43	0.34	0.34	0.36	0.36
Past Year MDE Disorder	6.34	0.44	0.36	0.35	0.38	0.37
Past Year Manic Disorder	0.31	0.10	0.09	0.09	0.09	0.09
Past Year Dysthymic Disorder	1.70	0.27	0.25	0.24	0.26	0.26
Past Year Bipolar Disorder	0.39	0.10	0.09	0.09	0.10	0.10
Past Year Any Mood Disorder	7.40	0.52	0.39	0.39	0.42	0.42
Psych Screen	0.58	0.15	0.14	0.14	0.15	0.15
Posttraumatic Stress Disorder	0.74	0.10	0.10	0.10	0.10	0.09
Panic Disorder without Agoraphobia	0.89	0.14	0.13	0.13	0.13	0.13
Agoraphobia without History of Panic Disorder	0.21	0.08	0.07	0.07	0.08	0.08
Social Phobia	0.96	0.20	0.15	0.16	0.21	0.21
Specific Phobia	1.61	0.81	0.44	0.44	0.71	0.71
Obsessive-Compulsive Disorder	0.29	0.07	0.06	0.06	0.06	0.06
Generalized Anxiety Disorder	1.79	0.22	0.19	0.20	0.21	0.21
Any Anxiety Disorder	5.65	0.89	0.55	0.54	0.80	0.80
Explosive Disorder	0.39	0.09	0.08	0.09	0.09	0.09
SMI*	3.94	0.29	0.25	0.25	0.26	0.26
AMI*	17.95	0.97	0.75	0.75	0.87	0.87
Alcohol Abuse	3.07	0.45	0.48	0.50	0.44	0.44
Alcohol Dependence	3.28	0.39	0.38	0.38	0.41	0.41
Alcohol Dependence or Abuse	6.36	0.66	0.65	0.65	0.64	0.64
Illicit Drug Abuse	0.92	0.18	0.17	0.17	0.18	0.18
Illicit Drug Dependence	2.06	0.39	0.44	0.45	0.39	0.39
Illicit Drug Dependence or Abuse	2.98	0.43	0.47	0.48	0.44	0.44
Any SUD	7.77	0.70	0.71	0.71	0.68	0.68
Adjustment Disorder	6.89	0.50	0.48	0.49	0.50	0.50
Any Disorders (excluding SUD or Adjustment)	11.49	0.95	0.66	0.65	0.84	0.84
1 Disorder	8.01	0.92	0.60	0.60	0.83	0.83
2 Disorders	1.76	0.19	0.17	0.17	0.18	0.18
3+ Disorder	0.87	0.12	0.11	0.11	0.11	0.11
Any Disorders (excluding SUD)	17.11	1.11	0.85	0.86	0.98	0.98
1 Disorder	10.97	0.97	0.68	0.70	0.88	0.89
2 Disorders	3.18	0.42	0.45	0.46	0.42	0.42
3+ Disorders	1.91	0.28	0.25	0.22	0.27	0.27

(continued)

Table 1. Alternative Standard Error Measures for Mental Health Prevalence Estimates: NSDUH Adult Clinical Interview Data File, 2008 to 2012 (continued)

Variable	Prevalence Estimate	Standard Error Measures				
		Not Corrected	Fully Corrected		Corrected, But Not by Year	
			Internal	External	Internal	External
Any Disorders	22.52	1.16	0.92	0.93	1.04	1.04
1 Disorder	14.90	0.98	0.75	0.75	0.91	0.92
2 Disorders	4.10	0.47	0.48	0.50	0.45	0.45
3+ Disorders	2.19	0.29	0.27	0.26	0.28	0.28

AMI = any mental illness; MDD = major depressive disorder; MDE = major depressive episode; SMI = serious mental illness; SUD = substance use disorder.

* The prevalence estimates here are not based on a model like those in Substance Abuse and Mental Health Services Administration (2013). Moreover, these estimates scale the contributions from the component years (2008 through 2012) for statistical efficiency, which was neither necessary nor appropriate for the model-based estimates.

Any mood disorder is defined as having major depressive disorder, bipolar disorder (type I only), or dysthymic disorder in the past year.

Substance abuse and dependence are mutually exclusive. If a respondent is classified as having substance dependence (alcohol or illicit drugs), then he cannot be classified as abusing that substance regardless of responses to the abuse criteria questions.

Any disorder is defined as having one of the measured mood disorders, anxiety disorders, substance use disorders (included or excluded as specified in the header), eating disorders, adjustment disorder (included or excluded as specified in the header), or intermittent explosive disorder. A respondent can be classified as having any disorder even if the number of disorders is not able to be determined.

Combined variables are set to "Yes" if one or more source variable is "Yes," to "No" if all of the source variables are "No," and "missing" otherwise. Cases with missing values in the MHSS variables are excluded from the analyses.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

To summarize the differences between the fully corrected internal standard error measure and an alternative measure, the following statistic is calculated for each prevalence estimate:

$$D = \log(\text{alternative } se \text{ measure} \,/\, \text{fully-corrected-internal } se \text{ measure}) \times 100\%, \qquad (5)$$

Observe that D is very close to the percentage difference between the two *se* measures when that difference is within 10 percent. Unlike a standard percentage difference, however, D treats the numerator and denominator values in its internal ratio

$$\text{alternative } se \text{ measure} \,/\, \text{corrected-internal } se \text{ measure}$$

symmetrically. For example, when the numerator is twice the denominator, D is roughly 69 percent, although when the numerator is half the denominator, D is roughly - 69 percent. Therefore, a positive D value indicates the alternative *se* measure tends to overestimate the *se*; while a negative value indicates the alternative *se* measure tends to underestimate the *se*.

The average D value across the 43 estimates in Table 1 using the not corrected *se* measure is 13.5 percent. The median D *value* is 8.9 percent, with half the values ranging between 3.1 percent to 22.9 percent. Not all the D values are positive, however.

Table 2 summarizes the biases of the alternative standard error measures by measuring the differences between each and the fully corrected internal standard error measure using the D statistic in equation (5). On average, correcting, but not by year, removes less than half of the

bias in the not corrected standard error measure relative to the fully corrected version (13.5 percent is reduced to 8.1 percent or 7.9 percent). Using the external versions of the standard error measures tends to be slightly higher than their internal analogues. We will explore a possible reason for this anomalous result in Chapter 7.

Table 2. Summarizing Differences Between Alternative Standard Error Measure and the Fully Corrected Internal Method (in Percent, using the *D* Statistic*)

Method	Mean	Median	First Quartile	Last Quartile	Minimum	Maximum
Not Corrected	13.5	8.9	3.1	22.9	-14.0	61.2
Fully Corrected, External	0.4	0.1	-0.5	1.2	-1.8	4.2
Corrected, but not by Year						
External	7.9	5.1	1.4	11.3	-14.0	48.5
Internal	8.0	5.3	1.9	11.3	-14.2	48.5

* D = log(alternative se measure / fully corrected internal *se* measure) ×100%.

See Table 1 for other definitions.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Tables 1 and 2 show that the external and internal versions of a standard error measure are usually close. The biggest absolute difference in the two fully corrected measures was .02 (for alcohol abuse), whose D statistic was within 4.2 percent. The average percentage difference computed using the D statistic was 0.4 percent. SAMHSA only publishes estimated mental health prevalences and their standard errors to one decimal place.

This page intentionally left blank

6. Using Different Stratum Identifiers

As noted in the Chapter 1, the standard-error measures in Table 1 were computed using 100 variance strata. These are identified on Mental Health Surveillance Survey (MHSS) datasets by MHVESTR, with the two variance primary sampling units (PSUs) within each identified by MHVEREP. An alternative set of 50 variance strata, identified by MHVSTR09, were used for the methodological work described in Liao et al. (2014) and in testing for differences across the years. This was because not all variance PSUs would be represented in 2009 sample had the MHVESTR variance strata been used.

Table 3 displays the impact of using the alternative stratum identifiers on the not corrected standard error measures and the fully corrected internal standard error measures.

Using MHVSTR09 in place of MHVESTR resulted in an average decrease in the not corrected standard error measure of 0.5 percent computed using the *D* statistic and in the fully corrected internal measure of 1.5 percent, both small amounts. Asymptotic theory suggests that using the alternative stratum identifier should have no effect on the expectations of the variance estimates because we retained two PSUs per stratum. The differences we do see may reveal a limitation of applying asymptotic theory to finite samples or simply be the result of random noise.

Table 3. Standard Error Measures for Mental Health Prevalence Estimates Using Different Stratum Identifiers

Variable	Prevalence Estimate	Standard Error Measures			
		Not Corrected		Fully Corrected, Internal	
		MHVESTR	MHVSTR09	MHVESTR	MHVSTR09
Lifetime MDD	19.79	0.95	0.93	0.82	0.71
Lifetime MDE Disorder	20.68	0.97	0.94	0.84	0.71
Lifetime Manic Disorder	0.71	0.16	0.15	0.15	0.15
Lifetime Bipolar Disorder	0.69	0.16	0.16	0.15	0.15
Lifetime MDE or Manic Disorder	20.78	0.97	0.94	0.84	0.72
Past Year MDD Disorder	5.99	0.43	0.44	0.34	0.35
Past Year MDE Disorder	6.34	0.44	0.45	0.36	0.36
Past Year Manic Disorder	0.31	0.10	0.10	0.09	0.10
Past Year Dysthymic Disorder	1.70	0.27	0.25	0.25	0.24
Past Year Bipolar Disorder	0.39	0.10	0.10	0.09	0.10
Past Year Any Mood Disorder	7.40	0.52	0.52	0.39	0.39
Psych Screen	0.58	0.15	0.15	0.14	0.15
Posttraumatic Stress Disorder	0.74	0.10	0.10	0.10	0.10
Panic Disorder without Agoraphobia	0.89	0.15	0.14	0.15	0.13
Agoraphobia without History of Panic Disorder	0.21	0.08	0.07	0.07	0.06
Social Phobia	0.96	0.20	0.17	0.15	0.12
Specific Phobia	1.61	0.81	0.81	0.44	0.43
Obsessive-Compulsive Disorder	0.29	0.07	0.05	0.06	0.05
Generalized Anxiety Disorder	1.79	0.22	0.20	0.19	0.19
Any Anxiety Disorder	5.75	0.89	0.87	0.55	0.56
Explosive Disorder	0.39	0.09	0.08	0.08	0.08

(continued)

Table 3. Standard Error Measures for Mental Health Prevalence Estimates Using Different Stratum Identifiers (continued)

Variable	Prevalence Estimate	Standard Error Measures			
		Not Corrected		Fully Corrected-Internal	
		MHVESTR	MHVSTR09	MHVESTR	MHVSTR09
SMI*	3.94	0.29	0.27	0.25	0.24
AMI*	17.95	0.97	0.93	0.75	0.66
Alcohol Abuse	3.07	0.45	0.45	0.48	0.50
Alcohol Dependence	3.28	0.39	0.40	0.38	0.40
Alcohol Dependence or Abuse	6.36	0.66	0.65	0.65	0.66
Illicit Drug Abuse	0.92	0.18	0.21	0.17	0.20
Illicit Drug Dependence	2.06	0.39	0.42	0.44	0.49
Illicit Drug Dependence or Abuse	2.98	0.43	0.47	0.47	0.51
Any SUD	7.77	0.70	0.69	0.71	0.71
Adjustment Disorder	6.89	0.50	0.48	0.48	0.46
Any Disorders (excluding SUD or Adjustment)	11.57	0.95	0.95	0.65	0.62
1 Disorder	8.06	0.92	0.96	0.60	0.58
2 Disorders	1.76	0.19	0.18	0.17	0.16
3+ Disorder	0.91	0.12	0.11	0.11	0.11
Any Disorders (excluding SUD)	17.19	1.10	1.22	0.85	0.87
1 Disorder	11.02	0.98	1.11	0.68	0.77
2 Disorders	3.17	0.43	0.38	0.45	0.42
3+ Disorder	1.95	0.28	0.31	0.25	0.28
Any Disorders	22.58	1.16	1.25	0.92	0.91
1 Disorder	14.91	0.98	1.07	0.75	0.79
2 Disorders	4.12	0.47	0.46	0.48	0.48
3+ Disorder	2.22	0.29	0.32	0.27	0.30

AMI = any mental illness; MDD = major depressive disorder; MDE = major depressive episode; SMI = serious mental illness; SUD = substance use disorder.

* The prevalence estimates here are not based on a model like those in Substance Abuse and Mental Health Services Administration (2013). Moreover, these estimates scale the contributions from the component years (2008 through 2012) for statistical efficiency, which was neither necessary nor appropriate for the model-based estimates.

Any mood disorder is defined as having major depressive disorder, bipolar disorder (type I only), or dysthymic disorder in the past year.

Substance abuse and dependence are mutually exclusive. If a respondent is classified as having substance dependence (alcohol or illicit drugs), then he cannot be classified as abusing that substance regardless of responses to the abuse criteria questions.

Any disorder is defined as having one of the measured mood disorders, anxiety disorders, substance use disorders (included or excluded as specified in the header), eating disorders, adjustment disorder (included or excluded as specified in the header), or intermittent explosive disorder. A respondent can be classified as having any disorder even if the number of disorders is not able to be determined.

Combined variables are set to "Yes" if one or more source variable is "Yes," to "No" if all of the source variables are "No," and "missing" otherwise. Cases with missing values in the MHSS variables are excluded from the analyses.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

7. Discussion

The standard error measures labeled "internal" were designed to capture the added variability of the calibration targets based on estimates from the National Survey on Drug Use and Health (NSDUH) main survey. Nevertheless, the results in Table 2 suggest (as was the case) that they were, on average, smaller than otherwise analogous measures that treated those targets as fixed (the "external" measures). Exploring this apparently anomalous result, observe that

$$p = \left(\frac{\sum_S \omega_k y_k}{\sum_S \omega_k} - \frac{\sum_S w_k \mathbf{z}_k{}^T \mathbf{b}^*}{\sum_S w_k} \right) + \frac{\sum_S w_k \mathbf{z}_k{}^T \mathbf{b}^*}{\sum_S w_k}$$

$$\text{implies } Var(p) = Var(p - p_1) + Var(p_1) + 2Cov(p - p_1, p_1),$$

$$\text{where } p_1 = \frac{\sum_S w_k \mathbf{z}_k{}^T \mathbf{b}^*}{\sum_S w_k},$$

and $\mathbf{b}^*$ is the probability limit of $\mathbf{b}$ as the clinical sample size grows arbitrarily large, which we assume exists ($\mathbf{b}^*$, unlike $\mathbf{b}$, is not a random variable) because the NSDUH main sample is itself calibrated, $\sum_S \omega_k = \sum_S w_k$.

In the above formulation, $p = \sum_S \omega_k y_k / \sum_S \omega_k$ is a prevalence estimate based on the clinical sample, and $Var(p)$ is its variance. The expression $Var(p)$ is the full variance estimator used in the fully corrected internal standard error measure; $Var(p - p_1)$ is asymptotically the variance of p treating the calibration targets as fixed, which is used in the fully corrected external standard error measure, and $Var(p_1)$ is the direct contribution to the variance of p from the calibration targets themselves being based on a random sample. This is because $2Cov(p - p_1, p_1)$ tended to be negative and dominate $Var(p_1)$.

The negativity of $Cov(p - p_1, p_1)$ was largely a happy byproduct of inserting the a_k in equation (2), which was done to increase statistical efficiency. Recall that a_k is a product of all the weighting factors applied to the NSDUH main sample analysis weight w_k before poststratification. Removing the a_k before computing the fully corrected internal version of the standard error measures caused, on average, a 2.7 percent increase in the estimated coefficient of variation (CV). The estimated means also changed slightly, hence the use of estimated CV here (and the D statistic). As can be seen in Table 4, not all CV measures decreased from inserting the a_k. In fact, over a quarter decreased, and the median decrease was only 0.5 percent, which is still positive.

Removing the a_k from equation (2) but otherwise mimicking the production of the fully corrected internal version of the standard error measures tends to make $Cov(p - p_1, p_1)$ disappear. When external targets replaced internally computed ones with the a_k in equation

(2) removed from both, the standard error measure decreased 0.4 percent on average. The median decrease was of roughly the same size.

Table 4. Summarizing the Impact of Removing the a_k from Equation (2) on the Fully Corrected Estimator (in Percent, Applying the *D* Statistic to CVs*)

Comparisons	Mean	Median	First Quartile	Last Quartile	Minimum	Maximum
Fully Corrected, Internal Without vs. With the a_k	2.7	0.5	-3.1	5.2	-17.6	32.3
Fully Corrected, External vs. Fully Corrected, Internal Both Without the a_k	-0.4	-0.4	-0.8	0.1	-.2.8	0.9

* D = log(CV using alternative se measure / CV using fully corrected internal *se* measure) ×100%.

See Table 3 for other definitions.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

It is important to appreciate that most of the variance of p comes from $Var(p - p_1)$ no matter what the sign of $Cov(p - p_1, p_1)$. It should also be noted that the square of the fully corrected internal version of the standard error measure is only an *asymptotically* unbiased estimator for $Var(p)$ under a host of assumptions. In addition, this variance estimator itself has a variance.

One of the assumptions implicitly made by the fully corrected internal version of the standard error measure was that the calibration targets themselves were pure probability estimates based on the NSDUH main sample. Some of these targets, like the populations of the six age/gender categories, were in fact provided by the Census Bureau, while the others benefited from the calibration weighting in the NSDUH main sample. Therefore, it is encouraging that whether the calibration targets were treated as fixed there was only a very modest impact on the standard error measures.

Another assumption was that the weighting adjustment for the clinical sample before poststratification not only removed selection biases because of undercoverage and nonresponse, but actually estimated the probabilities of a Hispanic responding to the main NSDUH survey in English and an adult participating in the clinical interview exactly. In fact, the presence of these additional calibration weighting adjustments would tend to bias the fully corrected internal standard error measure *upward* (i.e., make them overestimate the true standard error). A detailed argument for this can be found in Kott and Day (2014). Briefly, if the residuals $e_k = y_k - \mathbf{z}_k{}^T\mathbf{b}$ were correlated with the covariates used in the coverage adjustment or one of the two nonresponse adjustments, then using calibration weighting techniques in those adjustments would incorporate more information about the undercovered and/or nonrespondents in the prevalence estimation—and thus decrease the variance of the resulting prevalence estimator—than would be reflected in the standard error measure.

Returning to the results displayed in Tables 1 and 2 and keeping in mind that standard error measures are themselves estimates subject to both bias and variance, there appears to be little argument against using the external version of the fully corrected standard error measure. The corrected measures that were not corrected by year, in contrast, appear not to capture

adequately the reduction in standard error because of the poststratification of the clinical sample. Therefore, it is recommended that users of the 2008 to 2012 MHSS clinical interview data employ either the internal or external versions of the fully corrected standard error measure for estimated prevalences.

This page intentionally left blank

References

Aldworth, J., Colpe, L., Gfroerer, J., Novak, S., Chromy, J., Barker P., Barnett-Walker, K., Karg, R., Morton, K., & Spagnola, K. (2010). The National Survey on Drug Use and Health and Mental Health Surveillance Study: Calibration analysis. *International Journal of Methods in Psychiatric Research, 19*(S1), 61-87.

Center for Behavioral Health Statistics and Quality. (2015a). *2013 National Survey on Drug Use and Health: Methodological resource book (Section 13, statistical inference report)*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2015b). *National Survey on Drug Use and Health: 2008-2012 adult clinical interview data file and codebook*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *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.

Karg, R. S., Bose, J., Batts, K. R., Forman-Hoffman, V. L., Liao, D., Hirsch, E., Pemberton, M. R., Colpe, L. J., & Hedden, S. L. (2014). *CBHSQ Data Review: Past year mental disorders among adults in the United States: Results from the 2008–2012 Mental Health Surveillance Study*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S. L., Manderscheid, R. W., Walters, E. E., & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry, 60,* 184-189.

Kott, P. S. (2011). A nearly pseudo-optimal method for keeping calibration weights from falling below unity in the absence of nonresponse or frame errors. *Pakistan Journal of Statistics, 27,* 391-396.

Kott, P. S., & Day, C. D. (2014). Developing calibration weights and standard-error estimates for a survey of drug-related emergency-department visits. *Journal of Official Statistics, 30,* 521-532.

Kott, P. S., & Liao, D. (2015). One step or two? Calibration weighting from a complete list frame with nonresponse. *Survey Methodology, 41,* 165-181.

Leon, A. C., Olfson, M., Portera, L., Farber, L., & Sheehan, D. V. (1997). Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *International Journal of Psychiatry in Medicine, 27*(2), 93-105.

Liao, D., Kott, P., Aldworth, J., Yu, F., Morton, K., Chen, P. Shook-Sa, B., Kott, P., & Davis, T. (2014). *2012 Mental Health Surveillance Study: Design and estimation report* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, RTI 0212800.001.112.008.006). Research Triangle Park, NC: RTI International.

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*, 110-123.

RTI International. (2012). *SUDAAN language manual, Release 11.0.* Research Triangle Park, NC: Author.

Substance Abuse and Mental Health Services Administration. (2013). *Results from the 2012 National Survey on Drug Use and Health: Mental health findings* (HHS Publication No. SMA 13-4805, NSDUH Series H-47). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Substance Abuse and Mental Health Services Administration. (2014). *Results from the 2013 National Survey on Drug Use and Health: Mental health findings* (HHS Publication No. SMA 14-4887, NSDUH Series H-49). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Appendix A:
SAS-callable SUDAAN Code for Computing Standard Error Measures with WTADJX

This appendix assumes the reader has access to 2008 to 2012 National Survey on Drug Use and Health (NSDUH) adult clinical interview data file and its codebook (CBHSQ, 2015b), which contains the names and definitions of the variables used in this appendix. This appendix demonstrates the sample SAS-callable SUDAAN codes that are used to compute standard error measures with WTADJX in Tables 1 and 2.

First, we need to create some additional variables that are unavailable in the 2008 to 2012 NSDUH adult clinical interview data file for computing the standard error measures.

Additional Variables Needed for All New Standard Error Measures

(i.e., excluding the not corrected ones):

```
AGESEX = CATAGMH2*10 + IRSEX;        * an age/sex categorical variable;

FUNCIMP = WHODASC3; IF CALV1 = 2 THEN FUNCIMP = SDSSC3;
                                     * create the functional impairment variable;

IF AMDELT = 1 THEN AMDEL2_U = 1; ELSE AMDEL2_U = 0;
                                     * recode 2 and missings with 0 for lifetime MDE;

MHVESTR = ceil((VESTR – 30000)/9);   * needed for the NSDUH main adult sample;
```

Needed for External Standard Error Measures:

```
CALFAC = MHNEWWGT/(ANALWT_A * MHADJ_4A**2 * MHADJ_4B**2);
                                       * helps create model variables for WTADJX;

FUNCIMPC = FUNCIMP * CALFAC;

WSPDSC2C = WSPDSC2 * CALFAC;
```

Needed for Fully Corrected Measures, both External and Internal:

```
CAGESEX    =  CALV1 * 100 + AGESEX;

CRACE4     =  CALV1 * 100 + RACE4;

CAMDEL2_U  =  CALV1 * 100 + AMDEL2_U ;

CAMDEY2_U  =  CALV1 * 100 + AMDEY2_U;

CMHSUTK_U  =  CALV1 * 100 + MHSUTK_U  ;

CCATAGMH2  =  CALV1 * 100 +  CATAGMH2;
```

Needed for Corrected-But-Not-By-Year-Internal Measures:

WSTART_N = MHFNLWGT if from clinical sample;
WSTART_N = [a scaled version of ANALWT created for the NSDUH main sample, available to users upon request]

CALFAC1 = 1 if from clinical sample;
CALFAC1= -1 if from NSDUH main sample;

CALFAC = MHNEWWGT/(ANALWT_A * MHADJ_4A**2 * MHADJ_4B**2) if from clinical sample;
CALFAC = 0 if from NSDUH main sample

FUNCIMP1 = FUNCIMP * CALFAC1;

WSPDSC21 = WSPDSC2 * CALFAC1;

FUNCIMPC = FUNCIMP * CALFAC;

WSPDSC2C = WSPDSC2 * CALFAC;

Needed for Fully Corrected Internal Measures:

BOUND is the total of the ANALWT_A in NSDUH main adult sample × .05 in 2008A;
BOUND is the total of the ANALWT_A in NSDUH main adult sample × .04 otherwise;

WSTART = MHFNLWGT/MHADJ_4B if from clinical sample;
WSTART = [a scaled version of ANALWT created for the NSDUH main sample, available to users upon request]

CALFAC1 = 1 if from clinical sample;
CALFAC1 = -1 if from NSDUH main sample

CALFAC2 = MHNEWWGT/(ANALWT_A * MHADJ_4A**2 * MHADJ_4B**2) if from clinical sample;
CALFAC2 = 0 if from NSDUH main sample

FUNCIMP1 = FUNCIMP * CALFAC1;

WSPDSC21 = WSPDSC2 * CALFAC1;

FUNCIMP2 = FUNCIMP * CALFAC2;

WSPDSC22 = WSPDSC2 * CALFAC2;

Second, we can use WTADJX in SUDAAN 11 to compute the standard error measures using the four alternative methods.

Sample Program for Fully Corrected External Measures (Data: clinical sample)

```
PROC WTADJX NOPRINT DATA = [clinical sample]  DESIGN = WR  NOTSORTED
ADJUST = POST;  NEST MHVESTR  MHVEREP; WEIGHT MHFNLWGT;
CLASS  IRSEX  RACE4  /*  Includes categorical variables for which one wants  estimates */
        CAGESEX  CRACE4 CMHSUTK_U  CAMDEL2_U CAMDEY2_U
        CCATAGMH2 /NOFREQS;   /* Must include categorical variables used in MODEL
                                 and CALVARS statements */
VAR  SCID_SMI SCID_AMI [and so on – variables for which one wants standard errors];
MODEL _ONE_ = CAGESEX*CALFAC CRACE4*CALFAC   CMHSUTK_U*CALFAC
              CAMDEL2_U *CALFAC CAMDEY2_U*CALFAC
              CCATAGMH2*FUNCIMPC CCATAGMH2*WSPDSC2C/NOINT;
CALVARS       CAGESEX          CRACE4          CMHSUTK_U
              CAMDEL2_U              CAMDEY2_U
              CCATAGMH2*FUNCIMP   CCATAGMH2*WSPDSC2/NOINT;
POSTWGT   /*The calibration  targets calculated based on the adult NSDUH main sample*/
  994000    982000   2963000   3027000   2557000   2973000    994000    982000
 2963000   3027000   2557000   2973000    675000    668000   1967000   2008000
 1746000   2024000   2420000   2351000   6818000   6982000   6298000   7230000
 6013000   5993000  16819000  17492000  16319000  18783000   6071000   6036000
16819000  17462000  16715000  19191000   9284000   1522000    869000   1820000
 9284000   1522000    869000   1820000   6216000   1055000    572000   1245000
21834000   3734000   2070000   4460000  54336000   9360000   5829000  11893000
54547000   9510000   6054000  12183000  13001000    494000  12988000    508000
 8750000    339000  30877000   1221000  78430000   2989000  79132000   3161000
11781000   1714000  11643000   1852000   7911000   1177000  28012000   4087000
71100000  10319000  71476000  10817000  12634000    861000  12525000    971000
 8495000    593000  29920000   2178000  76115000   5303000  76684000   5609000
 2351000   5330000   3599000    795000   2026000   1058000   1579000   3746000
 2415000   5625000  13032000   9029000  14075000  32196000  23332000  14816000
33683000  25616000   4791000   9227000   4458000   4758000   9830000   5153000
 3122000   6530000   3199000  11158000  22526000  12401000  27944000  55698000
;
OUTPUT MEAN SE_MEAN TOTAL SE_TOTAL/FILENAME = SE_OUTPUT  REPLACE;
RUN;
```

/* Creates output dataset SE_OUTPUT with means and standard errors for all the variable in the VAR statement within each category in the CLASS statement (e.g., the four race/ethnicities in RACE4). Given the size of the CLASS statement, printing the output is not advised, hence the NOPRINT in the first line of code */

Sample Program for Fully Corrected Internal Measures

```
(Data: clinical sample + NSDUH main adult sample)
PROC WTADJX NOPRINT DATA = [clinical sample + NSDUH main adult sample]
DESIGN = WR  NOTSORTED ADJUST = POST;
NEST MHVESTR  MHVEREP;WEIGHT WSTART;
CLASS IRSEX CAGESEX CRACE4 CMHSUTK_U CAMDEL2_U CAMDEY2_U
                                                    CCATAGMH2 /NOFREQS;
VAR  SCID_SMI SCID_AMI [and so on – variables for which one wants standard errors];
MODEL _ONE_ = CAGESEX*CALFAC2 CRACE4*CALFAC2   CMHSUTK_U*CALFAC2
                    CAMDEL2_U *CALFAC2 CAMDEY2_U*CALFAC2
                    CCATAGMH2*FUNCIMP2 CCATAGMH2*WSPDSC22/NOINT;
CALVARS           CAGESEX*CALFAC1 CRACE4*CALFAC1   CMHSUTK_U*CALFAC1
                    CAMDEL2_U *CALFAC1 CAMDEY2_U*CALFAC1
                    CCATAGMH2* FUNCIMP1 CCATAGMH2*WSPDSC21/NOINT;
POSTWGT
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
;
OUTPUT MEAN SE_MEAN TOTAL SE_TOTAL/FILENAME = SE_OUTPUT  REPLACE;
RUN;
```

Sample Program for Corrected-But-Not-By-Year-External Measures (Data: clinical sample)

```
PROC WTADJX NOPRINT DATA = [clinical sample]  DESIGN = WR  NOTSORTED
ADJUST = POST;  NEST MHVESTR  MHVEREP; WEIGHT MHFNLWGT;
CLASS IRSEX RACE4 AGESEX MHSUTK_U AMDEL2_U AMDEY2_U
                                                CATAGMH2 /NOFREQS;
VAR  SCID_SMI SCID_AMI [and so on – variables for which one wants standard errors];
MODEL _ONE_ = AGESEX*CALFAC RACE4*CALFAC   MHSUTK_U*CALFAC
                    AMDEL2_U *CALFAC AMDEY2_U*CALFAC
                    CATAGMH2*FUNCIMPC  CATAGMH2*WSPDSC2C/NOINT;
CALVARS           AGESEX RACE4 MHSUTK_U  AMDEL2_U AMDEY2_U
                    CATAGMH2*FUNCIMP  CATAGMH2*WSPDSC2/NOINT;
```

```
POSTWGT
 17167000    17011000     48349000    49999000    46191000    53174000
155501000    26703000     16263000    33422000   223178000     8712000
201923000    29966000    216374000    15516000    39241000    90013000
 65049000    81516000    160113000    91154000    ;
```

```
OUTPUT MEAN SE_MEAN TOTAL SE_TOTAL/FILENAME = SE_OUTPUT  REPLACE;
RUN;
```

Sample Program for Corrected-But-Not-By-Year-Internal Measures

```
(Data: clinical sample + NSDUH main adult sample)
PROC WTADJX NOPRINT DATA = clinical sample + NSDUH main adult sample]
DESIGN = WR  NOTSORTED ADJUST = POST;
NEST MHVESTR  MHVEREP; WEIGHT WSTART_N;
CLASS IRSEX RACE4 AGESEX MHSUTK_U AMDEL2_U AMDEY2_U
                                                    CATAGMH2 /NOFREQS;
VAR  SCID_SMI SCID_AMI [and so on – variables for which one wants standard errors];
MODEL _ONE_ = AGESEX*CALFAC RACE4*CALFAC   MHSUTK_U*CALFAC
                AMDEL2_U *CALFAC AMDEY2_U*CALFAC
                CATAGMH2*FUNCIMPC CATAGMH2*WSPDSC2C/NOINT;
CALVARS         AGESEX*CALFAC1 RACE4*CALFAC1   MHSUTK_U*CALFAC1
                AMDEL2_U *CALFAC1 AMDEY2_U*CALFAC1
                CATAGMH2*FUNCIMP1 CATAGMH2*WSPDSC21/NOINT;
POSTWGT
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
;
OUTPUT MEAN SE_MEAN TOTAL SE_TOTAL/FILENAME = SE_OUTPUT  REPLACE;
RUN;
```

An Example of an Output Dataset (Fully Corrected Internal Measures)

```
DATA SE_OUTPUT2;    SET SE_OUT PUT;
IF IRSEX >=0 OR RACE4 >= 0;
                /*  Restricts datasets to domains for these two class variables,
                    where IRACE4 = 0 in the dataset means across all race/ethnicities,
                     and    IRACE4 = -2 means race/ethnicity has been ignored */
MEAN = MEAN * 100; SE_MEAN = SE_MEAN*100;
                                    * converts variables to percentage form;
PROC PRINT; ID VARIABLE;   /* 1 for SCID_SMI;   2 FOR SCID_AMI   */

VAR  IRSEX RACE4  MEAN SE_MEAN;
FORMAT MEAN F5.2  SE_MEAN 5.2 ;
```

VARIABLE	IRSEX	RACE4	MEAN	SE_MEAN
1	-2	0	3.94	0.25
1	-2	1	4.38	0.30
1	-2	2	3.46	0.56
1	-2	3	4.55	1.20
1	-2	4	2.01	0.62
2	-2	0	17.95	0.75
2	-2	1	18.18	0.79
2	-2	2	15.83	2.46
2	-2	3	16.29	3.13
2	-2	4	19.40	3.67
1	0	-2	3.94	0.25
1	1	-2	2.96	0.32
1	2	-2	4.85	0.36
2	0	-2	17.95	0.75
2	1	-2	14.36	0.95
2	2	-2	21.29	1.14