Challenges of Using Prediction Models to Produce Nationally Representative Estimates of Serious Mental Illness

Sarra Hedden[1], Joseph Gfroerer[1], Jonaki Bose[1], Phillip Kott[2], Dan Liao[2], & Lisa Colpe[3]

[1]Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 1 Choke Cherry Rd, Rockville MD 20857
[2]RTI International; Research Triangle Park, NC 27709
[3]National Institute of Mental Health; 6001 Executive Boulevard, Bethesda, MD 20892

Overview

There has been a need for estimates of serious mental illness (SMI) at both the national and state level. Estimates of SMI are of particular importance for use in block grants used by States to apply for federal funds for mental health services. As such, the 1992 ADAMHA Reorganization Act required the Substance Abuse and Mental Health Services Administration (SAMHSA) to develop a definition of SMI and methodology for estimating SMI among adults. In response, in 1993 SAMHSA convened a technical advisory group that developed a definition of SMI, this definition was published in the *Federal Register*. Among adults, SMI was defined as

> "Persons aged 18 and over, who currently or at any time during the past year, have had diagnosable mental, behavioral, or emotional disorder of sufficient duration to meet diagnostic criteria specified within DSM-III-R that has resulted in functional impairment. "…" Functional impairment is defined as difficulties that substantially interfere with or limit role functioning in one or more major life activities including basic daily living skills; instrumental living skills; and functioning in social, family, and vocational/educational contexts."

In 2006, SAMHSA convened a technical advisory group consisting of mental health and statistical experts to seek guidance on how to measure and estimate serious mental illness among the general population of adults in the U.S. The group recommended that SAMHSA's National Survey on Drug Use and Health (NSDUH) be modified to produce estimates of SMI among adults[i]. Specifically, it was recommended that the past year Kessler 6 (K6) scale of psychological distress, already in the NSDUH, be supplemented with items on functional impairment. The K6 was chosen because it had already been demonstrated to be a good predictor of SMI in prior studies (Kessler et al, 2003). Since impairment is a component of the SMI definition, a scale measuring impairment was recommended to improve prediction and face validity, which would facilitate acceptance of the estimates by epidemiologists and the public. Furthermore, the group recommended that the data from short scales collected in the NSDUH be used to estimate SMI via a prediction model fit on data from clinical psychiatric interviews administered to a subsample of NSDUH respondents. A subsample was recommended because it was recognized that it was not feasible to administer a full clinical diagnostic interview to all 45,000 adult respondents of NSDUH due to burdens imposed on respondents because of the length of such an instrument and because the NSDUH interviewers are not clinicians.

Given the recommendations from the technical advisory group, in 2007 SAMHSA began the development of a program called the Mental Health Surveillance Study (MHSS). The MHSS included a clinical follow-up interview that was conducted on a subsample of respondents to the NSDUH by clinical interviewers over the telephone. The planned sample size for the clinical follow-up was 1,500 in 2008 and 500 in each subsequent year. The MHSS also involved the addition of impairment items, an abbreviated World Health Organization Disability Assessment Scale (WHODAS; Rehm et al., 1999; Novak, 2007), and other mental health items (e.g., items on suicidal thoughts and behavior) to the NSDUH. Furthermore, the MHSS involved the development of an estimation methodology for SMI using clinical interview data collected in 2008.

In order to produce estimates of SMI in a timely manner, an estimation methodology was developed using the clinical interview data collected in 2008[ii]. Details on the specific methods used to estimate SMI among adults in the U.S. are described later in the manuscript. The methodology, which used data from approximately 750 of the respondents to the clinical interview in 2008, was used to produce annual estimates of SMI for the 2008, 2009, 2010, and 2011 NSDUHs. Using the same estimation methods to produce SMI estimates each year allowed for assessment of trend in SMI over time. That is, using different estimation methods each year based on the small annual clinical samples (only 500 interviews planned each year for 2009 and later) would have resulted in large

changes in prevalence rates due to sampling error, making it difficult to detect true variation in SMI over time. Although SAMHSA used the estimation methods based on the 2008 clinical follow-up data for producing annual estimates of SMI, the clinical data collection continued from 2009 to 2012. SAMHSA decided to use this data to validate the 2008 estimation methods and to only revise these methods if it could be shown that revisions would produce more accurate estimates of SMI. Also, the clinical sample was supplemented in 2011 and 2012 by an extra 1,000 cases each year. This expansion of the sample was funded through an interagency agreement with the National Institute of Mental Health. The interagency agreement also helped support studies that were conducted to determine if the estimation methodology for SMI could be improved.

Using combined data from the 2008 to 2012 MHSS clinical follow-up samples, it was determined that revisions to the 2008 estimation methods were warranted. Specifically, the 2008 estimation methods were demonstrated to produce estimates of SMI that were biased upward, particularly for young adults. As such, all previously released estimates of SMI from 2008 to 2011 were revised and this improved estimation methodology was used to produce SMI estimates in 2012.

The rest of this manuscript provides a general overview of the methods used to produce estimates of SMI among adults using the NSDUH and describes the revisions that were made to the estimation procedures to produce more accurate estimates. Specifically, the following topics are included: a brief description of the NSDUH, a description of the MHSS clinical follow-up interview procedures, the SMI estimation methodology developed from the 2008 clinical sample, revisions to the estimation methodology (referred to as the 2012 methods), comparisons of SMI estimates using the 2008 and 2012 estimation methods, steps taken to release revised estimates, and caveats on the estimation methods.

Description of the NSDUH

The NSDUH is the federal government's primary source of information on the nature and extent of substance use and abuse in the United States. The survey also includes several modules of questions that focus on mental health issues. Conducted since 1971, the survey collects data by administering questionnaires to a representative sample of about 67,500 persons in the United States at their place of residence. The respondent universe is the civilian, noninstitutionalized population aged 12 years old or older residing within the United States. 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 prisons and long-term hospitals. Young people are oversampled, with one-third of the sample in each state allocated to the following three age groups: 12-17, 18-25, and 26 and older. At each sampled address, a 5-minute screening procedure using a handheld computer lists all household members and their basic demographic data. To obtain the target sample sizes, a preprogrammed selection algorithm selects zero to two sample person(s), depending on the composition of the household.

The main interview data are collected through face-to-face computer-assisted interviewing (CAI), including audio computer-assisted self-interviewing (ACASI), on a laptop computer. The interviews average about an hour. Each respondent who completes a full interview is given a $30 cash payment. The questionnaire contains demographic items (which are interviewer-administered) and self-administered questions pertaining to the use of tobacco, alcohol, and illicit drugs, substance dependence and abuse, treatment for substance use problems, health conditions, and mental health (SAMHSA, 2012). Mental health items contained in the main interview data include short scales of psychological distress (the K6), functional impairment due to psychological distress (abbreviated WHODAS), items on major depressive episode (MDE), items on serious thoughts of suicide and suicidal behavior, and items on mental health service use.

MHSS Clinical Follow-Up Interview

For the 2008 to 2012 survey years, at the end of the main NSDUH interview, a nationally representative subsample of NSDUH adult respondents who completed the interview in English was asked to complete a follow-up interview on mental health. Once a selected respondent agreed to participate in the follow-up interview, they were provided with a $30 incentive. NSDUH respondents who agreed to participate in the follow-up interview were contacted by telephone within 2 to 4 weeks and administered a psychiatric diagnostic interview by trained clinical interviewers. The diagnostic instrument for this study was the Structured Clinical Interview for DSM-IV-TR Axis I Disorders Non-Patient Edition (SCID-I/NP). The SCID-I/NP (First et al., 2002) is a semi-structured diagnostic interview that

has been widely used in clinical components of studies such as the NCS-R (Kessler et al., 2004), the National Survey of American Life (Jackson et al., 2004), and the NSDUH substance-use disorders reappraisal study (Jordan, et al., 2008). The interview was modified to assess past 12-month mental health disorders and functioning via telephone. Table 1 indicates the particular mental disorders assessed using the SCID. Disorders used in the operational definition of SMI included past year mood, psychotic, anxiety, eating, impulse-control, and adjustment disorders. Functional impairment due to having a diagnosed mental disorder was assessed via the global assessment of functioning (GAF) scale. Functioning is rated on a scale from 1-100 with scores less than or equal to 50 considered 'serious'. Respondents to the clinical interview were deemed to have SMI if they were diagnosed with at least one of the mental disorders measured in the SCID and had serious functional impairment due to having the disorder. Detailed information on the data collection procedures are found elsewhere (Colpe et al., 2010; Blazei et al., in press)

Table 1. Modules included in the MHSS adapted Structured Clinical Interview for DSM-IV disorders (SCID)

MOOD DISORDERS	PAST YEAR EATING DISORDERS
*Past Year Major Depressive Episode	*Anorexia Nervosa
Lifetime Major Depressive Episode	*Bulimia Nervosa
*Past Year Manic Episode	
Lifetime Manic Episode	PAST YEAR IMPULSE CONTROL DISORDERS
*Past Year Dysthymic Disorder	*Intermittent Explosive Disorder
PAST YEAR PSYCHOTIC DISORDERS	PAST YEAR SUBSTANCE USE DISORDERS
*Psychotic Screen	Alcohol Abuse
	Alcohol Dependence
PAST YEAR ANXIETY DISORDERS	Non-Alcohol Substance Abuse
*Posttraumatic Stress Disorder	Non-Alcohol Substance Dependence
*Panic Disorder with and without Agoraphobia	
*Agoraphobia without History of Panic Disorder	PAST YEAR ADJUSTMENT DISORDERS
*Social Phobia	*Adjustment Disorder
*Specific Phobia	
*Obsessive Compulsive Disorder	GLOBAL ASSESSMENT OF FUNCTIONING
*Generalized Anxiety Disorder	*SMI Functional impairment = GAF $\leq$ 50

2008 MHSS Estimation Methods

Using the clinical interview data collected in 2008, a statistical model and cut point was used to produce estimates of SMI among all adults (Aldworth et al., 2010, Liao et al., in press). Specifically, a weighted logistic regression model was fit on the clinical diagnostic data collected in 2008 from a nationally representative subsample of approximately 750 clinical interview respondents. The dependent variable for the model was whether or not the respondents had a diagnosis of SMI based on the clinical diagnostic interview. The predictor variables were the psychological distress (K6) score and functional impairment score (abbreviated WHODAS) based on scales collected in the main NSDUH interview. The model was used to produce a predicted probability of having SMI for each of the respondents to the clinical interview. A cut point was established among the predicted probabilities such that if adults with probabilities at or above the cut point were predicted to have SMI and the rest were not, the weighted number of false positives (adults not diagnosed to have SMI but predicted to have SMI) would come as close as possible to equaling the weighted number of false negatives (adults diagnosed to have SMI but not predicted to have SMI). If these weighted numbers were exactly equal, then the estimated proportion of adults predicted to have SMI would be the same as the estimated proportion actually diagnosed to have SMI.

Given that the predictor variables in the model were variables collected on the NSDUH main interview, a probability of having SMI could be computed for every NSDUH adult respondent using the estimated model parameters. Applying the cut point (determined from the clinical sample) on the predicted probabilities estimated in the NSDUH adult sample, each NSDUH adult respondent can be classified as having or not having SMI. This dichotomous variable then can be used to compute prevalence estimates of SMI for adults.

2008 Weights

The weights used in the 2008 modeling process were developed using fairly simple methods given the small sample size available in 2008. The weights were the product of four factors: (1) the respondent's NSDUH person-level analysis weight, (2) the inverse of the probability that the respondent was selected for the clinical sample, (3) a simple nonresponse adjustment based on the K6 scores, and (4) a poststratification adjustment that forced the population estimates computed from the clinical sample for some demographic groups to match population totals derived from Census Bureau data.

2008 Model

The following is the fitted 2008 model for the probability of an adult having SMI (π):

$$\text{logit}(\pi) \equiv \log[\pi / (1 - \pi)] = -4.74999920 + 0.20977232\, X_k + 0.38388395\, X_w$$

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

X_k = *Alternative Past Year K6 Score*: Past year K6 score 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 2008 formula for the predicted probability of SMI (SMIPP) can then be expressed using the model parameter estimates above as follows:

$$\text{SMIPP} = 1/(1 + \exp[-(-4.74999920 + 0.20977232 * X_k + 0.38388395 * X_w)]).$$

If SMIPP was greater than or equal to 0. (SMI cut point), then the respondent was predicted as having past year SMI. Although the primary objective was to derive model-based estimates of SMI, an estimate of 'any mental illness' (AMI) also was derived. AMI, defined similarly to SMI with respect to the presence of a diagnosable mental disorder, does not require functional impairment due to the disorder. After assessing a variety of models, the original SMI model was chosen to estimate AMI, using a lower cutpoint of 0.02400.

Revisions to the MHSS Estimation Methods

By the end of 2012, the combined sample from the MHSS collection was approximately 5,500 adults. Using the larger sample, two components of the estimations methods were investigated to determine whether revisions to the 2008 estimation methodology could be made that would increase the accuracy of the SMI estimates. The two components investigated were the model that was fit in the clinical interview data and the weights that were applied to the clinical interview data during the modeling process.

Weight Revisions

As described previously, the weights used in the 2008 modeling were fairly simple. Using the larger sample available and the added resources (time and funding) to continue an investigation into improving the weights, three adjustments were made to better account for coverage and nonresponse biases and to improve statistical efficiency of the estimates (Liao et al., in press).

A potential for coverage bias in the MHSS clinical sample resulted from not including adults who responded to the main NSDUH interview in Spanish. Although the NSDUH main interview includes the opportunity to complete the interview in either English or Spanish, the MHSS sampling excluded those respondents who had completed the NSDUH main interview in Spanish as the MHSS clinical interviews were only conducted in English. To compensate for the potential coverage bias in the SMI estimates due to excluding the NSDUH respondents who completed the main NSDUH interview in Spanish, the weights of Hispanic respondents who were eligible for the clinical follow-up were adjusted to account for those who completed the interview in Spanish. This was done by forcing the total of the adjusted weights among Hispanics to equal the total of the adult NSDUH Hispanic weights, both overall and within categories shown to be related to whether a Hispanic NSDUH respondent completed the main NSDUH interview in English. Variables collected on the NSDUH shown to be associated with whether or not a respondent completes the interview in Spanish included whether or not the respondent was born in the U.S., how many years they lived in the U.S., education level, and age.

Because the MHSS clinical interview sample is a subsample of respondents from the NSDUH interview, data collected from the NSDUH main interview could be used to demonstrate differences between respondents and nonrespondents to the clinical follow-up interview. Adjustments for potential nonresponse biases in the clinical sample were applied in two stages that separately accounted for: (1) *initial nonresponse,* caused by failing to agree to a clinical follow-up after the completion of the NSDUH main interview, and (2) *final nonresponse,* caused by not completing the clinical follow-up after having agreed to participate. Approximately, 15 percent of the selected sample (unweighted) did not agree to a clinical interview. An additional 20 percent of adults selected agreed to but did not ultimately participate in the clinical interview.

NSDUH variables found to be correlated with nonresponse to the MHSS and assumed to be related to key outcome measures of SMI were used to adjust the clinical subsample weighting. That is, the weights first were adjusted for initial nonresponse using variables found to be correlated with initial refusal. The weights then were adjusted for final nonresponse using variables found to be correlated with final nonresponse. Similar to the coverage adjustment, the sum of the adjusted weights among respondents after each stage of nonresponse adjustment was forced to equal the sum of the weight before adjustment among both respondents and nonrespondents.

Table 2 presents an evaluation of the effect of the revised weights on selected mental health estimates. Specifically, the table presents prevalence estimates of the mental health characteristics for the adult NSDUH sample, the MHSS subsample that was selected from the NSDUH, nonrespondents and respondents to the MHSS, respondents using the 2008 weights and respondents using the 2012 weights. In general, those selected for the MHSS are a representative sample of the adult NSDUH sample and therefore should have similar characteristics as the entire adult NSDUH sample. Among the MHSS sample, respondents are different from nonrespondents; respondents are more likely to have indicators of mental health issues than nonrespondents. A comparison of the characteristics when applying the 2008 versus the 2012 weights indicates that use of the 2012 weights over the 2008 weights may better account for nonresponse bias because the estimates for the respondents based on the 2012 weights are closer to the estimates for the adult NSDUH sample compared to the estimates based on the 2008 weights. A more detailed description of the revised weighting procedures can be found elsewhere (Liao, et al., in press).

Table 2. Characteristics by Sample and Response Status among Persons Aged 18 or Older, 2008-2011 Adult NSDUH Main Study and Clinical Sample

	Adult NSDUH Main Sample (1)	Selected MHSS Clinical Sample				
		Selected Sample (2)	Nonrespondents (3)	Respondents before Adjustment (4)	Respondents after Adjustment (2008 Weights) (5)	Respondents after Adjustment (Revised 2012 Weights) (6)
Selected Mental Health Related Variables	Estimate (%)	Estimate (%)	Estimate (%)	Estimate (%)	Estimate (%)	Estimate (%)
Major Depressive Episode (MDE)						
Lifetime MDE	12.8	12.6	9.5[ab]	14.7[c]	14.6[a]	12.9[d]
Past Year MDE	6.6	6.4	5.3[ab]	7.2	7.2	6.6[d]
Serious Thoughts of Suicide in the Past Year	3.7	3.5	2.8[ab]	4.1	3.8	3.7

MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health.
NOTES: Weights used for the prevalence estimates included: Column 1 weight= NSDUH weight, Columns 2-4 MHSS weight prior to the nonresponse adjustments and poststratification; Column 5 weight= 2008 MHSS adjusted weight. Column 6 weight= 2012 MHSS revised weight without scaling over years

a Difference between this estimate (from columns 2-6) and corresponding estimate from full CAI sample (column 1) is statistically significant at $p < 0.10$.

b Difference between this estimate from nonrespondents (column 3) and corresponding estimate from respondents (column 4) using design-based weights is statistically significant at $p < 0.10$.

c Difference between this estimate (column 4) using design-based weights and corresponding estimate from respondents (column 5) using final adjusted weights is statistically significant at $p < 0.10$.

d Difference between this estimate (column 5) using final adjusted weights and corresponding estimate from respondents (column 6) using initial adjusted weights is statistically significant at $p < 0.10$.

A poststratification adjustment also was added as a final step to the weighting procedures to improve statistical efficiency of the SMI estimators. This adjustment forced the sum of the final weights for the MHSS clinical data to equal the sum of the final weights for the adult NSDUH main interview sample for a set of variables that were shown to be associated with SMI. Detailed information on the revised weighting procedures can be found elsewhere (Liao et al., in press).

Model Revisions

Three criteria were used for model selection and to determine whether a revised model would be an improvement over the 2008 model. These were (1) minimization of misclassification, (2) minimization of subpopulation bias, and (3) model parsimony. A revised model was considered an improvement over the 2008 model if it had lower misclassification error, that is, lower rates of the number of false positives and false negatives. An overall misclassification metric or error rate also was established for this comparison which was defined as the weighted sum of the false positives and the false negatives among the clinical follow-up sample. In order to reduce misclassification errors, several covariates that were indicators of mental illness, such as depression and suicidal thoughts, were considered for the 2012 prediction model.

Because a cut point estimator is based on a model, the estimator can be systematically biased if the model is incorrectly specified. The 2008 model cut point was chosen so that for the overall population of adults, the total number of estimated false positives came as close as possible to the total number estimated false negatives. Using this cut point minimized the difference between the direct estimate (model- free estimates) and the predicted estimate of SMI prevalence derived from the clinical interview. Use of the cut point that balanced false negative and false positive rates removed the possibility of bias in the cut point estimator of SMI prevalence for all adults.[iii]

However, SAMHSA produces annual estimates of SMI prevalence not only among all adults but also within subpopulations of interest such as state, age group, gender, and education level. Although a cut point chosen for the overall adult population eliminates the potential for bias in the estimate of SMI for all adults, this cut point does not necessarily produce unbiased prevalence estimates within particular adult subpopulations. Consequently, a bias measure was developed and used to assess whether SMI cut point estimates developed using a particular model were biased within key subpopulations[1]. This bias measure is the difference between the weighted proportions of clinical respondents predicted to have SMI and those actually diagnosed to have SMI within the subpopulation of interest (this is equal to the difference between the false positive rate and the false negative rate in the subpopulation).[iv]

The third criterion for model selection was parsimony. Only a limited number of predictor variables could reasonably be included in a logistic model for SMI, since the clinical data being analyzed derived from a complex survey had, at most 100 effective degrees of freedom (100 variance strata with two variance replicates each). We adhered to a guideline that there should be no more than 10 (i.e.,$\approx\sqrt{100}$) predictor variables in the statistical model used for prediction of SMI. Furthermore, demographic variables used to define key subpopulations for which SMI prevalence were to be estimated, such as gender, race/ethnicity, and employment, were avoided as candidate covariables in the initial model investigation. Adding such a demographic variable as a covariable in the model would tend to fix the relationship between SMI and the subpopulation defined by the particular demographic variable added. This would make it difficult to determine changes in the relationship between SMI and the particular demographic variable included in the model over time. Since SAMHSA produces annual estimate of SMI among all adult and among various subpopulations, demographic variables were only considered for the model if their inclusion produced significant decreases in bias for estimates of SMI within subpopulations specified.

Given these criteria, various models using different combinations of predictor variables were compared to the 2008 model using the revised 2012 weights. A final "best" model that included K6 and WHODAS scores along with items on depression, serious thoughts of suicide and age as covariates was chosen. The 2012 model produced lower misclassification errors and subpopulation level biases than the 2008 model.

The following is the final fitted 2012 model for the probability of adult having SMI (π):

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

where the X_k, X_w, X_s, X_m, and X_a terms are defined as follows:

X_k = *Alternative Past Year K6 Score:* Past year K6 score 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.

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

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

X_a = *AGE1830:* Coded as age minus 18 if aged 18 to 30; coded as 12 otherwise.

[1] The key subpopulations examined for bias were gender, four age groups, four race/ethnicity groups, four regions, three levels of county urbanicity, four levels of employment, four levels of education, three levels of household income, health insurance (yes/no), and whether the adult received mental health services (yes/no).

The 2012 formula for the predicted probability of SMI (SMIPP) can then be expressed using the model parameter estimates as follows:

$$SMIPP = 1/(1 + \exp[-(-5.972664 + 0.0873416*X_k + 0.3385193*X_w + 1.9552664*X_s + 1.1267330*X_m + \quad *X_a)]).$$

If SMIPP was greater than or equal to 0.260573529 (SMI cut point), then the respondent was predicted as having past year SMI. If SMIPP was great than or equal to 0.0192519810 (AMI cut point), then the respondent was predicted as having past year AMI.

Table 3 indicates the misclassification rates (false positive, false negative, and error rates) and bias measure for the 2008 model and the 2012 model among all adults with the 2008 model refit using the 2008-2012 sample and revised 2012 weights. For both the 2008 and 2012 models the SMI estimate among all adults is 3.89 percent. Using the 2012 model decreases the misclassification error for SMI compared to the 2008 model. Both models produce a fairly unbiased estimate of SMI since the cut point for both the 2008 and 2012 estimation methods was chosen such that the estimated numbers of false positives and false negatives were equal among all adults.

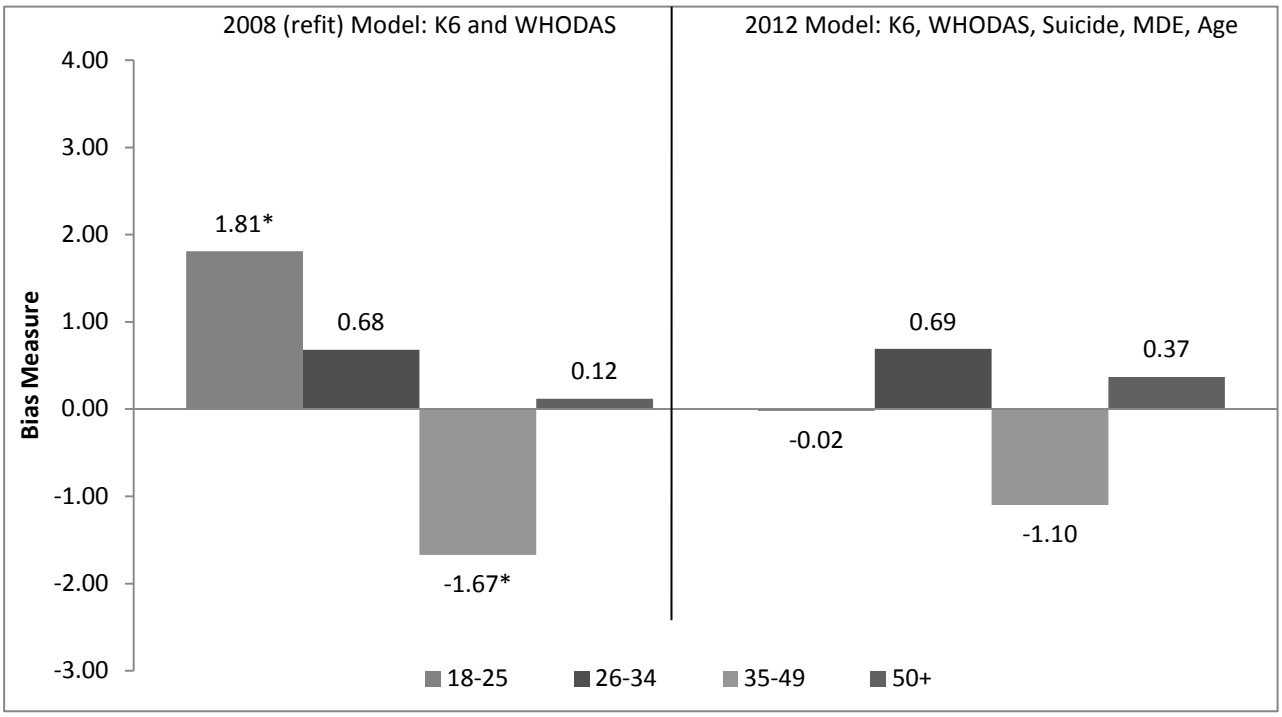
Table 3. Misclassification Statistics for SMI Estimates Computed with Different Models:2008a-2012

Model	NSDUH Main Survey Cut Point Estimate	False Rate		Error Rate	Bias: FP/FN
		Positive (FP)	Negative (FN)		
2008 (refit) model	3.89	2.21	2.21	4.42	1.00
2012 model	3.89	1.92	1.93	3.84	0.99

NOTE: The 2008 model was refit on the 2008-2012 MHSS data using the revised 2012 weights
NOTE: The false positive, false negative, and error rates were calculated using the clinical subsample weights that were scaled when combined across years for efficiency (Liao et al., in press).

Figure 1 displays the bias measures for SMI by age group calculated from the combined 2008-2012 clinical interview sample (n=5,500) for the 2008 and the 2012 models. That is, the 2008 model was refit on the 2008-2012 MHSS data using the revised 2012 weights and compared to the 2012 model. As demonstrated by Figure 1, the 2012 model which includes an age covariate in the model, produces less biased estimates of SMI by age group than the (refit) 2008 model which does not include an age covariate. This decrease in the bias measure is most prominent for the 18 to 25 year old age group. Moreover, the bias in the 35 to 49 year old group is no longer significant at the 0.05 level.

Figure 1. Bias Measure for Estimates of SMI by Age Groups and Model: 2008a-2012 Clinical Interview Sample

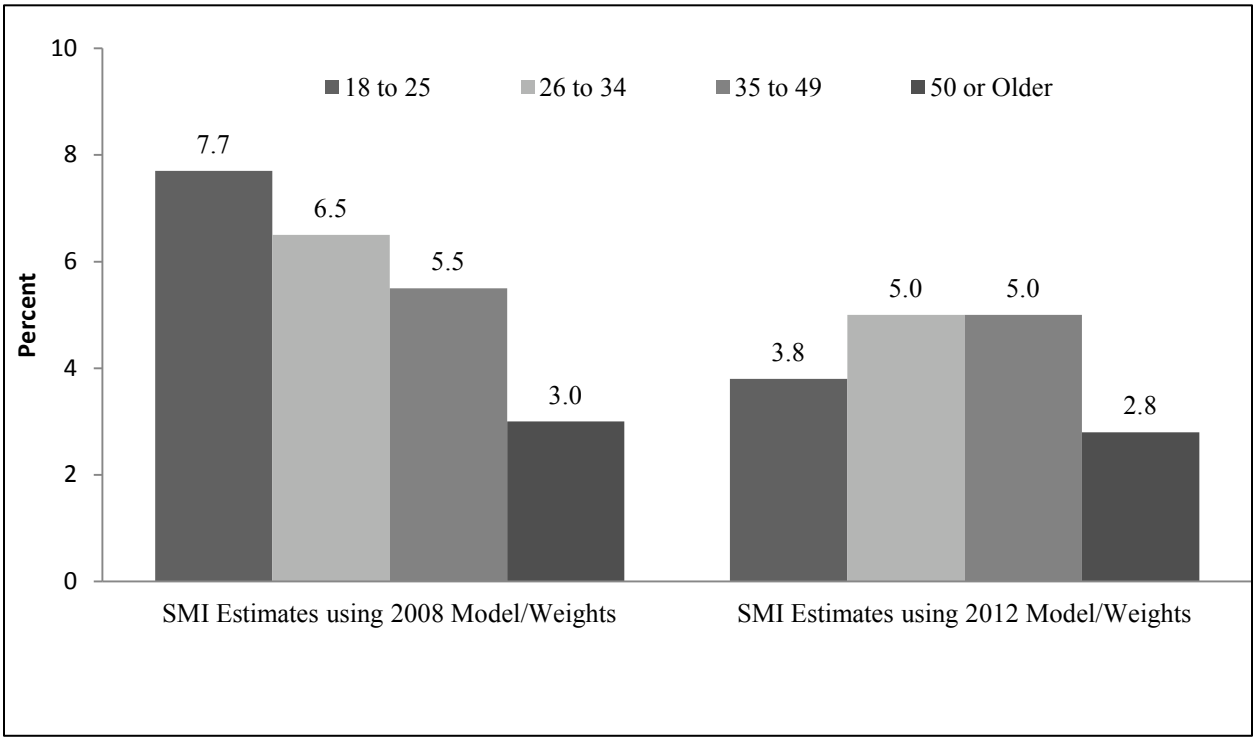


NOTE: The 2008 model was refit on the 2008-2012 MHSS data using the revised 2012 weights
NOTE: Bias measures are computed from the clinical subsample using 2012 weights that were scaled when combined across years for efficiency. See Chapter 5 of Liao et al. (in press) for details.
*p<0.05

Comparisons of 2012 NSDUH SMI Estimates using the 2008 and 2012 Estimation Methods

Comparisons between the direct estimates from the MHSS clinical interview sample and the model based estimates from the NSDUH adult sample were made to determine whether use of the 2012 estimation methods reduced biases in estimates of SMI among all adults and adult subpopulations. Figure 2 displays the SMI estimates by age group computed using the 2008 and 2012 estimation methods over the combined 2008 to 2012 NSDUH main sample. As shown in the figure, the SMI estimates using the 2012 estimation methods are lower than the SMI estimates computed using the 2008 estimation methods; this is mainly due to the revision in the weighting which better accounted for the fact that nonrespondents to the clinical interview component of the survey were less likely to have mental health problems. Also, the distribution of the SMI estimates across the four age groups differs between the methods. That is, using the 2008 estimation method, the 18 to 25 year old age group had the highest prevalence of SMI; using the 2012 estimation method the 26 to 49 year olds have the highest prevalence of SMI. Comparisons of SMI estimates by type of estimation method also were made for various other subpopulation including, race/ethnicity, gender, employment status, education level, poverty status, health insurance status. In most instances, use of the 2012 estimation method produces less biased SMI estimates compared to estimates computed using the 2008 estimation method. A more detailed discussion of these comparisons may be found in previously published documents (Kott et al. 2013).

Figure 2. Estimates of Serious Mental Illness (SMI) in Past Year among Persons Aged 18 or Older, by Age Group and Estimation Method: NSDUH 2008-2012



NOTE: Unlike estimates in pervious tables and figures, the estimates in this table include data from a NSDUH half sample in 2008 that were assigned questions based on an alternative set of functional impairment questions rather than the WHODAS. See Liao et al. (in press) for more details.

Releasing Revised Estimates of SMI

Once it was determined that revising the estimation methods would produce more accurate estimates of SMI, a plan for releasing the revised SMI estimates was formed and implemented. The plan included communication strategies to inform key stakeholders of the revisions and to help gain public acceptance of the revisions, documentation of the revisions, release of the revised estimates to the public, revisions to previously published estimates, and revisions to data files.

In order to inform key stakeholders and the public of the revisions to the methodology several conference presentations were made on the revised estimation procedures. These presentations also served to provide feedback on the methodology from external methodologists. Presentation included those at the American Public Health Association, the Eastern North American Region of the International Biometric Society, the Joint Statistical Meetings, and the Federal Committee on Statistical Methods. Feedback on the revisions was further obtained by contracting with several mental health experts to independently review the estimation methods and to determine whether the revised methods produced more accurate estimates of SMI. Briefing presentations on the revisions to the estimation methodology that included feedback received by the mental health experts were given to inform key stakeholders of the changes in the estimation methodology for SMI.

In addition to communicating the changes in the estimation methodology via presentations, several manuscripts were prepared on the estimation methodology for submission to survey methods journals. Complete documentation of the revisions included the preparation of a detailed report on the design and estimation procedures of the MHSS from 2008 to 2012.

The first release of the revised estimates included a brief, nontechnical report presenting the 2011 revised SMI estimates (SAMHSA, 2013). This report was accompanied by a short methodological report that described the technical aspects of the revisions to the estimation methods (Kott et al., 2013). Shortly after this report was released,

NSDUH's 2012 Mental Health Detailed Tables and Mental Health Findings report was released and included revised estimates of SMI (SAMHSA, 2013).

Early in the process of commencing the estimation methodology work, it was determined that if a revision to the SMI estimation method was warranted, previously released estimates of SMI would need to be revised in order to assess trends over time. This included annual estimates of SMI published from 2008 to 2011. Since 2008, SMI estimates were published by SAMHSA in the form of annual detailed tables, an annual mental health findings report, state-level estimates, and various topical reports. The decision on which products to revise was prioritized based on cost and the likelihood that the product was widely utilized. As such, it was determined that all detailed tables that included estimates of SMI from 2008 to 2011 would be revised and would replace previously published tables. Previously published mental health findings reports from 2008 to 2011 would not be revised but would include a disclaimer indicating that revised estimates of mental illness that are more accurate are available in the mental health detailed tables. Reports and tables that contained state-level estimates of SMI also were revised. Any analytic report that had not yet been released was revised to include the more accurate estimates of SMI and previously published analytic reports that contained estimates of SMI were not revised but contained a disclaimer indicating that the estimates had been revised. Furthermore, any data file that contained SMI variables were revised.

Caveats on the Methods

Various caveats on the methods from producing model-based estimates of SMI should be noted. Although effort went into adjusting for the potential biases in those adults responding to the clinical interview, there is no guarantee that the adjustments for nonresponse to the clinical and the undercoverage of Hispanics who chose to respond to the main survey in Spanish were completely successful. There is likewise no guarantee that the nonresponse and coverage adjustments in the NSDUH were completely successful in removing all of the biases.

In addition, the mental illness estimates were based on a weighted logistic model for SMI. Although statistical tests did not uncover significant biases in the SMI prevalence estimates within most of the key subpopulations of interest that does not guarantee that other subpopulation level biases do not exist. Although many of the subpopulation level estimates commonly produced were investigated, there are many more possible subpopulations for which SMI estimates may be computed given the vast number of variables available in the NSDUH data. Also, it should be noted that SMI prevalence estimates for a subpopulation closely related to a variable used in the SMI model is likely to be biased. Such variables include suicidal thoughts, the experience of major depressive episode (MDE), and the various components of the Kessler-6 or WHODAS scales. An ongoing research study is evaluating several alternate models that do not use the suicidal thoughts and the experience of MDE as predictors and may produce unbiased estimates for the corresponding subpopulations.

Another caveat concerns the standard errors of the SMI estimates. By treating model-predicted SMI indicators as true values of SMI when estimating the standard error of a cut point estimate, standard errors are underestimated. This underestimation is due to not accounting for the error in model fitting. Nevertheless, standard errors calculated this way, can be useful when estimating the difference in SMI prevalence between subpopulations or differences in estimates over time because the same model fit is used for both subpopulations and therefore the error due to model fitting when estimating differences is effectively cancelled out.

Finally, the MHSS clinical data collection ended in 2012, and the 2012 SMI model will be applied to NSDUH data going forward to produce mental illness estimates under the assumption that the relationship between SMI and the predictor variables is stable over time. Without continuing the clinical interview, the validity of this assumption cannot be assessed, and one cannot test whether covariates other than those specified in the 2012 model may be needed in the future to reduce misclassification or subpopulation-level biases.

References

Aldworth J., Colpe L. J., Gfroerer J. C., Novak S. P., Chromy J. R., Barker P. R., Barnett-Walker K., Karg R. S., Morton K. B., & Spagnola K. (2010). The National Survey on Drug Use and Health Mental Health Surveillance Study: calibration analysis. *International Journal of Methods in Psychiatric Research, 19*(Suppl.1), 61-87.

Blazei, R., Coffey, L. E., Gordon, R., Granger, R., Karg, R., Panzer, K., & Shook-Sa, B. (2014). Mental Health Surveillance Study operations report (2008-2012). In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 16a, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Deliverable No. 41, RTI/0212800.001.112.001). Research Triangle Park, NC: RTI International.

Colpe L. J., Barker P. R., Karg R. S., Batts K. R., Morton K. B., Gfroerer J. C., Stolzenberg S. J., Cunningham D. B., First M. B., & Aldworth J. (2010). The National Survey on Drug Use and Health Mental Health Surveillance Study: calibration study design and field procedures. *International Journal of Methods in Psychiatric Research, 19*(Suppl.1), 36-48.

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

Jackson, J. S., Torres, M., Caldwell, C. H., Neighbors, H. W., Nesse, R. M., Taylor, R. J., Trierweiler, S. J., & Williams, D. R. (2004). The National Survey of American Life: A study of racial, ethnic and cultural influences on mental disorders and mental health. *International Journal of Methods in Psychiatric Research, 13*, 196-207.

Jordan, B. K., Karg, R. S., Batts, K. R., Epstein, J. F., & Wiesen, C. A. (2008). A Clinical Validation of the National Survey on Drug Use and Health Assessment of Substance Use Disorders. *Addictive Behaviors, 33*, 782-798.

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.

Kessler, R. C., Abelson, J., Demler, O., Escobar, J. I., Gibbon, M., Guyer, M. E., Howes, M. J., Jin, R., Vega, W. A., Walters, E. E., Wang, P., Zaslavsky, A., & Zheng, H. (2004). Clinical Calibration of DSM-IV Diagnoses in the World Mental Health (WMH) Version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMH-CIDI). *International Journal of Methods in Psychiatric Research, 13*, 122–139.

Kott, P., Hedden, S., Aldworth, J., Bose, J., Chromy, J., Gfroerer, J., & Liao, D. (2013, October). *2012 National Survey on Drug Use and Health: A revised strategy for estimating the prevalence of mental illness* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, RTI 0212800.002.120.008.002.004). Research Triangle Park, NC: RTI International.

Liao, D., Kott, P., Aldworth, J., Yu, F., Karg, R., Shook-Sa, & Davis, T. (in press). *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.

Novak, S. P., Colpe, L. J., Barker, P. R., & Gfroerer, J. C. (2010). Development of a Brief Mental Health Impairment Scale using a Nationally Representative Sample in the USA. *International Journal of Methods in Psychiatric Research, 19*(Suppl.1), 49-60.

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.

Substance Abuse and Mental Health Services Administration, Center for Mental Health Services. (1993, May 20). Final notice [Final definitions for: (1) Children with a serious emotional disturbance, and (2) adults with a serious mental illness]. *Federal Register, 58*(96), 29422-29425.

Substance Abuse and Mental Health Services Administration (2013). Results from the 2012 National Survey on Drug Use and Health: Mental Health Findings. NSDUH Series H-47, DHHS Publication No. SMA 13-4805. Center for Behavioral Health Statistics and Quality: Rockville, MD.

Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. (November 19, 2013). *The NSDUH Report: Revised Estimates of Mental Illness from the National Survey on Drug Use and Health*. Rockville, MD.

---

[i] Although the primary objective was to derive model-based estimates of SMI, an estimate of "any mental illness" (AMI) also was of interest. AMI, defined similarly to SMI with respect to the presence of a diagnosable mental disorder, does not require functional impairment from the disorder. After assessing a variety of models, the SMI model was chosen to estimate AMI using a lower cutpoint. Use of the SMI model to produce estimates of AMI by using a lower cut point was implemented for both the 2008 and 2012 estimation methods.

[ii] In 2008, the NSDUH main interview adult sample was partitioned into two half samples to test alternative impairment measures: the abbreviated WHODAS and SDS (Sheehan Disability Scale). MHSS clinical subsamples were selected from both half-samples and yielded approximately 750 respondents for each impairment measure (WHODAS sample [sample 'a'] and SDS sample [sample 'b']).An initial objective of the 2008 clinical sample was to select one of the two impairment measures for inclusion in future years' surveys. The investigation in 2008 led to selection of a model with a WHODAS scale rather than an SDS scale. For more information on the investigations incorporating the SDS half sample see Liao et al. (in press). Tables and figures in this report that incorporated the WHODAS half sample and excluded the SDS half sample are denoted as 2008a.

[iii] Assuming both the NSDUH and clinical sample are weighted properly, the estimated proportion of adults predicted to have SMI based on the 2008 NSDUH and the proportion based on the 2008 clinical sample are nearly unbiased estimators for the same target. Because the estimated proportion of predicted positives based on the clinical sample will approximately equal the estimated proportion actually diagnosed with SMI based on the clinical sample and the latter is an unbiased estimator for the actual population proportion having SMI, the cut point estimator based on the NSDUH sample is also a nearly unbiased estimator. Mathematically, let y_k , c_k, w_k , and ω_k be the actual SMI diagnosis (1 for yes, 0 for no), the model-based SMI prediction, the NSDUH weight, and the clinical sample weight, respectively, for respondent k. Then under probability sampling theory, $\Sigma w_k c_k/\Sigma w_k$ and $\Sigma \omega_k c_k/\Sigma \omega_k$ are nearly unbiased estimators of the same parameter. The SMI cut point, which was chosen to equalize the false positives and false negatives, forces $\Sigma \omega_k c_k/\Sigma \omega_k \approx \Sigma \omega_k y_k/\Sigma \omega_k$. Furthermore, $\Sigma \omega_k y_k/\Sigma \omega_k$ is a nearly unbiased estimator for the proportion of the adult population who would be diagnosed with SMI using the clinical interview.

[iv] Mathematically, the bias measure is $B = \Sigma_D \omega_k(c_k - y_k)/\Sigma_D \omega_k$, where Σ_D denotes summation over a subpopulation of interest. The measure B can be viewed as a simple weighted mean, and its standard error is computed by taking the sample design into consideration. The ratio of the bias measure over the estimated standard error is asymptotically standard normal under the null hypothesis of no bias at the subpopulation level. This ratio (and the normality assumption) was used in testing for bias at the subpopulation level.