

# **NATIONAL SURVEY ON DRUG USE AND HEALTH: STATISTICAL MODELS TO PREDICT MENTAL ILLNESS FROM 2005 TO 2007**

## **DISCLAIMER**

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

Substance Abuse and Mental Health Services Administration  
Center for Behavioral Health Statistics and Quality  
Rockville, Maryland 20857

May 2017

*This page intentionally left blank*

# **NATIONAL SURVEY ON DRUG USE AND HEALTH: STATISTICAL MODELS TO PREDICT MENTAL ILLNESS FROM 2005 TO 2007**

Contract No. HHSS283201000003C  
RTI Project No. 0213984.100.001.022

**RTI Authors:**

Jeremy Aldworth  
Phillip Kott

**SAMHSA Authors:**

Jonaki Bose  
Joseph Gfroerer\*  
Sarrah Hedden  
Arthur Hughes\*

**Project Director:**

David Hunter

**SAMHSA Project Officer:**

Peter Tice

\* Retired from SAMHSA.

For questions about this report, please e-mail [Peter.Tice@samhsa.hhs.gov](mailto: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

May 2017

Recommended Citation: Center for Behavioral Health Statistics and Quality.  
(2017). *National Survey on Drug Use and Health: Statistical Models to  
Predict Mental Illness from 2005 to 2007*. Substance Abuse and Mental  
Health Services Administration, Rockville, MD.

## **Acknowledgments**

This publication was prepared with the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality (CBHSQ), by RTI International (a registered trademark and a trade name of Research Triangle Institute). At RTI, the report was edited by Richard S. Straw and formatted by Debbie Bond. Two of the SAMHSA authors, Joseph Gfroerer and Arthur Hughes, have retired from SAMHSA.

# Table of Contents

Chapter	Page
List of Figures .....	v
List of Tables .....	vii
1. Introduction.....	1
1.1 Background .....	2
1.2 Objective and Content of Report.....	2
2. Estimating SMI and AMI from the 2012 SMI Prediction Model.....	5
3. Alternative Models for Estimating SMI and AMI.....	7
3.1 Background .....	7
3.2 SMIPP_U as Potential Response Variable for Alternative Models .....	10
3.3 Alternative Models for Predicting SMI.....	11
3.3.1 Candidate SMI Models .....	11
3.3.2 ROC Analysis of SMI Candidate Models.....	14
3.3.3 Bias of SMI Candidate Models, by Domain.....	16
3.3.4 Estimates of SMI Derived from Model S4 and SMIYR_U, by Domain: 2005 to 2014 .....	17
3.4 Alternative Models for Predicting AMI .....	20
3.4.1 Candidate AMI Models.....	20
3.4.2 ROC Analysis of AMI Candidate Models .....	34
3.4.3 Bias of AMI Candidate Models, by Domain .....	36
4. Summary and Conclusions .....	39
4.1 SMI Estimation.....	39
4.2 AMI Estimation.....	40
4.3 Concluding Remarks .....	40
References.....	43

## Appendix

A	Predictor Variables Used in Alternative Models .....	A-1
B	Plots of SMI Estimates Derived from Model S4 and SMIYR_U, by Domain .....	B-1

*This page intentionally left blank*

## List of Figures

Figure	Page
B.1 SMI Estimates Derived from Model S4 and SMIYR_U: 2005 to 2014 .....	B-3
B.2 SMI Estimates Derived from Model S4 and SMIYR_U, by Gender: 2005 to 2014 .....	B-4
B.3 SMI Estimates Derived from Model S4 and SMIYR_U, by Age Group: 2005 to 2014.....	B-5
B.4 SMI Estimates Derived from Model S4 and SMIYR_U, by Race/Ethnicity (Five Largest Groups): 2005 to 2014 .....	B-6
B.5 SMI Estimates Derived from Model S4 and SMIYR_U, by Race/Ethnicity (Three Smallest Groups): 2005 to 2014 .....	B-7
B.6 SMI Estimates Derived from Model S4 and SMIYR_U, by Region: 2005 to 2014 .....	B-8
B.7 SMI Estimates Derived from Model S4 and SMIYR_U, by County Type (Large and Small Metro Areas): 2005 to 2014.....	B-9
B.8 SMI Estimates Derived from Model S4 and SMIYR_U, by County Type (Nonmetro Areas): 2005 to 2014 .....	B-10
B.9 SMI Estimates Derived from Model S4 and SMIYR_U, by Receipt of Mental Health Services: 2005 to 2014 .....	B-11
B.10 SMI Estimates Derived from Model S4 and SMIYR_U, by Employment Status: 2005 to 2014 .....	B-12
B.11 SMI Estimates Derived from Model S4 and SMIYR_U, by Education Level: 2005 to 2014 .....	B-13
B.12 SMI Estimates Derived from Model S4 and SMIYR_U, by Poverty Level: 2005 to 2014 .....	B-14
B.13 SMI Estimates Derived from Model S4 and SMIYR_U, by Health Insurance Coverage: 2005 to 2014 .....	B-15

*This page intentionally left blank*



## List of Tables

Table	Page
3.1 List of Predictor Variables Used, with Brief Descriptions .....	10
3.2 Four Alternative Candidate Models for Predicting SMI.....	11
3.3 Model S1 (20 <i>DF</i> ): Predictor Variables, <i>DF</i> , and Wald <i>F</i> and Beta Statistics .....	12
3.4 Model S2 (15 <i>DF</i> ): Predictor Variables, <i>DF</i> , and Wald <i>F</i> and Beta Statistics .....	13
3.5 Model S3 (11 <i>DF</i> ): Predictor Variables, <i>DF</i> , and Wald <i>F</i> and Beta Statistics .....	13
3.6 Model S4 (6 <i>DF</i> ): Predictor Variables, <i>DF</i> , and Wald <i>F</i> and Beta Statistics .....	14
3.7 ROC Statistics of SMI Cut Point Estimates Based on Four Candidate Models .....	15
3.8 Bias of SMI Cut Point Estimates of Four Candidate Models, by Domain: 2008A-2012.....	18
3.9a Estimates of SMI Based on Model S4 and SMIYR_U, by Domain: 2005-2010 .....	21
3.9b Estimates of SMI Based on Model S4 and SMIYR_U, by Domain: 2011-2014 .....	23
3.10a Standard Errors of Estimates of SMI Based on Model S4 and SMIYR_U, by Domain: 2005-2010 .....	25
3.10b Standard Errors of Estimates of SMI Based on Model S4 and SMIYR_U, by Domain: 2011-2014 .....	27
3.11 ROC Statistics of AMI Cut Point Estimates Based on Four SMI Candidate Models.....	29
3.12 Four Alternative Candidate Models for Predicting AMI.....	30
3.13 Model A1 (22 <i>DF</i> ): Predictor Variables, <i>DF</i> , and Wald <i>F</i> and Beta Statistics.....	31
3.14 Model A2 (21 <i>DF</i> ): Predictor Variables, <i>DF</i> , and Wald <i>F</i> and Beta Statistics .....	32
3.15 Model A3 (20 <i>DF</i> ): Predictor Variables, <i>DF</i> , and Wald <i>F</i> and Beta Statistics .....	33
3.16 Model A4 (7 <i>DF</i> ): Predictor Variables, <i>DF</i> , and Wald <i>F</i> and Beta Statistics .....	34
3.17 ROC Statistics of AMI Cut Point Estimates Based on Four Candidate Models .....	35
3.18 Bias of AMI Cut Point Estimates of Four Candidate Models by Domain: 2008A-2012.....	37
A.1 Description of Predictor Variables Used in Alternative Models: Predictor Variables Closely Related to Those of 2012 Model .....	A-3
A.2 Description of Predictor Variables Used in Alternative Models: Other Predictor Variables .....	A-5

*This page intentionally left blank*

# 1. Introduction

In 2008, the Substance Abuse and Mental Health Services Administration (SAMHSA) implemented a new program to produce national and state estimates of the prevalence of mental illness. The Mental Health Surveillance Study (MHSS) was conducted as part of the National Survey on Drug Use and Health (NSDUH) from 2008 to 2012 for the primary purpose of developing models that estimate the prevalence of serious mental illness (SMI) and any mental illness (AMI) in the adult (18 or older) civilian, noninstitutionalized population of the United States. The MHSS consisted of two major components:

- a new questionnaire module administered to all adult respondents that captured data on psychological distress and impairment caused by mental health issues, and
- a subsample of adults selected from the main NSDUH study for follow-up clinical interviews.

Information from the clinical interviews was used to determine the "gold-standard" measure of SMI status ("yes" or "no") for each subsample respondent. The clinical interview data were linked with the data obtained from the new module to develop a prediction model to estimate SMI and AMI for the full NSDUH sample (Center for Behavioral Health Statistics and Quality [CBHSQ], 2014). A prediction model was developed that included gold-standard SMI status (obtained from the clinical interviews) as the response variable. The predictor variables included variables such as psychological distress and impairment measures that were collected in the main NSDUH questionnaire for adults. The resulting prediction model developed from the MHSS clinical data was then applied to all adult NSDUH respondents to obtain the predicted probability of SMI for each respondent. Next, a cut point was determined so that if the predicted probability of SMI for a respondent met or exceeded the cut point, then he or she was *predicted* to be SMI positive; otherwise, he or she was predicted to be SMI negative. That is, a dichotomy of *SMI status* was computed. Receiver operating characteristic (ROC) analysis was applied to determine the cut point that resulted in the weighted number of false-positive and false-negative counts being (at least approximately) equal, thus ensuring unbiased estimates. Using the same model, a cut point for producing unbiased estimates of AMI was also determined.

SAMHSA has produced estimates of SMI and AMI each year since 2008 using models developed with 2008 to 2012 clinical interview data, along with data from the main survey mental health module that has remained in NSDUH. Unfortunately, these models could not be applied to the 2007 and earlier NSDUH data because some of the predictors in the 2008 to 2012 models were not collected before 2008. However, the 2005 to 2007 NSDUHs did collect some of the necessary predictors, along with other mental health variables that could potentially be good predictors of SMI and AMI. Given the analytic benefits of having comparable indicators of SMI and AMI for additional years of NSDUH data, this report explores whether new models could be developed from the MHSS data to generate 2005 to 2007 estimates of SMI and AMI that would be comparable with the 2008 and later estimates. This report also describes the process and specific models evaluated and chosen for use in creating new mental illness indicator variables for 2005 to 2007.

## 1.1 Background

In 2008, the first year in which the MHSS was conducted, all adult respondents to the main NSDUH interview received a six-item scale (Kessler-6 or K6) that captured information on psychological distress in the past 12 months (Kessler et al., 2003). Although it had been determined that an indicator of impairment was needed for the predictive model, there was not a consensus about which indicator was best. Thus, two alternative impairment scales were tested. They were assigned according to a split-sample design as follows:

- A random half of the sample (Sample A) received an abbreviated eight-item version of the World Health Organization Disability Assessment Schedule (WHODAS; Novak, Colpe, Barker, & Gfroerer, 2010; Rehm et al., 1999).
- The other half (Sample B) received the four-item Sheehan Disability Scale (SDS; Leon, Olfson, Portera, Farber, & Sheehan, 1997).

For the subsample of approximately 1,500 adult NSDUH participants who completed the follow-up MHSS clinical interview, the randomization of the impairment scales was maintained, so that about half of the MHSS clinical sample participants were administered the WHODAS and half were administered the SDS (i.e., there were approximately 750 completed interviews from each half sample).<sup>1</sup>

Based on an analysis of the 2008 MHSS data, it was determined that the WHODAS was a better predictor of SMI and would therefore be administered as the sole impairment scale in subsequent NSDUHs (CBHSQ, 2009). Using only the 750 clinical interview cases from the WHODAS sample, an SMI prediction model was constructed (CBHSQ, 2009) that contained only two predictor variables: (1) a variable based on the K6, and (2) a variable based on the WHODAS. ROC analysis determined the appropriate cut points for estimating SMI and AMI. This model was initially used as the basis of SMI and AMI estimates for 2008 and subsequent years.

The sample size for the clinical interview subsample was approximately 500 in 2009 and 2010 and approximately 1,500 in 2011 and 2012. By 2012, the combined 2008A to 2012 clinical sample included approximately 5,000 respondents. Based on this much larger sample, the SMI prediction model was revised and improved. As a result of the revision, the 2012 SMI prediction model added to the two terms already in the 2008 model three predictor variables related to age, past year major depressive episode (MDE), and suicidal thoughts in the past year. Analyses indicated that some of these added terms reduced bias at the domain (i.e., subpopulation) level of various demographic and geographic variables, while others reduced the overall error rate (i.e., the sum of false positives and false negatives expressed as rates).

## 1.2 Objective and Content of Report

The objective of this methodological study is to develop and assess potential predictive models that can be applied to the 2005 to 2007 NSDUH data and generate estimates of SMI and AMI that are consistent with the 2008 and later estimates. These new models will include only

---

<sup>1</sup> Samples A and B of 2008 are henceforward referred to as 2008A and 2008B, respectively.

predictors that were collected consistently from 2005 to 2012. If successful, the study will make it possible for analysts to examine trends in mental illness from 2005 through 2015 and later, as well as conduct pooled analyses that will include the 2005 to 2007 data.

This report is organized as follows. The 2012 SMI prediction model and method for determining estimates of SMI and AMI are briefly discussed in Chapter 2. Alternative models for determining estimates of SMI and AMI based on variables contained in all NSDUH years from 2005 to 2012 are described, compared, and assessed in Chapter 3. Concluding remarks and guidance to analysts on the use of these results are provided in Chapter 4. Two appendices provide the predictor variables used in the alternative models (Appendix A) and plots of SMI estimates derived from the models, by domain (Appendix B).

*This page intentionally left blank*

## 2. Estimating SMI and AMI from the 2012 SMI Prediction Model

The 2012 serious mental illness (SMI) prediction model is used to provide national estimates of SMI and any mental illness (AMI) based on data from the 2008A (i.e., Sample A of 2008) to 2012 National Surveys on Drug Use and Health (NSDUHs) for adults aged 18 or older who were assigned to the World Health Organization Disability Assessment Schedule (WHODAS) questions (Center for Behavioral Health Statistics and Quality [CBHSQ], 2014).<sup>2</sup> Specifically, let  $\pi$  represent the probability that an adult has SMI. Then the 2012 model can be expressed as follows:

$$\text{logit}(\hat{\pi}) \equiv \log\left[\frac{\hat{\pi}}{1-\hat{\pi}}\right] = -5.9726644 + 0.0873416 * X_k + 0.3385193 * X_w + 1.9552664 * X_s + 1.1267330 * X_m + 0.1059137 * X_a,$$

where  $\hat{\pi}$  is the estimated probability that an adult had SMI, and the predictor variables  $X_k$ ,  $X_w$ ,  $X_s$ ,  $X_m$ , and  $X_a$  are defined as follows:

$X_k$  = *Alternative Past Year Kessler-6 (K6) Score*: Past year K6 score of less than 8 recoded as 0; past year K6 score of 8 to 24 recoded as 1 to 17.

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

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

$X_m$  = *Past Year Major Depressive Episode (MDE) on NSDUH*: Coded as 1 if the criteria for past year MDE were met;<sup>3</sup> coded as 0 otherwise.

$X_a$  = *Adjusted Age (i.e., AGE1830)*: Coded as age minus 18 if aged 18 to 30; coded as 12 otherwise.

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

$$\text{SMIPP\_U} = 1 / \left( 1 + \exp \left[ - \left( -5.9726644 + 0.0873416 * X_k + 0.3385193 * X_w + 1.9552664 * X_s + 1.1267330 * X_m + 0.1059137 * X_a \right) \right] \right).$$

---

<sup>2</sup> A separate model was developed for respondents assigned to the Sheehan Disability Scale (SDS) in 2008B (i.e., Sample B of 2008).

<sup>3</sup> See Section B.4.5 of CBHSQ (2015a) for a detailed discussion of the past year MDE criteria.

If SMIPP\_U was greater than or equal to 0.2605735292 (the SMI cut point), then the respondent was predicted as having past year SMI (i.e., SMIYR\_U = 1); otherwise, the respondent was predicted as not having past year SMI (SMIYR\_U = 0). If SMIPP\_U was greater than or equal to 0.0192519810 (the AMI cut point), then the respondent was predicted as having past year AMI (AMIYR\_U = 1); otherwise, the respondent was predicted as not having past year AMI (AMIYR\_U = 0).<sup>4</sup>

---

<sup>4</sup> Before the 2012 SMI prediction model had been developed, SMI predicted probabilities and SMI and AMI cut point estimates were produced from the 2008 SMI model, and the corresponding variables were called SMIPP, SMIYR, and AMIYR, respectively. These variables are not comparable with their analogues produced from the 2012 SMI model (i.e., the variables SMIPP\_U, SMIYR\_U, and AMIYR\_U).



## 3. Alternative Models for Estimating SMI and AMI

### 3.1 Background

As noted in Section 1.2 in Chapter 1, the objective of this study is to develop and assess potential predictive models that can be applied to the 2005 to 2007 National Survey on Drug Use and Health (NSDUH) data and generate estimates of serious mental illness (SMI) and any mental illness (AMI) that are consistent with the 2008 and later estimates. These new models will include only predictors that were collected consistently from 2005 to 2012. This chapter describes these new models, including the response variables that were considered for these new models, and which of them proved to be more suitable than others.

Chapter 2 includes a discussion of how the 2012 model for providing national estimates of SMI and AMI was constructed using data from the Mental Health Surveillance Study (MHSS) clinical study. The *response variable* of the 2012 model consisted of gold-standard SMI diagnoses based on information collected in the clinical study from 2008A (i.e., Sample A of 2008) to 2012, and the *predictor variables* were derived from information collected in the 2008A to 2012 NSDUHs.<sup>5</sup>

However, for the alternative models, the goal is to generate estimates of SMI and AMI that are consistent (i.e., comparable with the 2008 and later estimates *based on* SMIYR\_U and AMIYR\_U, respectively) from the main NSDUH data.<sup>6</sup> To achieve this, SMIYR\_U, AMIYR\_U, and their parent variable SMIPP\_U were considered as response variables for these new models. Under this scenario, the available sample size based on 2008A to 2012 data to investigate the new models increased from approximately 5,000 (MHSS clinical data) to approximately 200,000 (adult NSDUH data).

Trend measurements in SMI and AMI can be made beyond 2012 because the predictor-variable information required for the 2012 model has been collected in 2013 and later surveys. However, the 2012 model cannot be used to estimate SMI and AMI before 2008 because not all of the required predictor-variable information was collected in those earlier surveys. Therefore, to approximate SMI and AMI trends before 2008, alternative models need to be constructed from suitable predictor-variable information available in NSDUHs before 2008. Because of changes to the survey questionnaire in 2004, this investigation is limited to developing alternative models from 2005 to 2007 that can be used in trend analyses in combination with the 2008 to 2012 data. Therefore, the alternative models need to include *predictor variables* that can be constructed from information common to all NSDUHs from 2005 to 2012.

---

<sup>5</sup> To simplify the alternative model development and assessment process, the 2008B (i.e., Sample B of 2008) data containing information from the Sheehan Disability Scale (SDS) rather than the World Health Organization Disability Assessment Schedule (WHODAS) is excluded from further analysis.

<sup>6</sup> See Chapter 2 for a description of SMIYR\_U and AMIYR\_U, noting that each of these variables was derived from the parent variable SMIPP\_U, which is also described in Chapter 2.

A list of potential predictor variables was compiled using as a starting point variables that are identical (or whose 2005 to 2007 versions are at least comparable) to the variables in the 2012 model that are also available in the 2005 to 2007 NSDUH data. Three such variables were identified:

1. WSPDSC2: recoded alternative past year Kessler 6 (K6) score (2005 to 2007 version comparable with  $X_k$  in the 2012 model),
2. AMDEYR\_U: recoded major depressive episode (MDE) in the past year (2005 to 2007 version comparable with  $X_m$  in the 2012 model), and
3. AGE1830: recoded age ( $X_a$  in the 2012 model).

Because of changes to the questionnaire in 2008, there were breaks in trends, and the WHODAS and MDE variables listed above that occur in the 2005 to 2007 NSDUHs are not fully *comparable* with those in 2008 and onward. However, adjustments were applied to the 2005 to 2007 variables so that these adjusted variables are assumed to be *comparable* with the corresponding variables in the 2008A to 2012 NSDUHs.

The changes in the WHODAS and MDE variables listed above are described as follows:

- In 2005 to 2007, respondents were asked the six K6 questions for only one reference period—the one month in the past year when they were at their worst emotionally. In 2008, the same questions were asked, but the initial reference period was in the past month. Respondents were then asked if there was a month in the past year that was worse than the past month. If they responded "yes," then the questions were asked again with the worst month in the past year as the reference period. The maximum score from the two reference periods was then used as the past year K6 total score. The changes to the reference period of the K6 questions in 2008 appeared to have caused significant effects with respect to K6 scores.
- The adult depression section of the questionnaire did not change in 2008, but changes to the mental health questions immediately preceding the MDE questions in 2008 appeared to have caused significant context effects<sup>7</sup> with respect to estimates of lifetime and past year MDE between the two time periods of 2005 to 2007 and 2008 to 2009 (the latest year studied). In addition, significant effects were detected in the MDE estimates between the WHODAS and SDS half samples of 2008 (i.e., the 2008A and 2008B samples, respectively).

As a result of the changes in the questionnaire in 2008, adjustments were applied to (1) the 2005 to 2007 version of the past year K6 score variable, and (2) the 2005 to 2007 and 2008B version of the past year MDE variable. The adjustments to the variables were made to the 2005 to 2008A data to make them comparable with data from 2008B, 2009 data, and onward. Another NSDUH methodological report (Center for Behavioral Health Statistics and Quality

---

<sup>7</sup> Changes to questions between survey years can affect how respondents answer questions in subsequent modules (i.e., "context effects"). A context effect may be said to take place when the response to a question is affected by information that is not part of the question itself.

[CBHSQ], 2012) provides further details about the comparability adjustments that were applied to these variables and the reasons that such adjustments were required.

Two variables in the 2005 to 2007 NSDUHs were assumed to measure similar concepts as the remaining two predictor variables in the 2012 model:

1. ADSDSSC3: recoded alternative SDS<sup>8</sup> score specifically associated with depression and embedded in the adult depression section of the questionnaire (some similarity with  $X_w$  in the 2012 model), and
2. ADWRSTHK\_U: recoded serious thoughts of suicide in the past year specifically associated with depression and embedded in the adult depression section of the questionnaire (similar to  $X_s$  in the 2012 model).

Tests of association were conducted between these two variables and their "similar" counterparts in the 2012 model using the weighted 2008A to 2012 data. The following results were obtained: (1) noting that the alternative SDS score has five values (0 to 4) and  $X_w$  has nine values (0 to 8), Pearson's correlation coefficient was 0.50, which indicates a moderate level of association; and (2) noting that the two versions of serious thoughts of suicide in the past year are binary variables, Pearson's correlation coefficient was 0.43, which also indicates a moderate level of association (in addition, the simple kappa coefficient was 0.43, indicating a moderate level of agreement). In both cases, chi square tests of association were significant at the .0001 level of significance, but this may not be surprising given the large sample size.

Additional variables under consideration included suitable key mental health variables identified in Appendix I of the 2012 public use file and codebook (CBHSQ, 2013).

Many of the potential predictor variables have missing values; consequently, recodes had to be constructed to allow records with missing values to be available for modeling analyses. The complete set of candidate variables was reduced to a smaller subset of variables that were actually used in the various models investigated in this analysis. This subset of variables is listed and briefly described in [Table 3.1](#); more details are provided in Appendix A, including details on the construction of recodes where applicable. [Table A.1](#) in Appendix A describes the subset of five variables listed above that were identical to, comparable with, or at least similar to the predictor variables in the 2012 model, and [Table A.2](#) describes all of the other variables.

---

<sup>8</sup> The SDS was not directly associated with the 2012 (WHODAS) model, but a predictor variable based on information from the SDS was used in the 2008B model (CBHSQ, 2009). This predictor variable acted analogously to the predictor variable based on information from the WHODAS in the 2012 model (i.e., they both captured information about the severity of impairment).

**Table 3.1 List of Predictor Variables Used, with Brief Descriptions**

<b>Predictor Variable</b>	<b>Brief Description</b>
WSPDSC2	Alternative worst K6 total score (2005 to 2007 version comparable with $X_k$ in 2012 model)
AMDEYR_U	Adult PY MDE (2005 to 2007 version comparable with $X_m$ in 2012 model)
AGE1830	Age recode used in 2012 model ( $X_a$ in 2012 model)
ADSDSSC3	Alternative SDS score from adult depression questionnaire section (similar to $X_w$ in 2012 model)
ADWRSTHK_U	PY serious thoughts of suicide from adult depression questionnaire section (similar to $X_s$ in 2012 model)
AMDELT_U	Adult LT MDE
AMHINP2_U	Received inpatient mental health treatment in PY
AMHRX2_U	Received prescribed medications for mental health treatment in PY
MHLMNT_U	Received outpatient mental health treatment at mental health clinic/center in PY
MHLTHER_U	Received outpatient mental health treatment at private therapist office in PY
MHRCOST2_U	No mental health treatment in PY because could not afford cost
MHRHAND2_U	No mental health treatment in PY because thought could handle problem without treatment
MHRWHER2_U	No mental health treatment in PY because did not know where to go
MHRENUF2_U	No mental health treatment in PY because insurance would not pay enough
MHRNMBRS2_U	No mental health treatment in PY because of fear of neighbors negative opinion
MHRCMIT2_U	No mental health treatment in PY because of fear of being committed or medicated
ANXDLIF_U	Had anxiety in LT from recoded pregnancy and health care questionnaire section
DEPRSLIF_U	Had depression in LT from recoded pregnancy and health care questionnaire section
DEPRSYR_U	Had depression in PY from recoded pregnancy and health care questionnaire section
WRKEDYR_U	Worked at any job in PY
JBSTAT14	Work situation in past week
HEALTH_U	Overall health in core demographics questionnaire section (1 = excellent to 5 = poor; 0 = missing recode)

K6 = Kessler 6; LT = lifetime; MDE = major depressive episode; PY = past year; SDS = Sheehan Disability Scale.

NOTE: See Appendix A for details on the predictor variables, including the construction of recodes, where applicable.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2005 to 2012.

### 3.2 SMIPP\_U as Potential Response Variable for Alternative Models

As shown in Chapter 2, the variable SMIPP\_U represents the predicted probability of having SMI based on the 2012 model. This variable can be considered as a continuous variable with values ranging between 0 and 1 and hence contains more information than the dichotomized variables SMIYR\_U and AMIYR\_U that were derived from SMIPP\_U to indicate SMI and AMI status ("yes" or "no"), respectively. For this reason, SMIPP\_U was considered first as a potential response variable for the development of alternative models.

Because SMIPP\_U was treated as a continuous variable, a linear regression model was used. Predictions from this model would then be dichotomized by an appropriate cut point to provide a predictor of SMI (or AMI) status.

The use of linear regression models makes it possible for predicted values to fall below 0 or above 1. However, after dichotomization, this did not present problems in practice because all predicted values less than the cut point (including negative values) were simply transformed to 0s, and all other predicted values (including those greater than 1) were transformed to 1s.

A more serious problem arose that did not appear to have an obvious solution. In the 2012 logistic regression model, the age term in the model served to reduce bias within the age group domains. However, it was found that the inclusion of this age term in a linear regression model adjusted only the *means* of the predicted values within those age groups and not the part of the distribution in the neighborhood of the cut point (i.e., approximately the 96th percentile of the weighted distribution). Therefore, inclusion of the age term had no effect in reducing bias within those age groups. In addition, it was subsequently found that models with SMIYR\_U (or AMIYR\_U) as the response variable resulted in lower estimates of bias and total error rate relative to similar models with SMIPP\_U as the response variable. Consequently, the use of SMIPP\_U as a potential response variable for the alternative models was dropped.

### 3.3 Alternative Models for Predicting SMI

#### 3.3.1 Candidate SMI Models

Alternative logistic regression models for predicting SMI were examined by using SMIYR\_U (i.e., the dichotomous variable of having or not having SMI as the response variable) and the predictor variables discussed in Section 3.1 and described in Table 3.1. Four candidate models emerged from a process where an increasingly stringent criterion for including predictor variables was used. Table 3.2 displays the predictor variable inclusion criterion and the corresponding numerator degrees of freedom (*DF*) of each of the four candidate models.<sup>9</sup>

**Table 3.2 Four Alternative Candidate Models for Predicting SMI**

Model	<i>DF</i>	Predictor Variable Inclusion Criterion
S1	20	Wald <i>F</i> test <i>p</i> value < .05
S2	15	Wald <i>F</i> test <i>p</i> value < .01
S3	11	Wald <i>F</i> test statistic > 10
S4	6	Wald <i>F</i> test statistic > 100

*DF* = degrees of freedom; SMI = serious mental illness.

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

Tables 3.3 to 3.6 provide details of the four models listed in Table 3.2, namely, the predictor variable, *DF*, Wald *F* test statistic, Wald *F* test *p* value, beta, beta standard error (SE), and beta *p* value. The tables indicate that all four candidate models included the three predictor variables that were identical to or comparable with those from the 2012 model, plus the two predictor variables with similarity to other variables in the 2012 model (see Section 3.1 for details). Perhaps not surprisingly, these five predictor variables have the largest Wald *F* statistics in each of the models, which is indicative of greater predictive power.

<sup>9</sup> For all of the models, the denominator *DF* = 900.

**Table 3.3 Model S1 (20 *DF*): Predictor Variables, *DF*, and Wald *F* and Beta Statistics**

Predictor Variable	<i>DF</i>	Wald <i>F</i> Statistic	Wald <i>F</i> <i>p</i> Value	Beta	Beta SE	Beta <i>p</i> Value
Intercept				-8.70	0.137	0.000
WSPDSC2	1	2,460.3	0.000	0.33	0.007	0.000
AMDEYR_U	1	663.9	0.000	2.10	0.082	0.000
AGE1830	1	458.8	0.000	0.17	0.008	0.000
ADSDSSC3	1	248.9	0.000	0.46	0.029	0.000
ADWRSTHK_U	1	474.2	0.000	1.61	0.074	0.000
AMHINP2_U	1	6.4	0.012	0.45	0.177	0.012
AMHRX2_U	1	6.8	0.010	0.22	0.086	0.010
MHLTHER_U	1	10.4	0.001	0.34	0.104	0.001
MHRCOST2_U	1	30.5	0.000	0.53	0.095	0.000
MHRHAND2_U	1	8.9	0.003	0.38	0.129	0.003
MHRCMIT2_U	1	26.3	0.000	1.05	0.204	0.000
ANXDLIF_U	1	6.6	0.010	0.20	0.077	0.010
DEPRSYR_U	1	22.0	0.000	0.42	0.090	0.000
WRKEDYR_U	1	6.7	0.010	0.26	0.102	0.010
JBSTAT14	1	9.3	0.002	0.37	0.122	0.002
HEALTH_U	5	2.5	0.029			
Level 0				-1.08	0.476	0.023
Level 2				0.07	0.098	0.448
Level 3				0.21	0.105	0.041
Level 4				0.18	0.118	0.128
Level 5				0.20	0.181	0.268

*DF* = degrees of freedom; SE = standard error.

NOTE: See [Table 3.1](#) for brief descriptions of predictor variables, and Appendix A for further details.

NOTE: The variable HEALTH\_U has six levels (1 = excellent to 5 = poor; 0 = missing recode); Level 1 is the reference level in the model.

Response variable = SMIYR\_U.

Dataset = 2008A to 2012 adult NSDUH data.

Analysis weight = ANALWT\_A/5.

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

**Table 3.4 Model S2 (15 *DF*): Predictor Variables, *DF*, and Wald *F* and Beta Statistics**

<b>Predictor Variable</b>	<b><i>DF</i></b>	<b>Wald <i>F</i> Statistic</b>	<b>Wald <i>F</i> <i>p</i> Value</b>	<b>Beta</b>	<b>Beta SE</b>	<b>Beta <i>p</i> Value</b>
Intercept				-8.61	0.119	0.000
WSPDSC2	1	2,483.2	0.000	0.33	0.007	0.000
AMDEYR_U	1	667.7	0.000	2.11	0.082	0.000
AGE1830	1	463.8	0.000	0.17	0.008	0.000
ADSDSSC3	1	254.4	0.000	0.46	0.029	0.000
ADWRSTHK_U	1	478.4	0.000	1.61	0.074	0.000
AMHINP2_U	1	7.0	0.008	0.47	0.177	0.008
AMHRX2_U	1	6.9	0.009	0.22	0.086	0.009
MHLTHER_U	1	9.6	0.002	0.32	0.103	0.002
MHRCOST2_U	1	31.7	0.000	0.53	0.095	0.000
MHRHAND2_U	1	8.6	0.003	0.38	0.128	0.003
MHRCMIT2_U	1	26.9	0.000	1.06	0.204	0.000
ANXDLIF_U	1	7.0	0.008	0.20	0.077	0.008
DEPRSYR_U	1	23.0	0.000	0.43	0.090	0.000
WRKEDYR_U	1	6.8	0.009	0.26	0.101	0.009
JBSTAT14	1	15.8	0.000	0.42	0.105	0.000

*DF* = degrees of freedom; SE = standard error.

NOTE: See Table 3.1 for brief descriptions of predictor variables, and Appendix A for further details.

Response variable = SMIYR\_U.

Dataset = 2008A to 2012 adult NSDUH data.

Analysis weight = ANALWT\_A/5.

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

**Table 3.5 Model S3 (11 *DF*): Predictor Variables, *DF*, and Wald *F* and Beta Statistics**

<b>Predictor Variable</b>	<b><i>DF</i></b>	<b>Wald <i>F</i> Statistic</b>	<b>Wald <i>F</i> <i>p</i> Value</b>	<b>Beta</b>	<b>Beta SE</b>	<b>Beta <i>p</i> Value</b>
Intercept				-8.54	0.116	0.000
WSPDSC2	1	2,598.1	0.000	0.34	0.007	0.000
AMDEYR_U	1	695.5	0.000	2.12	0.081	0.000
AGE1830	1	477.9	0.000	0.17	0.008	0.000
ADSDSSC3	1	262.0	0.000	0.47	0.029	0.000
ADWRSTHK_U	1	482.3	0.000	1.62	0.074	0.000
MHLTHER_U	1	12.7	0.000	0.37	0.103	0.000
MHRCOST2_U	1	33.4	0.000	0.55	0.096	0.000
MHRCMIT2_U	1	29.9	0.000	1.11	0.204	0.000
ANXDLIF_U	1	12.9	0.000	0.28	0.077	0.000
DEPRSYR_U	1	44.8	0.000	0.53	0.080	0.000
JBSTAT14	1	17.4	0.000	0.44	0.105	0.000

*DF* = degrees of freedom; SE = standard error.

NOTE: See Table 3.1 for brief descriptions of predictor variables, and Appendix A for further details.

Response variable = SMIYR\_U.

Dataset = 2008A to 2012 adult NSDUH data.

Analysis weight = ANALWT\_A/5.

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

**Table 3.6 Model S4 (6 *DF*): Predictor Variables, *DF*, and Wald *F* and Beta Statistics**

Predictor Variable	<i>DF</i>	Wald <i>F</i> Statistic	Wald <i>F</i> <i>p</i> Value	Beta	Beta SE	Beta <i>p</i> Value
Intercept				-8.48	0.114	0.000
WSPDSC2	1	2,935.6	0.000	0.34	0.006	0.000
AMDEYR_U	1	692.1	0.000	2.13	0.081	0.000
AGE1830	1	553.8	0.000	0.17	0.007	0.000
ADSDSSC3	1	317.5	0.000	0.51	0.028	0.000
ADWRSTHK_U	1	541.0	0.000	1.69	0.073	0.000
DEPRSYR_U	1	115.1	0.000	0.77	0.072	0.000

*DF* = degrees of freedom; SE = standard error.

NOTE: See [Table 3.1](#) for brief descriptions of predictor variables, and Appendix A for further details.

Response variable = SMIYR\_U.

Dataset = 2008A to 2012 adult NSDUH data.

Analysis weight = ANALWT\_A/5.

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

### 3.3.2 ROC Analysis of SMI Candidate Models

The four models described in [Tables 3.3 to 3.6](#) were run using the 2008A to 2012 adult NSDUH data and the appropriate analysis weight divided by the number of years represented (i.e., ANALWT\_A/5), yielding a predicted probability of SMI for each respondent and model.

Predicted probabilities from a candidate model were dichotomized into a cut point estimator of SMI as follows. A cut point was established among the predicted values such that if adults with values at or above the cut point were predicted to have SMI and the rest were not, the weighted number of false positives (i.e., adults for whom SMIYR\_U = 0 but predicted to have SMI) would come as close as possible to equaling the weighted number of false negatives (adults for whom SMIYR\_U = 1 but not predicted to have SMI). The cut point estimation method relies on being able to identify a cut point that equalizes the weighted false-positive and false-negative rates (or counts) to provide unbiased estimates of SMI. But the variable nature of unequal weights makes it almost impossible to achieve exact equality between the two rates, thereby resulting in some bias. However, for a model that yields a sufficient number of distinct, realizable predicted probabilities, it is typically not difficult to minimize the bias to a reasonable extent.

Receiver operating characteristic (ROC) statistics of the cut point estimators associated with the four models are displayed in [Table 3.7](#). The following ROC statistics are included in the table:

- false-positive rate, which is the proportion of all respondents who were predicted to have SMI but for whom SMIYR\_U = 0;
- false-negative rate, which is the proportion of all respondents who were predicted not to have SMI but for whom SMIYR\_U = 1;
- bias, which is the difference between the false-positive and false-negative rates;
- total error rate, which is the sum of the false-positive and false-negative rates;



**Table 3.7 ROC Statistics of SMI Cut Point Estimates Based on Four Candidate Models**

Model	Sample Size	Pop Size (1,000s)	Cut Point	SMIYR_U Estimate	Cut Point Estimate	SE	Bias	False-Positive Rate	False-Negative Rate	Total Error Rate	Sensitivity	Specificity	AUC
S1	206,510	229,830	0.377876	3.89	3.88	0.049	−0.0026	0.94	0.94	1.88	0.758	0.990	0.874
S2	206,510	229,830	0.377621	3.89	3.89	0.049	−0.0001	0.93	0.93	1.87	0.759	0.990	0.875
S3	206,510	229,830	0.371252	3.89	3.89	0.049	0.0008	0.94	0.94	1.88	0.758	0.990	0.874
S4	206,510	229,830	0.372556	3.89	3.89	0.048	0.0000	0.93	0.93	1.87	0.760	0.990	0.875

AUC = area under curve (i.e., average of sensitivity and specificity); Pop = population; ROC = receiver operating characteristic; SE = standard error of cut point estimate; SMI = serious mental illness.

NOTE: Models S1 to S4 are described in Section 3.3.1.

NOTE: Bias = false-positive rate – false-negative rate.

NOTE: Total error rate = false-positive rate + false-negative rate.

Response variable = SMIYR\_U.

Dataset = 2008A to 2012 adult NSDUH data.

Analysis weight = ANALWT\_A/5.

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

- sensitivity, which is the proportion of respondents for whom  $\text{SMIYR\_U} = 1$  who were predicted to have SMI;
- specificity, which is the proportion of respondents for whom  $\text{SMIYR\_U} = 0$  who were predicted not to have SMI; and
- area under the curve (AUC) (i.e., the ROC curve), which is based on predicted responses dichotomized by the cut point, which is the average of the sensitivity and specificity.

The ROC statistics displayed in [Table 3.7](#) indicate that all four models showed levels of bias similarly close to zero and similarly low levels of total error rate. However, the most parsimonious model (i.e., Model S4 with 6 *DF*) performed very slightly better than the other three models with respect to bias and total error rate. If bias, total error rate, and model parsimony were the only criteria used to select the "best" model, then Model S4 would be selected. However, the models need to be assessed for performance in terms of bias across a specified set of domains. Therefore, all four models were retained to see how they compare with respect to domain-level bias (i.e., the bias at the subpopulation level).

### 3.3.3 Bias of SMI Candidate Models, by Domain

As noted in Section 3.3.2, the cut point estimation method relies on being able to identify a cut point that at least approximately equalizes the weighted false-positive and false-negative rates to provide unbiased estimates of SMI, where bias is defined as the difference between the cut point estimate of SMI and the estimate of SMI based on  $\text{SMIYR\_U}$ . Even if this can be achieved, there is no guarantee that the same cut point will provide approximately unbiased estimates for various demographic and geographic domains. In fact, domain-level bias was a major consideration in the development of the 2012 prediction model, which led to the addition of the age variable in the model (CBHSQ, 2014).

Domain-level bias was assessed for all four candidate models described in [Tables 3.3](#) to [3.6](#) across the following demographic and geographic domains:

- gender (male, female);
- age (18 to 25, 26 to 34, 35 to 49, 50 or older);
- race/ethnicity (non-Hispanic, non-Hispanic white, non-Hispanic black, non-Hispanic American Indian or Alaska Native, non-Hispanic Native Hawaiian or Other Pacific Islander, non-Hispanic Asian, non-Hispanic two or more races, Hispanic);
- region (Northeast, Midwest, South, West);
- county type (large metro, small metro, small metro [250,000 to 1 million population], small metro [less than 250,000 population], nonmetro, nonmetro [urbanized], nonmetro [less urbanized], nonmetro [completely rural]);
- received mental health services ("yes," "no");
- employment (full time, part time, unemployed, other);

- education (less than high school, high school graduate, some college, college graduate);
- poverty level (less than 100 percent, at least 100 percent but less than 200 percent, at least 200 percent); and
- health insurance<sup>10</sup> (private, Medicaid/CHIP,<sup>11</sup> other,<sup>12</sup> no coverage).

The domain-level bias results for the four candidate models are displayed in [Table 3.8](#). For all four candidate models, [Table 3.8](#) indicates that the "yes" and "no" domains of the received mental health services variable showed significant bias at the .01 level. In addition, among the four models a scattering of other domains indicated significant bias at the .05 level. Model S4 had the smallest number of domains indicating significant bias.

Therefore, because of the better overall and domain-level bias, lower total error rate, and model parsimony exhibited by Model S4, it was selected for further evaluation. In the next section, Model S4's national estimates of SMI from 2005 to 2014 are compared with those derived from SMIYR\_U from 2008 to 2014.

### 3.3.4 Estimates of SMI Derived from Model S4 and SMIYR\_U, by Domain: 2005 to 2014

As noted in Section 3.3.3, Model S4 was selected from among the four candidate models for further evaluation. National estimates of SMI from 2005 to 2014 from the model were compared with corresponding estimates derived from SMIYR\_U from 2008 to 2014.

National estimates of SMI were obtained from Model S4 as follows. For each adult respondent from the 2005 to 2014 NSDUHs, his or her responses to each of the six predictor variables in Model S4 were applied to obtain a predicted probability of SMI. If the predicted probability was greater than or equal to the cut point of Model S4 shown in [Table 3.7](#) (i.e., 0.372556), then that respondent was predicted as having past year SMI; otherwise, that respondent was predicted as not having past year SMI. Then, for each year from 2005 to 2014, national estimates of the prevalence of SMI were derived from the weighted average of SMI status ("yes" or "no") predictions.

Note that the first two variables in Model S4, WSPDSC2 and AMDEYR\_U, are identical to the terms  $X_w$  and  $X_a$  in the 2012 model for some years of NSDUH, but for other years adjusted versions that were comparable with them had to be used. Specifically, an adjusted version of the alternative past year K6 score was required for the 2005 to 2007 NSDUHs, and an adjusted version of past year MDE was required for the 2005 to 2007 and 2008B NSDUHs. For details, see CBHSQ (2012).

---

<sup>10</sup> Respondents could indicate multiple types of health insurance; thus, these response categories are not mutually exclusive.

<sup>11</sup> CHIP (Children's Health Insurance Program) is for individuals aged 19 or younger.

<sup>12</sup> Other health insurance is defined as having Medicare, Civilian Health and Medical Program of the Uniformed Services (CHAMPUS), TRICARE, Civilian Health and Medical Program of the Veterans Administration (CHAMPVA), the Department of Veterans Affairs (VA), military health care, or any other type of health insurance.

**Table 3.8 Bias of SMI Cut Point Estimates of Four Candidate Models, by Domain: 2008A to 2012**

Demographic Variable	SMIYR_U Estimate	Model S1 (20 <i>DF</i> )			Model S2 (16 <i>DF</i> )			Model S3 (11 <i>DF</i> )			Model S4 (6 <i>DF</i> )		
		Estimate	Bias	SE Bias	Estimate	Bias	SE Bias	Estimate	Bias	SE Bias	Estimate	Bias	SE Bias
<b>Total</b>	3.89	3.88	0.00	0.05	3.89	0.00	0.05	3.89	0.00	0.05	3.89	0.00	0.05
<b>Gender</b>													
Male	2.89	2.86	-0.03	0.06	2.86	-0.04	0.06	2.85	-0.05	0.06	2.85	-0.05	0.06
Female	4.81	4.83	0.02	0.08	4.85	0.03	0.08	4.86	0.04	0.08	4.86	0.04	0.08
<b>Age in Years</b>													
18 to 25	3.86	3.96	0.10	0.06	3.98	0.12	0.06	<b>4.01<sup>a</sup></b>	<b>0.15</b>	<b>0.06</b>	<b>4.02<sup>a</sup></b>	<b>0.16</b>	<b>0.06</b>
26 to 34	4.98	4.88	-0.10	0.11	4.88	-0.10	0.11	4.91	-0.07	0.11	4.87	-0.11	0.11
35 to 49	5.01	5.00	-0.01	0.10	5.01	0.00	0.10	5.03	0.02	0.10	5.00	-0.01	0.10
50 or Older	2.77	2.77	0.00	0.09	2.76	-0.01	0.09	2.73	-0.04	0.09	2.76	0.00	0.08
<b>Race/Ethnicity</b>													
Not Hispanic	4.01	4.01	0.00	0.05	4.02	0.01	0.05	4.01	0.00	0.05	4.00	-0.01	0.05
White	4.25	4.27	0.03	0.06	4.29	0.04	0.06	4.28	0.03	0.06	4.26	0.01	0.06
Black	3.09	3.11	0.02	0.13	3.09	0.00	0.13	3.13	0.04	0.13	3.10	0.01	0.12
American Indian or Alaska Native	6.64	6.25	-0.38	0.56	6.04	-0.59	0.58	6.04	-0.59	0.58	6.25	-0.39	0.50
Native Hawaiian or Other Pacific Islander	2.39	1.73	-0.66	0.56	1.66	-0.73	0.56	1.73	-0.66	0.56	2.07	-0.32	0.67
Asian	2.19	<b>1.74<sup>a</sup></b>	<b>-0.45</b>	<b>0.23</b>	<b>1.73<sup>a</sup></b>	<b>-0.45</b>	<b>0.22</b>	<b>1.74<sup>a</sup></b>	<b>-0.45</b>	<b>0.22</b>	1.87	-0.32	0.23
Two or More Races	5.73	6.37	0.64	0.36	6.27	0.54	0.35	6.20	0.47	0.37	5.76	0.03	0.39
Hispanic	3.14	3.10	-0.04	0.13	3.09	-0.05	0.13	3.12	-0.02	0.14	3.20	0.06	0.14
<b>Region</b>													
Northeast	3.56	3.66	0.10	0.10	3.65	0.09	0.10	3.63	0.07	0.11	3.65	0.09	0.10
Midwest	4.24	4.18	-0.06	0.08	4.19	-0.05	0.08	4.20	-0.05	0.08	4.23	-0.02	0.08
South	3.71	3.82	0.10	0.09	3.81	0.10	0.09	3.82	0.11	0.09	3.82	0.11	0.09
West	4.09	3.89	-0.20	0.11	3.91	-0.18	0.11	3.90	-0.18	0.11	<b>3.86<sup>a</sup></b>	<b>-0.23</b>	<b>0.11</b>
<b>County Type</b>													
Large Metro	3.62	3.57	-0.05	0.06	3.58	-0.04	0.06	3.55	-0.07	0.06	3.58	-0.04	0.06
Small Metro	4.15	4.15	-0.01	0.08	4.15	0.00	0.08	4.17	0.01	0.08	4.13	-0.02	0.08
250K to 1M	4.12	4.13	0.01	0.10	4.14	0.02	0.10	4.15	0.04	0.10	4.11	0.00	0.10
Less than 250K	4.23	4.18	-0.05	0.15	4.18	-0.06	0.15	4.19	-0.04	0.15	4.18	-0.05	0.15
Nonmetro	4.26	4.43	0.18	0.15	4.41	0.15	0.15	4.45	0.20	0.15	4.43	0.17	0.15
Urbanized	4.49	4.65	0.16	0.20	4.60	0.11	0.20	4.66	0.17	0.20	4.60	0.11	0.20
Less Urbanized	4.00	4.30	0.30	0.24	4.29	0.29	0.24	4.32	0.33	0.24	4.30	0.30	0.24
Completely Rural	4.49	4.15	-0.34	0.30	4.17	-0.32	0.31	4.20	-0.29	0.31	4.31	-0.18	0.29

See notes at end of table.

(continued)

**Table 3.8 Bias of SMI Cut Point Estimates of Four Candidate Models, by Domain: 2008A to 2012 (continued)**

Demographic Variable	SMIYR_U Estimate	Model S1 (20 <i>DF</i> )			Model S2 (16 <i>DF</i> )			Model S3 (11 <i>DF</i> )			Model S4 (6 <i>DF</i> )		
		Estimate	Bias	SE Bias	Estimate	Bias	SE Bias	Estimate	Bias	SE Bias	Estimate	Bias	SE Bias
<b>Received Mental Health Services</b>													
Yes	18.54	<b>19.96<sup>b</sup></b>	<b>1.42</b>	<b>0.27</b>	<b>19.99<sup>b</sup></b>	<b>1.45</b>	<b>0.27</b>	<b>19.73<sup>b</sup></b>	<b>1.19</b>	<b>0.27</b>	<b>19.42<sup>b</sup></b>	<b>0.88</b>	<b>0.27</b>
No	1.55	<b>1.32<sup>b</sup></b>	<b>-0.23</b>	<b>0.04</b>	<b>1.32<sup>b</sup></b>	<b>-0.23</b>	<b>0.04</b>	<b>1.36<sup>b</sup></b>	<b>-0.19</b>	<b>0.04</b>	<b>.41<sup>b</sup></b>	<b>-0.14</b>	<b>0.04</b>
<b>Employment</b>													
Full Time	2.79	2.68	-0.11	0.06	<b>2.67<sup>a</sup></b>	<b>-0.11</b>	<b>0.06</b>	2.69	-0.10	0.05	2.79	0.00	0.06
Part Time	4.05	4.01	-0.03	0.16	4.05	0.00	0.16	4.07	0.02	0.16	4.16	0.12	0.16
Unemployed	6.42	6.48	0.06	0.21	6.50	0.07	0.20	6.34	-0.08	0.21	6.29	-0.13	0.21
Other	5.22	5.41	0.19	0.10	5.40	0.18	0.10	5.40	0.18	0.10	5.18	-0.04	0.10
<b>Education</b>													
< High School	4.01	<b>4.33<sup>a</sup></b>	<b>0.32</b>	<b>0.14</b>	<b>4.29<sup>a</sup></b>	<b>0.28</b>	<b>0.14</b>	<b>4.30<sup>a</sup></b>	<b>0.29</b>	<b>0.14</b>	4.20	0.19	0.13
High School Grad	4.14	4.04	-0.10	0.09	4.03	-0.11	0.09	4.05	-0.09	0.09	4.02	-0.12	0.09
Some College	4.44	4.48	0.04	0.10	4.50	0.06	0.10	4.48	0.04	0.10	4.47	0.03	0.10
College Grad	3.05	2.95	-0.10	0.09	2.97	-0.08	0.08	2.97	-0.08	0.08	3.06	0.01	0.09
<b>Poverty Level<sup>1</sup></b>													
< 100% Threshold	6.96	7.17	0.21	0.16	7.14	0.17	0.16	7.19	0.22	0.16	7.02	0.05	0.16
100% to 199% Threshold	4.77	4.98	0.22	0.13	4.99	0.23	0.13	4.99	0.22	0.14	4.87	0.10	0.13
≥ 200% Threshold	3.01	<b>2.91<sup>a</sup></b>	<b>-0.11</b>	<b>0.05</b>	2.91	-0.10	0.05	<b>2.91<sup>a</sup></b>	<b>-0.11</b>	<b>0.05</b>	2.97	-0.04	0.05
<b>Health Insurance<sup>2</sup></b>													
Private	2.79	<b>2.68<sup>a</sup></b>	<b>-0.10</b>	<b>0.05</b>	2.70	-0.09	0.05	<b>2.69<sup>a</sup></b>	<b>-0.10</b>	<b>0.05</b>	2.75	-0.04	0.05
Medicaid/CHIP <sup>3</sup>	8.49	<b>9.00<sup>a</sup></b>	<b>0.51</b>	<b>0.20</b>	<b>8.97<sup>a</sup></b>	<b>0.48</b>	<b>0.20</b>	<b>8.98<sup>a</sup></b>	<b>0.49</b>	<b>0.21</b>	8.70	0.22	0.20
Other <sup>4</sup>	3.77	3.71	-0.06	0.10	3.70	-0.07	0.10	3.70	-0.07	0.10	3.66	-0.11	0.10
No Coverage	5.14	5.20	0.06	0.15	5.22	0.08	0.15	5.23	0.09	0.15	5.15	0.01	0.15

*DF* = degrees of freedom; Grad = graduate; K = 1 thousand; M = 1 million; SE = standard error; SMI = serious mental illness.

<sup>a</sup> Bias is statistically significant from zero at the .05 level (also highlighted in bold).

<sup>b</sup> Bias is statistically significant from zero at the .01 level (also highlighted in bold).

<sup>1</sup> U.S. census poverty level threshold. Adults aged 18 to 22 in a college dormitory were excluded from the analysis.

<sup>2</sup> Respondents could indicate multiple types of health insurance; thus, these response categories are not mutually exclusive.

<sup>3</sup> CHIP is the Children's Health Insurance Program. Individuals aged 19 or younger are eligible for this plan.

<sup>4</sup> Other health insurance is defined as having Medicare, CHAMPUS, TRICARE, CHAMPVA, the VA, military health care, or any other type of health insurance.

NOTE: Models S1 to S4 are described in Section 3.3.1.

NOTE: Bias = Difference between cut point estimate from indicated model and estimate of SMI based on SMIYR\_U as computed in the adult NSDUH data.

Dataset = 2008A to 2012 adult NSDUH data.

Analysis weight = ANALWT\_A/5.

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

Tables 3.9a and 3.9b display SMI estimates derived from Model S4 from 2005 to 2010 and 2011 to 2014, respectively. Within the same tables, estimates derived from SMIYR\_U are placed alongside the Model S4 estimates from 2008 to 2014. SMI estimates are provided overall and across the same domains used in the domain-level bias analysis presented in Section 3.3.3. Tables 3.10a and 3.10b display the corresponding SEs of the SMI estimates. In addition, graphical representations of the estimates in Tables 3.9a and 3.9b are plotted in Figures B.1 to B.13 in Appendix B. The plotted overall estimates in Figure B.1 also display 95 percent confidence limits around the estimates, but confidence limits are not included in the domain-level plots to avoid visual clutter.

The results displayed in Tables 3.9a and 3.9b and in the figures in Appendix B indicate that overall the Model S4 and SMIYR\_U estimates of SMI agreed very well from 2008 to 2012, but appeared to diverge somewhat in 2013 (although the difference was not significant) and in 2014 (where the difference was significant at the .05 level). Note that the difference between the Model S4 and SMIYR\_U estimates appeared to be driven almost entirely by males in 2013 (where the difference was significant at the .05 level) and 2014 (where the difference was significant at the .01 level); there was very little difference among females for these 2 years.

In addition, the 2005 to 2007 Model S4 estimates appeared to follow the general trend pattern observed between 2008 and 2014. This pattern also was largely followed at the domain level, where only 9 differences between the Model S4 and SMIYR\_U estimates from 2008 to 2012 were significant at the .05 level; however, another 12 differences in 2013 and 2014 were significant at the .05 level. For domains with small sample sizes (e.g., Native Hawaiian or Other Pacific Islander), the estimates between 2005 and 2007 may have fluctuated somewhat erratically, but some of these estimates would meet criteria for precision-based suppression.

## 3.4 Alternative Models for Predicting AMI

### 3.4.1 Candidate AMI Models

Estimates of AMI from 2008 to 2012 are currently based on AMIYR\_U, which is a cut point estimator of AMI derived from the 2012 SMI model but using a different cut point (see Chapter 2). An *initial approach* investigating alternative models for predicting AMI followed the same process; that is, attempts were made to use the four alternative SMI models discussed in Section 3.3 but using a different cut point in each case to produce alternative AMI estimates.

ROC statistics of the resulting AMI cut point estimators associated with the four models are displayed in Table 3.11 (the statistics are similar to those displayed in Table 3.7, and are described in Section 3.3.2). Before discussing the four models with respect to the ROC statistics, first note that the cut points of the four models lie in a narrow range (0.004100 to 0.004576) and are about 4 times smaller than the AMI cut point for the 2012 model (0.019252). This arises from the fact that the distribution of predicted probabilities from the 2012 model is very different from those of the four alternative models, probably because the response variable, predictor variables, and sample size of the 2012 model are different from those of the alternative models. The actual shape of a distribution of predicted probabilities is immaterial, as long as the predicted probabilities are ordered in such a way that an appropriately situated cut point can act as an accurate discriminator between AMI positive and negative status.

Table 3.9a Estimates of SMI Based on Model S4 and SMIYR\_U, by Domain: 2005 to 2010

Demographic Variable	2005	2006	2007	2008		2009		2010	
	Model S4	Model S4	Model S4	SMIYR U	Model S4	SMIYR U	Model S4	SMIYR U	Model S4
<b>Total</b>	3.82	3.94	3.98	3.70	3.73	3.69	3.72	4.07	4.06
<b>Gender</b>									
Male	2.56	2.81	2.75	2.64	2.41*	2.59	2.61	3.02	2.95
Female	4.99	4.99	5.14	4.69	4.96	4.71	4.75	5.06	5.10
<b>Age in Years</b>									
18 to 25	4.23	3.72	3.81	3.75	3.76	3.35	3.76 <sup>b</sup>	3.92	3.96
26 to 34	4.78	5.48	4.85	4.73	4.99	5.04	4.97	5.09	4.89
35 to 49	4.89	4.88	5.01	4.80	4.68	4.80	4.82	5.27	5.17
50 or Older	2.44	2.71	2.96	2.53	2.57	2.55	2.48	2.97	3.07
<b>Race/Ethnicity</b>									
Not Hispanic	3.89	4.12	4.13	3.88	3.94	3.82	3.81	4.21	4.18
White	4.10	4.32	4.48	4.21	4.29	4.08	4.09	4.32	4.44
Black	2.89	3.05	3.01	2.55	2.68	2.73	2.74	3.85	3.26 <sup>a</sup>
American Indian or Alaska Native	6.73	7.63	5.27	3.01	4.07	4.82	4.37	8.09	6.00
Native Hawaiian or Other Pacific Islander	9.81*	2.73	1.15	1.16	1.05	3.53*	2.89*	2.61*	1.95
Asian	1.67	1.64	1.16	2.24	1.54	1.64	1.70	2.23	1.83
Two or More Races	6.38	11.73	6.23	4.93	6.02	7.52	6.78	7.41	7.49
Hispanic	3.33	2.75	3.01	2.56	2.34	2.88	3.09	3.23	3.31
<b>Region</b>									
Northeast	3.73	4.10	3.54	3.88	4.05	3.51	3.63	3.90	3.85
Midwest	4.29	3.87	4.07	3.98	3.85	4.32	4.16	4.55	4.39
South	3.55	4.31	4.13	3.40	3.53	3.40	3.62	3.71	4.01
West	3.84	3.31	4.04	3.78	3.67	3.69	3.51	4.33	4.00
<b>County Type</b>									
Large Metro	3.59	3.57	3.63	3.59	3.55	3.51	3.46	3.83	3.81
Small Metro	4.01	3.96	4.20	3.95	3.85	3.69	3.86	4.25	4.22
250K to 1M	3.78	3.98	4.14	3.93	3.83	3.65	3.87	4.50	4.39
Less than 250K	4.45	3.91	4.30	4.01	3.89	3.77	3.85	3.75	3.88
Nonmetro	4.20	5.09	4.75	3.61	4.08	4.27	4.28	4.53	4.61
Urbanized	4.89	5.39	4.85	4.11	4.53	4.07	3.71	4.88	4.57
Less Urbanized	4.29	4.64	4.82	3.44	4.05	4.30	4.63	3.95	4.42
Completely Rural	1.60	6.09	4.13	2.36	2.31	4.99	5.12	5.97	5.75

See notes at end of table.

(continued)

**Table 3.9a Estimates of SMI Based on Model S4 and SMIYR\_U, by Domain: 2005 to 2010 (continued)**

Demographic Variable	2005	2006	2007	2008		2009		2010	
	Model S4	Model S4	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4
<b>Received Mental Health Services</b>									
Yes	19.70	20.95	20.11	18.07	<b>19.68<sup>b</sup></b>	18.38	18.92	19.95	20.86
No	1.45	1.43	1.51	1.47	<b>1.24<sup>b</sup></b>	1.43	1.37	1.54	<b>1.38<sup>a</sup></b>
<b>Employment</b>									
Full Time	3.02	3.44	3.03	2.90	2.98	2.79	2.90	3.06	2.91
Part Time	3.91	3.66	3.90	3.73	4.07	3.54	3.64	4.10	4.07
Unemployed	6.98	6.93	7.30	6.70	6.63	5.39	5.61	6.25	6.16
Other	4.95	4.70	5.45	4.81	4.59	4.94	4.74	5.29	5.55
<b>Education</b>									
< High School	4.18	4.21	5.18	3.34	3.57	3.91	3.98	3.42	3.89
High School Grad	3.79	3.91	4.00	3.91	3.95	3.76	3.68	4.69	<b>4.22<sup>b</sup></b>
Some College	4.44	5.10	4.33	4.60	4.46	4.23	4.37	4.56	4.71
College Grad	3.03	2.74	2.92	2.87	2.90	3.01	3.04	3.32	3.41
<b>Poverty Level<sup>1</sup></b>									
< 100% Threshold	7.64	6.60	7.12	5.86	6.07	6.47	6.69	7.00	6.84
100% to 199% Threshold	4.42	4.52	4.69	4.95	5.05	4.54	4.20	4.71	4.89
≥ 200% Threshold	3.02	3.34	3.30	3.03	3.00	2.95	3.05	3.30	3.26
<b>Health Insurance<sup>2</sup></b>									
Private	2.63	2.99	2.88	2.89	2.81	2.65	2.72	3.16	3.08
Medicaid/CHIP <sup>3</sup>	9.76	8.59	8.73	7.81	8.40	8.71	8.63	8.01	8.42
Other <sup>4</sup>	3.16	3.54	3.75	3.30	3.23	3.41	3.18	4.06	3.98
No Coverage	5.64	5.42	5.68	4.93	5.06	4.72	4.82	4.95	5.08

\*Low precision; estimate would be suppressed due to not meeting the NSDUH suppression rule (also highlighted in italics).

Grad = graduate; K = 1 thousand; M = 1 million; SMI = serious mental illness.

<sup>a</sup> Difference between Model S4 and SMIYR\_U estimates is statistically significant from zero at the .05 level (also highlighted in bold).

<sup>b</sup> Difference between Model S4 and SMIYR\_U estimates is statistically significant from zero at the .01 level (also highlighted in bold).

<sup>1</sup> U.S. census poverty level threshold. Adults aged 18 to 22 in a college dormitory were excluded from the analysis.

<sup>2</sup> Respondents could indicate multiple types of health insurance; thus, these response categories are not mutually exclusive.

<sup>3</sup> CHIP is the Children's Health Insurance Program. Individuals aged 19 or younger are eligible for this plan.

<sup>4</sup> Other health insurance is defined as having Medicare, CHAMPUS, TRICARE, CHAMPVA, the VA, military health care, or any other type of health insurance.

NOTE: Model S4 is described in Section 3.3.1.

Dataset = 2005 to 2010 adult NSDUH data.

Analysis weight = ANALWT.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2005 to 2010.



Table 3.9b Estimates of SMI Based on Model S4 and SMIYR\_U, by Domain: 2011 to 2014

Demographic Variable	2011		2012		2013		2014	
	SMIYR_U	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4
<b>Total</b>	3.87	3.90	4.07	4.08	4.22	4.07	4.09	<b>3.91<sup>a</sup></b>
<b>Gender</b>								
Male	2.88	2.82	3.20	3.23	3.45	<b>3.10<sup>a</sup></b>	3.09	<b>2.75<sup>b</sup></b>
Female	4.80	4.89	4.88	4.87	4.93	4.98	5.01	4.98
<b>Age in Years</b>								
18 to 25	3.78	3.89	4.11	4.31	4.22	4.28	4.81	4.71
26 to 34	5.01	4.92	5.17	4.93	5.74	5.26	5.24	4.85
35 to 49	5.07	5.28	5.17	5.14	5.04	4.78	4.68	<b>4.29<sup>b</sup></b>
50 or Older	2.76	2.68	3.00	3.07	3.20	3.17	3.11	3.10
<b>Race/Ethnicity</b>								
Not Hispanic	4.11	4.13	4.01	4.00	4.30	4.14	4.19	<b>4.00<sup>a</sup></b>
White	4.44	4.40	4.24	4.17	4.61	4.48	4.43	4.25
Black	2.78	3.20	3.41	3.74	2.85	2.77	3.12	2.86
American Indian or Alaska Native	9.77	8.87	8.54	8.35	5.81	5.20	4.02	3.31
Native Hawaiian or Other Pacific Islander	3.50*	1.35	1.78	3.45	1.26	1.51	2.90	4.14
Asian	1.91	2.18	2.00	1.77	2.90	2.24	2.38	2.22
Two or More Races	4.61	4.41	4.20	4.47	7.07	6.34	8.94	8.74
Hispanic	2.50	2.54	4.40	4.54	3.74	3.70	3.50	3.36
<b>Region</b>								
Northeast	3.37	3.59	3.65	3.51	3.84	3.85	4.01	4.00
Midwest	4.20	4.13	3.93	4.23	4.45	4.41	4.51	<b>4.14<sup>a</sup></b>
South	3.62	3.60	4.10	4.22	4.39	4.11	3.94	3.83
West	4.38	4.40	4.48	4.18	4.01	3.88	4.01	3.75
<b>County Type</b>								
Large Metro	3.36	3.48	3.84	3.83	4.00	3.92	3.77	3.58
Small Metro	4.49	4.45	4.49	4.44	4.14	4.00	4.31	4.23
250K to 1M	4.26	4.27	4.19	4.23	4.07	4.12	4.27	4.24
Less than 250K	4.98	4.82	5.09	4.88	4.27	3.76	4.40	4.22
Nonmetro	4.37	4.21	4.05	4.24	5.10	4.71	4.79	4.44
Urbanized	4.54	4.87	3.61	3.98	5.73	5.24	5.02	4.51
Less Urbanized	4.39	3.85	4.40	4.47	4.62	4.47	4.52	4.53
Completely Rural	3.60	3.20	4.39	4.28	4.42	3.43	5.17	<b>3.76<sup>a</sup></b>

See notes at end of table.

(continued)

**Table 3.9b Estimates of SMI Based on Model S4 and SMIYR\_U, by Domain: 2011 to 2014 (continued)**

Demographic Variable	2011		2012		2013		2014	
	SMIYR_U	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4
<b>Received Mental Health Services</b>								
Yes	18.45	19.42 <sup>a</sup>	17.59	18.32	19.79	19.35	18.94	18.80
No	1.57	1.45	1.77	1.66	1.56	1.47	1.52	<b>1.33<sup>b</sup></b>
<b>Employment</b>								
Full Time	2.68	2.77	2.73	2.72	2.72	2.75	2.93	2.79
Part Time	4.61	4.58	3.91	3.69	4.82	<b>4.04<sup>a</sup></b>	4.20	4.04
Unemployed	5.42	5.43	7.82	7.52	6.62	6.62	6.71	5.71
Other	5.19	5.13	5.64	5.86	5.99	5.83	5.59	5.45
<b>Education</b>								
< High School	4.70	4.52	4.79	4.89	4.36	4.62	4.81	4.71
High School Grad	3.44	3.63	4.43	4.31	4.27	3.97	4.23	<b>3.83<sup>b</sup></b>
Some College	4.67	4.63	4.39	4.58	4.83	4.85	4.75	4.48
College Grad	3.19	3.20	3.06	2.99	3.57	3.25	3.06	3.11
<b>Poverty Level<sup>1</sup></b>								
< 100% Threshold	7.59	7.71	7.24	7.57	7.68	<b>7.01<sup>a</sup></b>	7.03	6.91
100% to 199% Threshold	4.60	4.49	5.16	5.40	5.14	5.32	5.11	5.02
≥ 200% Threshold	2.84	2.88	2.97	2.84	3.17	3.04	3.09	<b>2.86<sup>a</sup></b>
<b>Health Insurance<sup>2</sup></b>								
Private	2.61	2.66	2.69	2.69	3.05	2.96	2.93	2.79
Medicaid/CHIP <sup>3</sup>	8.74	8.78	8.46	8.93	8.28	8.26	8.23	8.45
Other <sup>4</sup>	3.48	3.30	4.36	4.46	4.14	4.15	4.10	4.06
No Coverage	5.20	5.27	6.07	<b>5.42<sup>b</sup></b>	5.93	5.50	5.25	<b>4.62<sup>b</sup></b>

\*Low precision; estimate would be suppressed due to not meeting the NSDUH suppression rule (also highlighted in italics).

Grad = graduate; K = 1 thousand; M = 1 million; SMI = serious mental illness.

<sup>a</sup> Difference between Model S4 and SMIYR\_U estimates is statistically significant from zero at the .05 level (also highlighted in bold).

<sup>b</sup> Difference between Model S4 and SMIYR\_U estimates is statistically significant from zero at the .01 level (also highlighted in bold).

<sup>1</sup> U.S. census poverty level threshold. Adults aged 18 to 22 in a college dormitory were excluded from the analysis.

<sup>2</sup> Respondents could indicate multiple types of health insurance; thus, these response categories are not mutually exclusive.

<sup>3</sup> CHIP is the Children's Health Insurance Program. Individuals aged 19 or younger are eligible for this plan.

<sup>4</sup> Other health insurance is defined as having Medicare, CHAMPUS, TRICARE, CHAMPVA, the VA, military health care, or any other type of health insurance.

NOTE: Model S4 is described in Section 3.3.1.

Dataset = 2011 to 2014 adult NSDUH data.

Analysis weight = ANALWT.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011 to 2014.

**Table 3.10a Standard Errors of Estimates of SMI Based on Model S4 and SMIYR\_U, by Domain: 2005 to 2010**

Demographic Variable	2005	2006	2007	2008		2009		2010	
	Model S4	Model S4	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4
<b>Total</b>	0.138	0.140	0.145	0.137	0.138	0.142	0.136	0.158	0.154
<b>Gender</b>									
Male	0.161	0.184	0.184	0.170	0.155	0.168	0.168	0.198	0.196
Female	0.221	0.211	0.218	0.213	0.222	0.225	0.214	0.226	0.225
<b>Age in Years</b>									
18 to 25	0.184	0.159	0.168	0.163	0.163	0.153	0.172	0.174	0.185
26 to 34	0.360	0.342	0.308	0.323	0.322	0.338	0.338	0.374	0.369
35 to 49	0.266	0.312	0.253	0.275	0.277	0.278	0.273	0.289	0.287
50 or Older	0.227	0.236	0.258	0.244	0.246	0.231	0.225	0.272	0.269
<b>Race/Ethnicity</b>									
Not Hispanic	0.146	0.153	0.154	0.148	0.153	0.158	0.150	0.167	0.166
White	0.164	0.173	0.177	0.174	0.183	0.184	0.175	0.186	0.185
Black	0.376	0.431	0.361	0.299	0.297	0.321	0.309	0.362	0.340
American Indian or Alaska Native	1.704	2.292	1.620	0.887	1.203	1.214	1.156	2.508	2.216
Native Hawaiian or Other Pacific Islander	5.287*	1.408	0.680	0.578	0.561	2.455*	2.344*	1.808*	0.757
Asian	0.504	0.419	0.393	0.683	0.527	0.429	0.449	0.587	0.510
Two or More Races	1.285	3.054	1.164	1.253	1.553	1.576	1.393	2.472	2.471
Hispanic	0.422	0.318	0.400	0.330	0.280	0.330	0.336	0.395	0.411
<b>Region</b>									
Northeast	0.335	0.335	0.306	0.317	0.334	0.291	0.317	0.311	0.313
Midwest	0.276	0.249	0.257	0.250	0.231	0.268	0.268	0.305	0.292
South	0.223	0.270	0.236	0.226	0.246	0.251	0.235	0.253	0.271
West	0.290	0.272	0.366	0.316	0.290	0.289	0.275	0.352	0.342
<b>County Type</b>									
Large Metro	0.203	0.192	0.205	0.188	0.181	0.206	0.192	0.214	0.209
Small Metro	0.231	0.244	0.249	0.252	0.239	0.226	0.239	0.245	0.250
250K to 1M	0.254	0.297	0.317	0.305	0.276	0.276	0.285	0.326	0.332
Less than 250K	0.422	0.462	0.428	0.455	0.448	0.389	0.422	0.331	0.361
Nonmetro	0.325	0.386	0.341	0.311	0.373	0.340	0.312	0.454	0.443
Urbanized	0.614	0.529	0.575	0.555	0.613	0.544	0.436	0.724	0.629
Less Urbanized	0.432	0.518	0.446	0.401	0.538	0.462	0.482	0.522	0.585
Completely Rural	0.424	1.872	0.884	0.665	0.689	1.150	1.053	2.237	2.201

See notes at end of table.

(continued)

**Table 3.10a Standard Errors of Estimates of SMI Based on Model S4 and SMIYR\_U, by Domain: 2005 to 2010 (continued)**

Demographic Variable	2005	2006	2007	2008		2009		2010	
	Model S4	Model S4	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4
<b>Received Mental Health Services</b>									
Yes	0.802	0.803	0.862	0.802	0.830	0.825	0.800	0.857	0.884
No	0.089	0.088	0.084	0.090	0.082	0.085	0.079	0.095	0.091
<b>Employment</b>									
Full Time	0.160	0.177	0.155	0.157	0.165	0.161	0.166	0.182	0.176
Part Time	0.329	0.338	0.375	0.352	0.421	0.288	0.308	0.323	0.317
Unemployed	0.788	0.943	0.997	0.853	0.829	0.572	0.575	0.649	0.637
Other	0.324	0.298	0.349	0.333	0.309	0.330	0.325	0.334	0.359
<b>Education</b>									
< High School	0.370	0.366	0.425	0.297	0.311	0.335	0.338	0.299	0.346
High School Grad	0.217	0.238	0.262	0.245	0.272	0.267	0.264	0.320	0.307
Some College	0.267	0.322	0.270	0.279	0.271	0.269	0.281	0.288	0.299
College Grad	0.278	0.228	0.245	0.240	0.231	0.249	0.253	0.248	0.260
<b>Poverty Level<sup>1</sup></b>									
< 100% Threshold	0.550	0.479	0.511	0.434	0.440	0.497	0.501	0.533	0.564
100% to 199% Threshold	0.322	0.346	0.354	0.371	0.396	0.316	0.275	0.328	0.348
≥ 200% Threshold	0.152	0.165	0.161	0.149	0.150	0.162	0.160	0.170	0.157
<b>Health Insurance<sup>2</sup></b>									
Private	0.132	0.150	0.144	0.154	0.150	0.137	0.141	0.167	0.160
Medicaid/CHIP <sup>3</sup>	0.773	0.649	0.767	0.647	0.638	0.672	0.656	0.686	0.742
Other <sup>4</sup>	0.315	0.330	0.343	0.303	0.286	0.339	0.316	0.338	0.340
No Coverage	0.372	0.416	0.373	0.341	0.405	0.332	0.355	0.377	0.386

\*Low precision; estimate would be suppressed due to not meeting the NSDUH suppression rule (also highlighted in italics).

Grad = graduate; K = 1 thousand; M = 1 million; SMI = serious mental illness.

<sup>1</sup> U.S. census poverty level threshold. Adults aged 18 to 22 in a college dormitory were excluded from the analysis.

<sup>2</sup> Respondents could indicate multiple types of health insurance; thus, these response categories are not mutually exclusive.

<sup>3</sup> CHIP is the Children's Health Insurance Program. Individuals aged 19 or younger are eligible for this plan.

<sup>4</sup> Other health insurance is defined as having Medicare, CHAMPUS, TRICARE, CHAMPVA, the VA, military health care, or any other type of health insurance.

NOTE: Model S4 is described in Section 3.3.1.

Dataset = 2005 to 2010 adult NSDUH data.

Analysis weight = ANALWT.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2005 to 2010.

**Table 3.10b Standard Errors of Estimates of SMI Based on Model S4 and SMIYR\_U, by Domain: 2011 to 2014**

Demographic Variable	2011		2012		2013		2014	
	SMIYR_U	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4
<b>Total</b>	0.140	0.141	0.145	0.141	0.160	0.151	0.120	0.115
<b>Gender</b>								
Male	0.170	0.169	0.187	0.190	0.227	0.202	0.159	0.142
Female	0.216	0.216	0.211	0.200	0.217	0.228	0.171	0.177
<b>Age in Years</b>								
18 to 25	0.171	0.173	0.171	0.178	0.175	0.180	0.212	0.212
26 to 34	0.355	0.355	0.371	0.351	0.409	0.391	0.297	0.262
35 to 49	0.262	0.288	0.306	0.298	0.299	0.284	0.224	0.209
50 or Older	0.223	0.219	0.247	0.247	0.262	0.250	0.194	0.199
<b>Race/Ethnicity</b>								
Not Hispanic	0.155	0.159	0.150	0.145	0.169	0.167	0.130	0.124
White	0.179	0.181	0.175	0.163	0.195	0.197	0.148	0.138
Black	0.315	0.366	0.372	0.394	0.357	0.331	0.294	0.281
American Indian or Alaska Native	3.342	3.287	2.518	2.498	1.509	1.505	1.104	0.908
Native Hawaiian or Other Pacific Islander	2.230*	0.974	0.762	1.675	0.715	0.950	1.205	2.144
Asian	0.523	0.575	0.502	0.426	0.836	0.628	0.430	0.634
Two or More Races	1.078	0.897	0.788	0.980	1.327	1.204	1.390	1.244
Hispanic	0.292	0.267	0.444	0.442	0.403	0.407	0.299	0.299
<b>Region</b>								
Northeast	0.296	0.308	0.316	0.314	0.306	0.297	0.266	0.266
Midwest	0.252	0.248	0.235	0.247	0.294	0.280	0.261	0.245
South	0.219	0.224	0.261	0.262	0.291	0.274	0.187	0.194
West	0.342	0.338	0.318	0.306	0.354	0.342	0.261	0.218
<b>County Type</b>								
Large Metro	0.190	0.193	0.205	0.205	0.233	0.227	0.164	0.153
Small Metro	0.249	0.244	0.277	0.261	0.277	0.263	0.224	0.224
250K to 1M	0.303	0.299	0.285	0.298	0.354	0.341	0.243	0.281
Less than 250K	0.413	0.427	0.585	0.497	0.443	0.388	0.438	0.365
Nonmetro	0.332	0.352	0.309	0.313	0.365	0.340	0.296	0.277
Urbanized	0.522	0.583	0.390	0.404	0.639	0.574	0.438	0.421
Less Urbanized	0.459	0.449	0.540	0.536	0.498	0.516	0.433	0.427
Completely Rural	0.740	0.698	0.979	0.997	1.029	0.738	0.852	0.697

See notes at end of table.

(continued)

**Table 3.10b Standard Errors of Estimates of SMI Based on Model S4 and SMIYR\_U, by Domain: 2011 to 2014**  
(continued)

Demographic Variable	2011		2012		2013		2014	
	SMIYR_U	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4	SMIYR_U	Model S4
<b>Received Mental Health Services</b>								
Yes	0.764	0.773	0.692	0.709	0.866	0.811	0.658	0.613
No	0.096	0.091	0.104	0.103	0.101	0.099	0.072	0.067
<b>Employment</b>								
Full Time	0.158	0.160	0.153	0.160	0.168	0.169	0.135	0.130
Part Time	0.398	0.385	0.330	0.310	0.416	0.347	0.269	0.262
Unemployed	0.551	0.539	0.789	0.738	0.708	0.727	0.576	0.490
Other	0.315	0.315	0.348	0.357	0.372	0.364	0.272	0.275
<b>Education</b>								
< High School	0.395	0.380	0.420	0.447	0.396	0.459	0.356	0.344
High School Grad	0.222	0.231	0.264	0.260	0.276	0.269	0.224	0.216
Some College	0.295	0.285	0.281	0.275	0.311	0.317	0.235	0.215
College Grad	0.241	0.248	0.243	0.233	0.319	0.249	0.189	0.204
<b>Poverty Level<sup>1</sup></b>								
< 100% Threshold	0.527	0.525	0.475	0.464	0.460	0.440	0.360	0.387
100% to 199% Threshold	0.310	0.305	0.346	0.343	0.335	0.382	0.295	0.292
≥ 200% Threshold	0.150	0.151	0.156	0.140	0.192	0.170	0.132	0.127
<b>Health Insurance<sup>2</sup></b>								
Private	0.141	0.140	0.135	0.131	0.172	0.165	0.121	0.115
Medicaid/CHIP <sup>3</sup>	0.651	0.662	0.659	0.624	0.572	0.598	0.444	0.493
Other <sup>4</sup>	0.286	0.281	0.378	0.387	0.372	0.348	0.265	0.258
No Coverage	0.370	0.375	0.408	0.377	0.406	0.398	0.329	0.303

\*Low precision; estimate would be suppressed due to not meeting the NSDUH suppression rule (also highlighted in italics).

Grad = graduate; K = 1 thousand; M = 1 million; SMI = serious mental illness.

<sup>1</sup> U.S. census poverty level threshold. Adults aged 18 to 22 in a college dormitory were excluded from the analysis.

<sup>2</sup> Respondents could indicate multiple types of health insurance; thus, these response categories are not mutually exclusive.

<sup>3</sup> CHIP is the Children's Health Insurance Program. Individuals aged 19 or younger are eligible for this plan.

<sup>4</sup> Other health insurance is defined as having Medicare, CHAMPUS, TRICARE, CHAMPVA, the VA, military health care, or any other type of health insurance.

NOTE: Model S4 is described in Section 3.3.1.

Dataset = 2011 to 2014 adult NSDUH data.

Analysis weight = ANALWT.

Source: SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2011 to 2014.

**Table 3.11 ROC Statistics of AMI Cut Point Estimates Based on Four SMI Candidate Models**

Model	Sample Size	Pop Size (1,000s)	Cut Point	AMIYR_U Estimate	Cut point Estimate	SE	Bias	False-Positive Rate	False-Negative Rate	Total Error Rate	Sensitivity	Specificity	AUC
S1	206,510	229,830	0.004226	18.06	18.04	0.111	-0.0165	4.26	4.28	8.54	0.763	0.948	0.856
S2	206,510	229,830	0.004176	18.06	18.07	0.111	0.0178	4.28	4.27	8.55	0.764	0.948	0.856
S3	206,510	229,830	0.004100	18.06	17.76	0.110	-0.2960	4.09	4.39	8.48	0.757	0.950	0.854
S4	206,510	229,830	0.004576	18.06	18.50	0.111	0.4416	4.54	4.10	8.64	0.773	0.945	0.859

AMI = any mental illness; AUC = area under curve (i.e., average of sensitivity and specificity); Pop = population; ROC = receiver operating characteristic; SE = standard error of cut point estimate; SMI = serious mental illness.

NOTE: Models S1 to S4 are described in Section 3.3.1.

NOTE: Bias = false-positive rate – false-negative rate.

NOTE: Total error rate = false-positive rate + false-negative rate.

Response variable = SMIYR\_U.

Dataset = 2008A to 2012 adult NSDUH data.

Analysis weight = ANALWT\_A/5.

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

Table 3.11 indicates that although the two least parsimonious SMI models (i.e., Models S1 and S2) show levels of bias close to zero, the two most parsimonious SMI models (i.e., Models S3 and S4) show substantial levels of bias. Before exploring Models S1 and S2 further, different approaches to providing estimates of AMI were considered to see if more parsimonious models with levels of bias close to zero (and with smaller total error rates) could be identified.

Similar to the SMI model identification process, a *second approach* considered was to develop candidate models directly from AMIYR\_U (i.e., by using AMIYR\_U as the response variable). However, this approach does not guarantee that the resulting AMI estimates would be *consistent* with the SMI estimates derived separately from the alternative SMI models. For example, a respondent might be predicted to have SMI from one of the candidate SMI models, but predicted not to have AMI from a candidate AMI model,<sup>13</sup> and in practice these inconsistencies were found to occur.

To prevent inconsistencies between predicted SMI and AMI, a *third approach* was used in which the respondents who were predicted as having SMI from one of the candidate SMI models (Model S4 was chosen) were automatically assigned to be predicted as also having AMI, and the corresponding records were removed from the data (about 4 percent of the records). Then alternative logistic regression models for predicting AMI were examined by using AMIYR\_U as the response variable in a subset of the data where respondents who were predicted to have SMI based on Model S4 were excluded from the data.

Using the third approach, the process of developing alternative AMI models was similar to that used in developing alternative SMI models discussed in Section 3.1. The same predictor variables shown in Table 3.1 were considered, and four candidate models were chosen by using an increasingly stringent criterion for including predictor variables. Table 3.12 displays the predictor variable inclusion criterion and the corresponding numerator *DF* of each of the four candidate models.<sup>14</sup>

**Table 3.12 Four Alternative Candidate Models for Predicting AMI**

Model	<i>DF</i>	Predictor Variable Inclusion Criterion
A1	22	Wald <i>F</i> test <i>p</i> value < .05
A2	21	Wald <i>F</i> test <i>p</i> value < .01
A3	20	Wald <i>F</i> test statistic > 10
A4	7	Wald <i>F</i> test statistic > 100

*DF* = degrees of freedom.

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

Tables 3.13 to 3.16 provide details of the four models listed in Table 3.12, namely, the predictor variable, *DF*, Wald *F* test statistic, Wald *F* test *p* value, beta, beta SE, and beta *p* value. The tables indicate that all four candidate models included the three predictor variables that were identical to or comparable with those from the 2012 model (see Section 3.1 for details); these

<sup>13</sup> By definition, AMI includes all three of the following categories of mental illness: SMI, moderate mental illness (MMI), and mild (low) mental illness (LMI) (CBHSQ, 2014; Chapter 1). Therefore, a respondent having SMI necessarily has AMI.

<sup>14</sup> For all models, the denominator *DF* = 900.



three predictor variables had the largest Wald  $F$  statistics in each of the models. Two of the less parsimonious models also included the variable that had similarities with  $X_s$  in the 2012 model (i.e., serious thoughts of suicide in the past year associated with the adult depression section of the questionnaire), but the Wald  $F$  statistic for this variable was much smaller in both cases.

**Table 3.13 Model A1 (22  $DF$ ): Predictor Variables,  $DF$ , and Wald  $F$  and Beta Statistics**

Predictor Variable	$DF$	Wald $F$ Statistic	Wald $F$ $p$ Value	Beta	Beta SE	Beta $p$ Value
Intercept				-5.49	0.066	0.000
WSPDSC2	1	7,958.2	0.000	0.58	0.006	0.000
AMDEYR_U	1	2,666.7	0.000	4.59	0.089	0.000
AGE1830	1	1,187.0	0.000	0.17	0.005	0.000
ADWRSTHK_U	1	9.5	0.002	0.27	0.088	0.002
AMDELT_U	1	195.0	0.000	0.87	0.063	0.000
AMHINP2_U	1	5.0	0.025	0.43	0.190	0.025
AMHRX2_U	1	71.8	0.000	0.49	0.058	0.000
MHLMNT_U	1	12.6	0.000	0.43	0.123	0.000
MHLTHER_U	1	37.1	0.000	0.50	0.082	0.000
MHRCOST2_U	1	70.0	0.000	0.83	0.099	0.000
MHRHAND2_U	1	42.2	0.000	0.78	0.120	0.000
MHRWHER2_U	1	13.7	0.000	0.69	0.187	0.000
MHRENUF2_U	1	10.4	0.001	0.82	0.254	0.001
MHRNMBRS2_U	1	14.1	0.000	0.94	0.249	0.000
ANXDLIF_U	1	17.7	0.000	0.23	0.054	0.000
DEPRSLIF_U	1	123.1	0.000	0.57	0.051	0.000
JBSTAT14	1	14.7	0.000	0.29	0.077	0.000
HEALTH_U	5	49.8	0.000			
Level 0				0.31	1.193	0.795
Level 2				0.27	0.045	0.000
Level 3				0.43	0.046	0.000
Level 4				0.76	0.062	0.000
Level 5				1.23	0.105	0.000

$DF$  = degrees of freedom; SE = standard error.

NOTE: See Table 3.1 for brief descriptions of predictor variables, and Appendix A for further details.

NOTE: The variable HEALTH\_U has six levels (1 = excellent to 5 = poor; 0 = missing recode); Level 1 is the reference level in the model.

Response variable = AMIYR\_U.

Dataset = 2008A to 2012 adult NSDUH data, excluding cases predicted as having SMI based on Model S4.

Analysis weight = ANALWT\_A/5.

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

**Table 3.14 Model A2 (21 *DF*): Predictor Variables, *DF*, and Wald *F* and Beta Statistics**

<b>Predictor Variable</b>	<b><i>DF</i></b>	<b>Wald <i>F</i> Statistic</b>	<b>Wald <i>F</i> <i>p</i> Value</b>	<b>Beta</b>	<b>Beta SE</b>	<b>Beta <i>p</i> Value</b>
Intercept				-5.48	0.066	0.000
WSPDSC2	1	7,964.4	0.000	0.58	0.006	0.000
AMDEYR_U	1	2,673.5	0.000	4.59	0.089	0.000
AGE1830	1	1,180.4	0.000	0.17	0.005	0.000
ADWRSTHK_U	1	9.6	0.002	0.27	0.088	0.002
AMDELT_U	1	193.9	0.000	0.87	0.063	0.000
AMHRX2_U	1	74.8	0.000	0.50	0.058	0.000
MHLMNT_U	1	15.1	0.000	0.47	0.121	0.000
MHLTHER_U	1	37.4	0.000	0.50	0.082	0.000
MHRCOST2_U	1	70.2	0.000	0.83	0.099	0.000
MHRHAND2_U	1	41.9	0.000	0.78	0.120	0.000
MHRWHER2_U	1	13.7	0.000	0.69	0.187	0.000
MHRENUF2_U	1	10.3	0.001	0.81	0.254	0.001
MHRNMBRS2_U	1	14.1	0.000	0.94	0.251	0.000
ANXDLIF_U	1	17.6	0.000	0.23	0.054	0.000
DEPRSLIF_U	1	121.7	0.000	0.56	0.051	0.000
JBSTAT14	1	15.6	0.000	0.30	0.076	0.000
HEALTH_U	5	50.1	0.000			
Level 0				0.31	1.194	0.797
Level 2				0.27	0.045	0.000
Level 3				0.43	0.046	0.000
Level 4				0.76	0.062	0.000
Level 5				1.24	0.105	0.000

*DF* = degrees of freedom; SE = standard error.

NOTE: See [Table 3.1](#) for brief descriptions of predictor variables, and Appendix A for further details.

NOTE: The variable HEALTH\_U has six levels (1 = excellent to 5 = poor; 0 = missing recode); Level 1 is the reference level in the model.

Response variable = AMIYR\_U.

Dataset = 2008A to 2012 adult NSDUH data, excluding cases predicted as having SMI based on Model S4.

Analysis weight = ANALWT\_A/5.

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

**Table 3.15 Model A3 (20 *DF*): Predictor Variables, *DF*, and Wald *F* and Beta Statistics**

Predictor Variable	<i>DF</i>	Wald <i>F</i> Statistic	Wald <i>F</i> <i>p</i> Value	Beta	Beta SE	Beta <i>p</i> Value
Intercept				-5.48	0.066	0.000
WSPDSC2	1	7,975.4	0.000	0.58	0.006	0.000
AMDEYR_U	1	2,819.3	0.000	4.54	0.086	0.000
AGE1830	1	1,177.0	0.000	0.16	0.005	0.000
AMDELT_U	1	395.1	0.000	0.96	0.049	0.000
AMHRX2_U	1	73.6	0.000	0.50	0.058	0.000
MHLMNT_U	1	15.7	0.000	0.48	0.121	0.000
MHLTHER_U	1	38.9	0.000	0.51	0.082	0.000
MHRCOST2_U	1	70.3	0.000	0.83	0.099	0.000
MHRHAND2_U	1	42.2	0.000	0.78	0.120	0.000
MHRWHERE2_U	1	13.6	0.000	0.69	0.187	0.000
MHRENUF2_U	1	10.5	0.001	0.82	0.253	0.001
MHRNMBRS2_U	1	14.3	0.000	0.95	0.252	0.000
ANXDLIF_U	1	17.4	0.000	0.23	0.054	0.000
DEPRSLIF_U	1	125.4	0.000	0.58	0.052	0.000
JBSTAT14	1	15.5	0.000	0.30	0.076	0.000
HEALTH_U	5	49.9	0.000			
Level 0				0.31	1.193	0.796
Level 2				0.27	0.045	0.000
Level 3				0.43	0.046	0.000
Level 4				0.76	0.062	0.000
Level 5				1.23	0.105	0.000

*DF* = degrees of freedom; SE = standard error.

NOTE: See [Table 3.1](#) for brief descriptions of predictor variables, and Appendix A for further details.

NOTE: The variable HEALTH\_U has six levels (1 = excellent to 5 = poor; 0 = missing recode); Level 1 is the reference level in the model.

Response variable = AMIYR\_U.

Dataset = 2008A to 2012 adult NSDUH data, excluding cases predicted as having SMI based on Model S4.

Analysis weight = ANALWT\_A/5.

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

**Table 3.16 Model A4 (7 DF): Predictor Variables, *DF*, and Wald *F* and Beta Statistics**

Predictor Variable	<i>DF</i>	Wald <i>F</i> Statistic	Wald <i>F</i> <i>p</i> Value	Beta	Beta SE	Beta <i>p</i> Value
Intercept				-5.26	0.057	0.000
WSPDSC2	1	8,385.4	0.000	0.60	0.006	0.000
AMDEYR_U	1	2,903.0	0.000	4.59	0.085	0.000
AGE1830	1	1,379.7	0.000	0.18	0.005	0.000
AMDELT_U	1	380.7	0.000	0.94	0.048	0.000
AMHRX2_U	1	172.7	0.000	0.70	0.053	0.000
MHRCOST2_U	1	176.9	0.000	1.20	0.090	0.000
DEPRSLIF_U	1	183.3	0.000	0.69	0.051	0.000

*DF* = degrees of freedom; SE = standard error.

NOTE: See [Table 3.1](#) for brief descriptions of predictor variables, and Appendix A for further details.

Response variable = AMIYR\_U.

Dataset = 2008A to 2012 adult NSDUH data, excluding cases predicted as having SMI based on Model S4.

Analysis weight = ANALWT\_A/5.

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

### 3.4.2 ROC Analysis of AMI Candidate Models

A ROC analysis was applied to AMI candidate models using the third approach to investigating alternative AMI models described in Section 3.4.1. The four AMI candidate models described in [Tables 3.13](#) to [3.16](#) were run using the 2008A to 2012 adult NSDUH data (where for consistency reasons, SMI positive cases already determined from Model S4 were excluded from the data) and the appropriate analysis weight divided by the number of years represented (i.e., ANALWT\_A/5). For each model, the response variable was the binary variable AMIYR\_U, and each model was designed to yield a *predicted probability* of AMI for each respondent.

Predicted probabilities from each of the four candidate models were then *dichotomized* into a binary cut point estimator of AMI as follows. A cut point was established such that respondents with predicted probabilities at or above the cut point were determined to have AMI and the rest were not. The value of the cut point was selected so that the weighted number of false positives (i.e., respondents for whom AMIYR\_U = 0 but whose predicted probabilities were at or above the cut point) would come as close as possible to equaling the weighted number of false negatives (respondents for whom AMIYR\_U = 1 but whose predicted probabilities were less than the cut point). Finally, respondents predicted as having AMI under each of the four candidate models were then combined with the SMI positive cases excluded from the data to form a complete category of cases predicted as having AMI, thus resulting in a consistent AMI cut point estimator.

ROC statistics of the AMI cut point estimators associated with the four AMI models are displayed in [Table 3.17](#). The following ROC statistics are included in the table:

- false-positive rate, which is the proportion of all respondents who were determined to have AMI under the AMI cut point estimator, but for whom the response variables for the models in question indicated the opposite;

**Table 3.17 ROC Statistics of AMI Cut Point Estimates Based on Four Candidate Models**

Model	Sample Size	Pop Size (1,000s)	Cut Point	AMIYR_U Estimate	Cut Point Estimate	SE	Bias	False-Positive Rate	False-Negative Rate	Total Error Rate	Sensitivity	Specificity	AUC
A1	206,510	229,830	0.234472	18.06	18.06	0.105	0.0001	3.97	3.97	7.94	0.780	0.952	0.866
A2	206,510	229,830	0.234264	18.06	18.06	0.105	0.0005	3.96	3.96	7.93	0.780	0.952	0.866
A3	206,510	229,830	0.233642	18.06	18.01	0.105	-0.0455	3.92	3.96	7.88	0.780	0.952	0.866
A4	206,510	229,830	0.216896	18.06	18.06	0.106	0.0001	4.06	4.06	8.12	0.775	0.950	0.863

AMI = any mental illness; AUC = area under curve (i.e., average of sensitivity and specificity); Pop = population; ROC = receiver operating characteristic; SE = standard error of cut point estimate; SMI = serious mental illness.

NOTE: Models A1 to A4 are described in Section 3.3.1. The AMI cut point estimates are derived by adding the corresponding cut point estimates of SMI based on Model S4 to those of AMI based on each of the four AMI candidate models, respectively, applied to the subset of data excluding cases predicted as having SMI based on Model S4. The cut point refers to the actual cut point of the AMI model in question, but all ROC statistics relate to the combined AMI cut point estimator.

NOTE: Bias = false-positive rate – false-negative rate.

NOTE: Total error rate = false-positive rate + false-negative rate.

Response variable = AMIYR\_U.

Dataset = 2008A to 2012 adult NSDUH data, excluding cases predicted as having SMI based on Model S4.

Analysis weight = ANALWT\_A/5.

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

- false-negative rate, which is the proportion of all respondents who were determined not to have AMI under the AMI cut point estimator, but for whom the response variables for the models in question indicated the opposite;
- bias, which is the difference between the false-positive and false-negative rates;
- total error rate, which is the sum of the false-positive and false-negative rates;
- sensitivity, which is the proportion of respondents for whom the response variables for the models in question indicated that they had AMI, who were predicted to have AMI by the AMI cut point estimator;
- specificity, which is the proportion of respondents for whom the response variables for the models in question indicated that they did not have AMI, who were predicted not to have AMI by the AMI cut point estimator; and
- AUC (i.e., under the ROC curve) based on predicted responses dichotomized by the cut point, which is the average of the sensitivity and specificity.

The ROC statistics displayed in [Table 3.17](#) indicate that all four models showed levels of bias similarly close to zero (although Model A3's absolute bias was a little larger), and Models A1 to A3 showed similarly low levels of total error rate. However, the most parsimonious model (Model A4 with 7 *DF*) appeared to have a larger total error rate. Therefore, based on both criteria of bias and total error rate, it appeared that Model A3 (20 *DF*) may be the "best" model among the four (assuming that the slight negative bias was acceptably low), but there was a possibility that a more parsimonious model with similar or better bias and total error rate values may exist.

Finally, note that the ROC statistics (particularly bias and total error rate) that resulted from using this approach showed fairly substantial improvements over those that resulted from using the first approach (i.e., using different cut points to estimate AMI from the SMI candidate models). Therefore, the AMI cut point estimators resulting from the third approach (i.e., where the four AMI models were used in combination with the SMI cut point estimator derived from SMI Model S4) were evaluated further to see how they performed with respect to domain-level bias.

### 3.4.3 Bias of AMI Candidate Models, by Domain

Domain-level bias results associated with the four AMI candidate models are displayed in [Table 3.18](#). Recall that the combined AMI estimates referred to in the table were obtained by adding the corresponding SMI estimates derived from Model S4 to the AMI estimates obtained from the AMI models applied to the subset of data excluding all cases predicted as having SMI based on Model S4. For all four candidate models, [Table 3.18](#) indicates significant bias at the .05 level (and also at the .01 level in many of those cases) across many domains, particularly those associated with age, receipt of mental health services, employment status, education level, poverty level, and health insurance.

Because all four models indicated domain-level bias across so many domains, it was decided that no further investigations into alternative models for AMI would be undertaken; hence, no estimates of AMI from 2005 to 2007 were derived from any of the candidate models.

Table 3.18 Bias of AMI Cut Point Estimates of Four Candidate Models, by Domain: 2008A to 2012

Demographic Variable	AMIYR_U Estimate	Model A1 (22 <i>DF</i> )			Model A2 (21 <i>DF</i> )			Model A3 (20 <i>DF</i> )			Model A4 (7 <i>DF</i> )		
		Estimate	Bias	SE Bias	Estimate	Bias	SE Bias	Estimate	Bias	SE Bias	Estimate	Bias	SE Bias
<b>Total</b>	18.06	18.06	0.00	0.10	18.06	0.00	0.10	18.01	-0.05	0.10	18.06	0.00	0.10
<b>Gender</b>													
Male	14.33	14.25	-0.08	0.14	14.25	-0.08	0.14	14.22	-0.11	0.14	14.14	-0.19	0.15
Female	21.53	21.60	0.07	0.15	21.61	0.08	0.15	21.54	0.01	0.15	21.71	0.18	0.15
<b>Age in Years</b>													
18 to 25	18.60	<b>20.98<sup>b</sup></b>	<b>2.38</b>	<b>0.13</b>	<b>20.99<sup>b</sup></b>	<b>2.39</b>	<b>0.13</b>	<b>21.04<sup>b</sup></b>	<b>2.44</b>	<b>0.13</b>	<b>21.39<sup>b</sup></b>	<b>2.79</b>	<b>0.14</b>
26 to 34	21.79	<b>22.33<sup>a</sup></b>	<b>0.55</b>	<b>0.23</b>	<b>22.35<sup>a</sup></b>	<b>0.56</b>	<b>0.23</b>	<b>22.35<sup>a</sup></b>	<b>0.56</b>	<b>0.23</b>	<b>22.82<sup>b</sup></b>	<b>1.04</b>	<b>0.23</b>
35 to 49	20.36	20.22	-0.14	0.18	20.21	-0.15	0.18	20.20	-0.16	0.18	20.28	-0.07	0.18
50 or Older	15.00	<b>14.05<sup>b</sup></b>	<b>-0.95</b>	<b>0.19</b>	<b>14.05<sup>b</sup></b>	<b>-0.95</b>	<b>0.19</b>	<b>13.93<sup>b</sup></b>	<b>-1.07</b>	<b>0.19</b>	<b>13.68<sup>b</sup></b>	<b>-1.31</b>	<b>0.19</b>
<b>Race/Ethnicity</b>													
Not Hispanic	18.51	18.43	-0.08	0.12	18.43	-0.08	0.12	18.37	-0.14	0.12	18.45	-0.06	0.12
White	19.00	18.92	-0.08	0.12	18.93	-0.07	0.12	18.86	-0.14	0.12	19.01	0.01	0.13
Black	16.66	16.71	0.05	0.31	16.68	0.02	0.31	16.67	0.01	0.31	16.40	-0.26	0.30
American Indian or Alaska Native	22.70	24.24	1.54	1.13	24.21	1.51	1.13	24.17	1.47	1.13	23.08	0.38	1.38
Native Hawaiian or Other Pacific Islander	23.20	23.45	0.25	2.16	23.47	0.27	2.16	23.40	0.20	2.16	22.87	-0.33	2.17
Asian	13.36	12.59	-0.77	0.66	12.48	-0.88	0.67	12.45	-0.91	0.66	12.60	-0.76	0.67
Two or More Races	25.10	25.68	0.59	0.90	25.68	0.59	0.90	25.88	0.78	0.91	25.27	0.17	0.93
Hispanic	15.29	15.79	0.50	0.25	15.78	0.49	0.25	<b>15.81<sup>a</sup></b>	<b>0.51</b>	<b>0.25</b>	15.68	0.39	0.26
<b>Region</b>													
Northeast	17.54	17.43	-0.11	0.22	17.42	-0.12	0.22	17.45	-0.09	0.22	17.43	-0.11	0.23
Midwest	18.38	18.51	0.13	0.18	18.53	0.14	0.18	18.53	0.15	0.18	18.56	0.18	0.18
South	17.83	17.82	-0.01	0.17	17.81	-0.02	0.17	17.73	-0.10	0.17	17.75	-0.08	0.17
West	18.52	18.50	-0.01	0.25	18.52	0.00	0.25	18.42	-0.10	0.25	18.57	0.05	0.25
<b>County Type</b>													
Large Metro	17.39	17.28	-0.11	0.15	17.26	-0.12	0.15	17.21	-0.18	0.15	17.33	-0.06	0.15
Small Metro	18.82	18.72	-0.09	0.17	18.75	-0.07	0.17	18.68	-0.13	0.17	18.79	-0.02	0.18
250K to 1M	18.69	18.50	-0.19	0.22	18.52	-0.17	0.22	18.46	-0.23	0.22	18.71	0.02	0.22
Less than 250K	19.07	19.16	0.09	0.27	19.20	0.13	0.27	19.14	0.07	0.27	18.96	-0.11	0.28
Nonmetro	18.82	<b>19.36<sup>a</sup></b>	<b>0.54</b>	<b>0.27</b>	<b>19.36<sup>a</sup></b>	<b>0.54</b>	<b>0.27</b>	<b>19.36<sup>a</sup></b>	<b>0.55</b>	<b>0.26</b>	19.05	0.23	0.26
Urbanized	18.91	19.60	0.69	0.42	19.63	0.72	0.42	19.65	0.74	0.42	19.36	0.45	0.39
Less Urbanized	19.34	19.72	0.38	0.38	19.72	0.37	0.38	19.68	0.34	0.37	19.34	-0.01	0.38
Completely Rural	16.04	16.69	0.65	1.05	16.62	0.58	1.05	16.70	0.66	1.04	16.41	0.37	1.03

See notes at end of table.

(continued)

**Table 3.18 Bias of AMI Cut Point Estimates of Four Candidate Models, by Domain: 2008A to 2012 (continued)**

Demographic Variable	AMIYR_U Estimate	Model A1 (22 <i>DF</i> )			Model A2 (21 <i>DF</i> )			Model A3 (20 <i>DF</i> )			Model A4 (7 <i>DF</i> )		
		Estimate	Bias	SE Bias	Estimate	Bias	SE Bias	Estimate	Bias	SE Bias	Estimate	Bias	SE Bias
<b>Received Mental Health Services</b>													
Yes	53.88	<b>59.89<sup>b</sup></b>	<b>6.01</b>	<b>0.39</b>	<b>59.88<sup>b</sup></b>	<b>6.00</b>	<b>0.39</b>	<b>59.45<sup>b</sup></b>	<b>5.57</b>	<b>0.39</b>	<b>60.35<sup>b</sup></b>	<b>6.47</b>	<b>0.41</b>
No	12.34	<b>11.37<sup>b</sup></b>	<b>-0.97</b>	<b>0.10</b>	<b>11.37<sup>b</sup></b>	<b>-0.97</b>	<b>0.10</b>	<b>11.39<sup>b</sup></b>	<b>-0.95</b>	<b>0.10</b>	<b>11.30<sup>b</sup></b>	<b>-1.03</b>	<b>0.10</b>
<b>Employment</b>													
Full Time	15.23	15.27	0.05	0.13	15.28	0.05	0.14	15.23	0.00	0.13	<b>15.65<sup>b</sup></b>	<b>0.42</b>	<b>0.14</b>
Part Time	19.68	20.02	0.34	0.30	20.05	0.37	0.30	20.01	0.32	0.30	<b>20.43<sup>a</sup></b>	<b>0.75</b>	<b>0.31</b>
Unemployed	23.68	<b>25.28<sup>b</sup></b>	<b>1.59</b>	<b>0.41</b>	<b>25.24<sup>b</sup></b>	<b>1.55</b>	<b>0.41</b>	<b>25.38<sup>b</sup></b>	<b>1.70</b>	<b>0.42</b>	<b>25.50<sup>b</sup></b>	<b>1.81</b>	<b>0.43</b>
Other	21.09	<b>20.54<sup>b</sup></b>	<b>-0.55</b>	<b>0.21</b>	<b>20.53<sup>b</sup></b>	<b>-0.56</b>	<b>0.21</b>	<b>20.46<sup>b</sup></b>	<b>-0.63</b>	<b>0.21</b>	<b>19.66<sup>b</sup></b>	<b>-1.43</b>	<b>0.21</b>
<b>Education</b>													
< High School	19.63	<b>20.58<sup>b</sup></b>	<b>0.95</b>	<b>0.28</b>	<b>20.52<sup>b</sup></b>	<b>0.89</b>	<b>0.28</b>	<b>20.50<sup>b</sup></b>	<b>0.87</b>	<b>0.28</b>	19.57	-0.06	0.28
High School Grad	17.82	17.82	0.00	0.18	17.84	0.02	0.18	17.82	0.00	0.18	17.67	-0.15	0.19
Some College	19.45	19.36	-0.09	0.19	19.38	-0.08	0.19	19.31	-0.14	0.19	19.62	0.16	0.19
College Grad	16.22	<b>15.81<sup>a</sup></b>	<b>-0.41</b>	<b>0.20</b>	<b>15.81<sup>a</sup></b>	<b>-0.41</b>	<b>0.20</b>	<b>15.75<sup>a</sup></b>	<b>-0.48</b>	<b>0.20</b>	16.27	0.04	0.20
<b>Poverty Level<sup>1</sup></b>													
< 100% Threshold	25.57	<b>27.37<sup>b</sup></b>	<b>1.80</b>	<b>0.30</b>	<b>27.34<sup>b</sup></b>	<b>1.77</b>	<b>0.30</b>	<b>27.39<sup>b</sup></b>	<b>1.82</b>	<b>0.30</b>	<b>26.56<sup>b</sup></b>	<b>0.99</b>	<b>0.29</b>
100% to 199% Threshold	21.06	21.44	0.37	0.25	21.44	0.37	0.25	21.46	0.39	0.25	21.04	-0.02	0.25
≥ 200% Threshold	15.68	<b>15.20<sup>b</sup></b>	<b>-0.48</b>	<b>0.13</b>	<b>15.20<sup>b</sup></b>	<b>-0.47</b>	<b>0.13</b>	<b>15.12<sup>b</sup></b>	<b>-0.56</b>	<b>0.13</b>	15.48	-0.20	0.13
<b>Health Insurance<sup>2</sup></b>													
Private	15.49	<b>15.00<sup>b</sup></b>	<b>-0.50</b>	<b>0.13</b>	<b>15.01<sup>b</sup></b>	<b>-0.48</b>	<b>0.13</b>	<b>14.95<sup>b</sup></b>	<b>-0.54</b>	<b>0.13</b>	15.33	-0.17	0.13
Medicaid/CHIP <sup>3</sup>	29.28	<b>31.60<sup>b</sup></b>	<b>2.32</b>	<b>0.39</b>	<b>31.54<sup>b</sup></b>	<b>2.26</b>	<b>0.39</b>	<b>31.57<sup>b</sup></b>	<b>2.29</b>	<b>0.39</b>	29.95	0.67	0.38
Other <sup>4</sup>	17.65	<b>16.18<sup>b</sup></b>	<b>-1.46</b>	<b>0.25</b>	<b>16.19<sup>b</sup></b>	<b>-1.45</b>	<b>0.24</b>	<b>16.13<sup>b</sup></b>	<b>-1.51</b>	<b>0.25</b>	<b>15.54<sup>b</sup></b>	<b>-2.11</b>	<b>0.25</b>
No Coverage	21.21	<b>22.35<sup>b</sup></b>	<b>1.14</b>	<b>0.27</b>	<b>22.34<sup>b</sup></b>	<b>1.13</b>	<b>0.27</b>	<b>22.31<sup>b</sup></b>	<b>1.10</b>	<b>0.27</b>	<b>22.29<sup>b</sup></b>	<b>1.08</b>	<b>0.27</b>

AMI = any mental illness; *DF* = degrees of freedom; Grad = graduate; K = one thousand; M = one million; SE = standard error; SMI = serious mental illness.

<sup>a</sup> Bias is statistically significant from zero at the .05 level (also highlighted in bold).

<sup>b</sup> Bias is statistically significant from zero at the .01 level (also highlighted in bold).

<sup>1</sup> U.S. census poverty level threshold. Adults aged 18 to 22 in a college dormitory were excluded from the analysis.

<sup>2</sup> Respondents could indicate multiple types of health insurance; thus, these response categories are not mutually exclusive.

<sup>3</sup> CHIP is the Children's Health Insurance Program. Individuals aged 19 or younger are eligible for this plan.

<sup>4</sup> Other health insurance is defined as having Medicare, CHAMPUS, TRICARE, CHAMPVA, the VA, military health care, or any other type of health insurance.

NOTE: Models A1 to A4 are described in Section 3.3.1. The AMI estimates are derived by adding the corresponding estimates of SMI based on Model S4 to those of AMI based on each of the four AMI candidate models, respectively, applied to the subset of data excluding cases predicted as having SMI based on Model S4.

NOTE: Bias = Difference between AMI cut point estimate from indicated model and estimate of AMI based on AMIYR\_U as computed in the adult NSDUH data.

Dataset = 2008A to 2012 adult NSDUH data.

Analysis weight = ANALWT\_A/5.

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



## 4. Summary and Conclusions

The objective of this methodological study was to investigate whether information from the 2005 to 2012 National Surveys on Drug Use and Health (NSDUHs) could be used to determine if models of mental illness could be fit to the 2008 to 2012 NSDUH data and applied to 2005 to 2007 NSDUH data to provide comparable estimates of 2005 to 2007 serious mental illness (SMI) and any mental illness (AMI) with those obtained from the 2008 to 2012 data.

Because the aim was to provide alternative estimates of SMI (AMI) that matched those derived from SMIYR\_U (AMIYR\_U), the SMIYR\_U (AMIYR\_U) variable(s) were designated as the *response variable(s)* in the alternative models. Therefore, the entire adult NSDUH data from 2008A<sup>15</sup> to 2012 was available for the development of those models. Candidate *predictor variables* were created from information contained in common in all NSDUHs from 2005 to 2012. A set of suitable predictor variables meeting this requirement was identified and then whittled down to a smaller subset of variables used in the models presented in this study (see [Table 3.1](#) for brief descriptions of the predictor variables and Appendix A for further details).

Based on the various analyses discussed in this methodological report, the following broad conclusions can be drawn about the alternative estimates of SMI and AMI that were created from the 2005 to 2012 NSDUHs.

### 4.1 SMI Estimation

The most parsimonious of the four candidate SMI models investigated (i.e., Model S4) appeared to provide SMI estimates with the "best" characteristics listed as follows:

- The receiver operating characteristic (ROC) statistics of the SMI cut point estimator associated with Model S4 exhibited a very low level of bias and a low total error rate overall.
- At the domain level, very few domains exhibited levels of bias significant at the .05 level.
- At the overall level, estimates of SMI derived from Model S4 were very similar to those derived from SMIYR\_U from 2008 to 2012, and the Model S4 estimates from 2005 to 2007 appeared to follow the general trend pattern observed between 2008 and 2012.
- At the domain level, Model S4 estimates were for the most part very similar to SMIYR\_U estimates from 2008 to 2012, where only nine differences were significant at the .05 level across all 5 years. For domains with small sample sizes, some estimates did appear to fluctuate, but some of those estimates met criteria for precision-based suppression.

---

<sup>15</sup> The 2008A NSDUH is Sample A of 2008. For explanatory details, see Section 1.1 in Chapter 1 and Chapter 2.

Model S4 was parsimonious with six predictor variables (providing 6 degrees of freedom [*DF*]) and shared many similarities with the 2012 SMI model (with 5 *DF*). For example, three of the six variables in Model S4 were identical to or comparable with those from the 2012 model, and two others were similar to the other two predictor variables in the 2012 model.

## **4.2 AMI Estimation**

The development of alternative models to estimate AMI was more challenging. The first approach where a different cut point was applied to the alternative SMI models to provide AMI estimates did not produce satisfactory results in terms of bias and total error rate. The second approach of modeling AMIYR\_U directly provided some inconsistent results (e.g., some respondents were predicted to have SMI under an SMI model but not AMI under a separate AMI model). The third approach was specifically designed to avoid inconsistencies and proceeded as follows: (1) respondents predicted as having SMI under a selected SMI model were assigned to be predicted as also having AMI and were removed from the data; (2) using the remaining data, candidate models using AMIYR\_U as the response variable were investigated; and (3) estimates from the models were combined with SMI estimates from the selected SMI model to provide estimates of AMI.

The most parsimonious of the four candidate AMI models that appeared to provide AMI estimates with the "best" characteristics was Model A3 (with 20 *DF*), but the results were not satisfactory. These are listed as follows:

- The ROC statistics of the AMI cut point estimator associated with Model A3 exhibited a fairly low level of bias, but also a fairly high total error rate overall.
- At the domain level, a large number of domains exhibited levels of bias significant at the .05 level (and in many cases, also at the .01 level), particularly those associated with age, receipt of mental health services, employment status, education level, poverty level, and health insurance.
- Model A3 was not parsimonious with 20 *DF*. Therefore, there is a greater likelihood that the model could include predictor variables that researchers might wish to cross-analyze with AMI, which is not recommended (Center for Behavioral Health Statistics and Quality, 2015b).

For the reasons listed above, it was decided that no further investigations into alternative models for AMI would be conducted. Hence, no estimates of AMI from 2005 to 2007 were derived from any of the candidate models.

## **4.3 Concluding Remarks**

National estimates of SMI based on Model S4 for 2005 to 2007 appear to be comparable with those based on the 2012 model for 2008 and later. Therefore, researchers could consider using the estimates of SMI for 2005 to 2007 based on Model S4 depending on their analysis. For this reason, future versions of the 2005 to 2007 data files will contain the following two variables:

1. SMIPP\_A: predicted probability of SMI based on Model S4 (2005 to 2007) and
2. SMIYR\_A: SMI status ("yes" or "no") after dichotomizing SMIPP\_A using the cut point 0.372556455 (i.e., if SMIPP\_A is less than the cut point, then SMIYR\_A = 0; if SMIPP\_A is greater than or equal to the cut point, then SMIYR\_A = 1).

The restricted use confidential files also can be used to create all of the predictor variables should they be of interest to researchers. [Tables A.1](#) and [A.2](#) in Appendix A provide guidance on how to create the predictor variables listed in [Table 3.1](#). The public use files (PUFs) can be used to create all of the predictor variables except the variables that require AGE, AGE1830, or JBSTATR as the source variables because these variables are not included in the PUFs. However, an approximation to AGE1830 may be created from the variable CATAG7 in the PUFs (i.e., where 4 = 18 to 20 years old, 5 = ages 21 to 25 years old, 6 = 26 to 34 years old, and 7 = 35 or older); such an approximation has not been tested in any model. The variable JBSTAT14 can be created from the PUFs exactly as described in [Table A.2](#) simply by using the PUF variable JBSTATR2 instead of JBSTATR.

It may also be possible for researchers to apply these models to other studies in which the same (or very similar) information is collected to enable the required predictor variables to be created. However, this approach should be applied with caution because these models were specifically developed from within NSDUH and other studies might have very different methodological characteristics that could affect the models and subsequent estimates.

*This page intentionally left blank*

# References

Center for Behavioral Health Statistics and Quality. (2009). *2008 National Survey on Drug Use and Health: Methodological resource book (Section 16, Measuring serious mental illness with the NSDUH: Results of 2008 12-month analysis)*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2012). *2010 National Survey on Drug Use and Health: Methodological resource book (Section 16b, Analysis of effects of 2008 NSDUH questionnaire changes: Methods to adjust adult MDE and SPD estimates and to estimate SMI in the 2005-2009 surveys)*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2013). *National Survey on Drug Use and Health: 2012 public use file and codebook*. Retrieved from <https://datafiles.samhsa.gov/>

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

Center for Behavioral Health Statistics and Quality. (2015a). *2014 National Survey on Drug Use and Health: Methodological summary and definitions*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2015b). *National Survey on Drug Use and Health: Alternative statistical models to predict mental illness*. 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. <https://doi.org/10.1001/archpsyc.60.2.184>

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. <https://doi.org/10.2190/t8em-c8yh-373n-luwd>

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. <https://doi.org/10.1002/mpr.313>

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. <https://doi.org/10.1002/mpr.61>

# **Appendix A: Predictor Variables Used in Alternative Models**

*This page intentionally left blank*



**Table A.1 Description of Predictor Variables Used in Alternative Models: Predictor Variables Closely Related to Those of 2012 Model**

Predictor Variable	Recode from NSDUH Analytic Variable(s)	NSDUH Analytic Variable	Codebook Description	Values	Questionnaire Section
WSPDSC2	if 2005 <= year <= 2007 then WSPDSC2 = max(K6SMXADJ – 7, 0)	K6SMXADJ (year = 2005-2007)	ADJUSTED WORST K6 TOTAL SCORE IN PAST YEAR	Range = 0-24	Recoded Psychological Distress
	else if 2008A <= year <= 2012 then WSPDSC2 = WSPDSC2;	WSPDSC2 (year = 2008A-2012)	ALTERNATIVE WORST K6 TOTAL SCORE	Range = 0-17	Recoded Mental Health
AMDEYR_U	if 2005 <= year <= 2007 then rename AJAMDEYR = AMDEYR_C;	AJAMDEYR (year = 2005-2007)	ADULT: ADJUSTED PAST YEAR MAJOR DEPRESSIVE EPISODE	. = missing; 1 = yes; 2 = no	Recoded Adult Depression
	else if 2008A <= year <= 2012 then rename AMDEYR = AMDEYR_C;	AMDEYR (year = 2008A-2012)	ADULT: PAST YEAR MAJOR DEPRESSIVE EPISODE (MDE)	. = missing; 1 = yes; 2 = no	Recoded Adult Depression
	if AMDEYR_C in (., 2) then AMDEYR_U = 0; else AMDEYR_U = AMDEYR_C;				
AGE1830	if 2005 <= year <= 2007 then do; if 18 <= AGE < 30 then AGE1830 = AGE – 18; else if 30 <= AGE then AGE1830 = 12;	AGE <sup>1</sup>	FINAL EDITED AGE	Range = 12-100	Core Demographics
	else if 2008A <= year <= 2012 then AGE1830 = AGE1830;	AGE1830 <sup>1</sup>	AGE CATEGORY USED FOR REVISED 2012 SMI MODEL	. = 12-17; 0-11 = 18-29; 12 = 30+	Recoded Core Demographics

See notes at end of table.

(continued)

**Table A.1 Description of Predictor Variables Used in Alternative Models: Predictor Variables Closely Related to Those of 2012 Model (continued)**

Predictor Variable	Recode from NSDUH Analytic Variable(s)	NSDUH Analytic Variable	Codebook Description	Values	Questionnaire Section
ADSDSSC3 <sup>2</sup>	if ADPSHMG3 > 10 then ADPSHMG3_U = 0; else ADPSHMG3_U = ADPSHMG3;	ADPSHMG3	HOW MUCH DID FEELINGS INTERFERE WITH HOME MNGMT	Scores = 0-10; missing = 94,97,98,99	Adult Depression
	if ADPSWORK > 10 then ADPSWORK_U = 0; else ADPSWORK_U = ADPSWORK;	ADPSWORK	HOW MUCH DID FEELING INTERFERE WITH ABLTY TO WORK	Scores = 0-10; missing = 94,97,98,99	Adult Depression
	if ADPSRELS > 10 then ADPSRELS_U = 0; else ADPSRELS_U = ADPSRELS;	ADPSRELS	HOW MUCH DID FEELING INTERFERE WITH RELATIONSHIPS	Scores = 0-10; missing = 94,97,98,99	Adult Depression
	if ADPSOC > 10 then ADPSOC_U = 0; else ADPSOC_U = ADPSOC;	ADPSOC	HOW MUCH DID FEELINGS INTERFERE WITH SOCIAL LIFE	Scores = 0-10; missing = 94,97,98,99	Adult Depression
	if ADPSHMG3_U < 7 then ADPSHMG3 = 0; else ADPSHMG3 = 1;				
	if ADPSWORK_U < 7 then ADPSWOR3 = 0; else ADPSWOR3 = 1;				
	if ADPSRELS_U < 7 then ADPSREL3 = 0; else ADPSREL3 = 1;				
	if ADPSOC_U < 7 then ADPSOC3 = 0; else ADPSOC3 = 1;				
	ADSDSSC3 = ADPSHMG3 + ADPSWOR3 + ADPSREL3 + ADPSOC3;				
ADWRSTHK_U	if ADWRSTHK >= 2 then ADWRSTHK_U = 0; else ADWRSTHK_U = ADWRSTHK;	ADWRSTHK	WHEN PRBLMS WORST THINK ABOUT COMMITTING SUICIDE	1 = yes; 2 = no; missing = 94,97,98,99	Adult Depression

SDS = Sheehan Disability Scale; SMI = serious mental illness; WHODAS = World Health Organization Disability Assessment Schedule.

NOTE: Variables displayed in uppercase letters are NSDUH analytic variables available in public use files (unless otherwise indicated).

<sup>1</sup> Variables AGE and AGE1830 are not available in public use files.

<sup>2</sup> Variable ADSDSSC3 was not directly associated with the 2012 (WHODAS) model, but it was used as a predictor variable based on information from the SDS in the 2008B model (Center for Behavioral Health Statistics and Quality, 2009). This predictor variable acted analogously to the predictor variable based on information from the WHODAS in the 2012 model (i.e., they both captured information about the severity of impairment).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2005 to 2012.

**Table A.2 Description of Predictor Variables Used in Alternative Models: Other Predictor Variables**

Predictor Variable	Recode from NSDUH Analytic Variable(s)	NSDUH Analytic Variable(s)	Codebook Description	Values	Questionnaire Section
AMDELT_U	if 2005 <= year <= 2007 then rename AJAMDELT = AMDELT_C;	AJAMDELT (year = 2005 to 2007)	ADULT: ADJUSTED LIFETIME MAJOR DEPRESSIVE EPISODE	. = missing; 1 = yes; 2 = no	Recoded Adult Depression
	else if 2008A <= year <= 2012 then rename AMDELT = AMDELT_C;	AMDELT (year = 2008A-2012)	ADULT: LIFETIME MAJOR DEPRESSIVE EPISODE (MDE)	. = missing; 1 = yes; 2 = no	Recoded Adult Depression
	if AMDELT_C in (., 2) then AMDELT_U = 0; else AMDELT_U = AMDELT_C;				
AMHINP2_U	if AMHINP2 in (., 2) then AMHINP2_U = 0; else AMHINP2_U = AMHINP2;	AMHINP2	RCVD INPATIENT MENTAL HEALTH TRT IN PST YR	. = missing; 1 = yes; 2 = no	Recoded Adult Mental Health Treatment
AMHRX2_U	if AMHRX2 in (., 2) then AMHRX2_U = 0; else AMHRX2_U = AMHRX2;	AMHRX2	RCVD PRESC MED FOR MENTAL HLTH TRT IN PST YR	. = missing; 1 = yes; 2 = no	Recoded Adult Mental Health Treatment
MHLMNT_U	if 2005 <= year <= 2008A then rename MHLMNT2 = MHLMNT_C;	MHLMNT2 (year = 2005 to 2008A)	RCVD OUTP MH TRT AT MH CLINIC/CENTER IN PST YR	. = missing; 0 = no; 1 = yes	Recoded Adult Mental Health Treatment
	else if 2009 <= year <= 2012 then rename MHLMNT3 = MHLMNT_C;	MHLMNT3 (year = 2009-2012)	RCVD OUTP MH TRT AT MH CLINIC/CENTER IN PST YR	. = missing; 0 = no; 1 = yes	Recoded Adult Mental Health Treatment
	if MHLMNT_C = . then MHLMNT_U = 0; else MHLMNT_U = MHLMNT_C;4				
MHLTHER_U	if 2005 <= year <= 2008A then rename MHLTHER2 = MHLTHER_C;	MHLTHER2 (year = 2005 to 2008A)	RCVD OUTP MH TRT AT PRIV THERAPIST OFC IN PST YR	. = missing; 0 = no; 1 = yes	Recoded Adult Mental Health Treatment
	else if 2009 <= year <= 2012 then rename MHLTHER3 = MHLTHER_C;	MHLTHER3 (year = 2009-2012)	RCVD OUTP MH TRT AT PRIV THERAPIST OFC IN PST YR	. = missing; 0 = no; 1 = yes	Recoded Adult Mental Health Treatment
	if MHLTHER_C = . then MHLTHER_U = 0; else MHLTHER_U = MHLTHER_C;				
MHRCOST2_U	if MHRCOST2 = . then MHRCOST2_U = 0; else MHRCOST2_U = MHRCOST2;	MHRCOST2	NO MH TRT IN PST YR B/C COULD NOT AFFORD COST	. = missing; 0 = no; 1 = yes	Recoded Adult Mental Health Treatment

See notes at end of table.

(continued)

**Table A.2 Description of Predictor Variables Used in Alternative Models: Other Predictor Variables (continued)**

Predictor Variable	Recode from NSDUH Analytic Variable(s)	NSDUH Analytic Variable(s)	Codebook Description	Values	Questionnaire Section
MHRHAND2_U	if MHRHAND2 = . then MHRHAND2_U = 0; else MHRHAND2_U = MHRHAND2;	MHRHAND2	NO MH TRT IN PST YR B/C THGT COULD HANDLE PRB W/O	. = missing; 0 = no; 1 = yes	Recoded Adult Mental Health Treatment
MHRWHERE2_U	if MHRWHERE2 = . then MHRWHERE2_U = 0; else MHRWHERE2_U = MHRWHERE2;	MHRWHERE2	NO MH TRT IN PST YR B/C DID NOT KNOW WHERE TO GO	. = missing; 0 = no; 1 = yes	Recoded Adult Mental Health Treatment
MHRENUF2_U	if MHRENUF2 = . then MHRENUF2_U = 0; else MHRENUF2_U = MHRENUF2;	MHRENUF2	NO MH TRT IN PST YR B/C INSUR NOT PAY ENOUGH	. = missing; 0 = no; 1 = yes	Recoded Adult Mental Health Treatment
MHRNMBRS2_U	if MHRNMBRS2 = . then MHRNMBRS2_U = 0; else MHRNMBRS2_U = MHRNMBRS2;	MHRNMBRS2	NO MH TRT IN PST YR B/C FEAR OF NEIGHBRS NEG OPIN	. = missing; 0 = no; 1 = yes	Recoded Adult Mental Health Treatment
MHRCMIT2_U	if MHRCMIT2 = . then MHRCMIT2_U = 0; else MHRCMIT2_U = MHRCMIT2;	MHRCMIT2	NO MH TRT IN PST YR B/C FEAR BEING COMMITTED/MED	. = missing; 0 = no; 1 = yes	Recoded Adult Mental Health Treatment
ANXDLIF_U	if ANXDLIF = . then ANXDLIF_U = 0; else ANXDLIF_U = ANXDLIF;	ANXDLIF	HAD ANXIETY IN LIFETIME	. = missing; 0 = no; 1 = yes	Recoded Pregnancy and Health Care
DEPRSLIF_U	if DEPRSLIF = . then DEPRSLIF_U = 0; else DEPRSLIF_U = DEPRSLIF;	DEPRSLIF	HAD DEPRESSION IN LIFETIME	. = missing; 0 = no; 1 = yes	Recoded Pregnancy and Health Care
DEPRSYR_U	if DEPRSYR = . then DEPRSYR_U = 0; else DEPRSYR_U = DEPRSYR;	DEPRSYR	HAD DEPRESSION IN PAST YEAR	. = missing; 0 = no; 1 = yes	Recoded Pregnancy and Health Care
WRKEDYR_U	if WRKEDYR >= 2 then WRKEDYR_U = 0; else WRKEDYR_U = WRKEDYR;	WRKEDYR	PAST 12 MOS, WORKED AT ANY JOB	1 = yes; 2 = no; missing = 94,97,98,99	Employment and Workplace
JBSTAT14	if JBSTATR = 14 then JBSTAT14 = 1; else JBSTAT14 = 0;	JBSTATR <sup>1</sup>	WORK SITUATION IN PAST WEEK - RECODE	Various codes; missing = 994,997,998,999	Employment and Workplace
HEALTH_U	if HEALTH > 5 then HEALTH_U = 0; else HEALTH_U = HEALTH;	HEALTH	OVERALL HEALTH	Scores = 1-5; missing = 94,97	Core Demographics

NOTE: Variables displayed in uppercase letters are NSDUH analytic variables available in public use files (unless otherwise indicated).

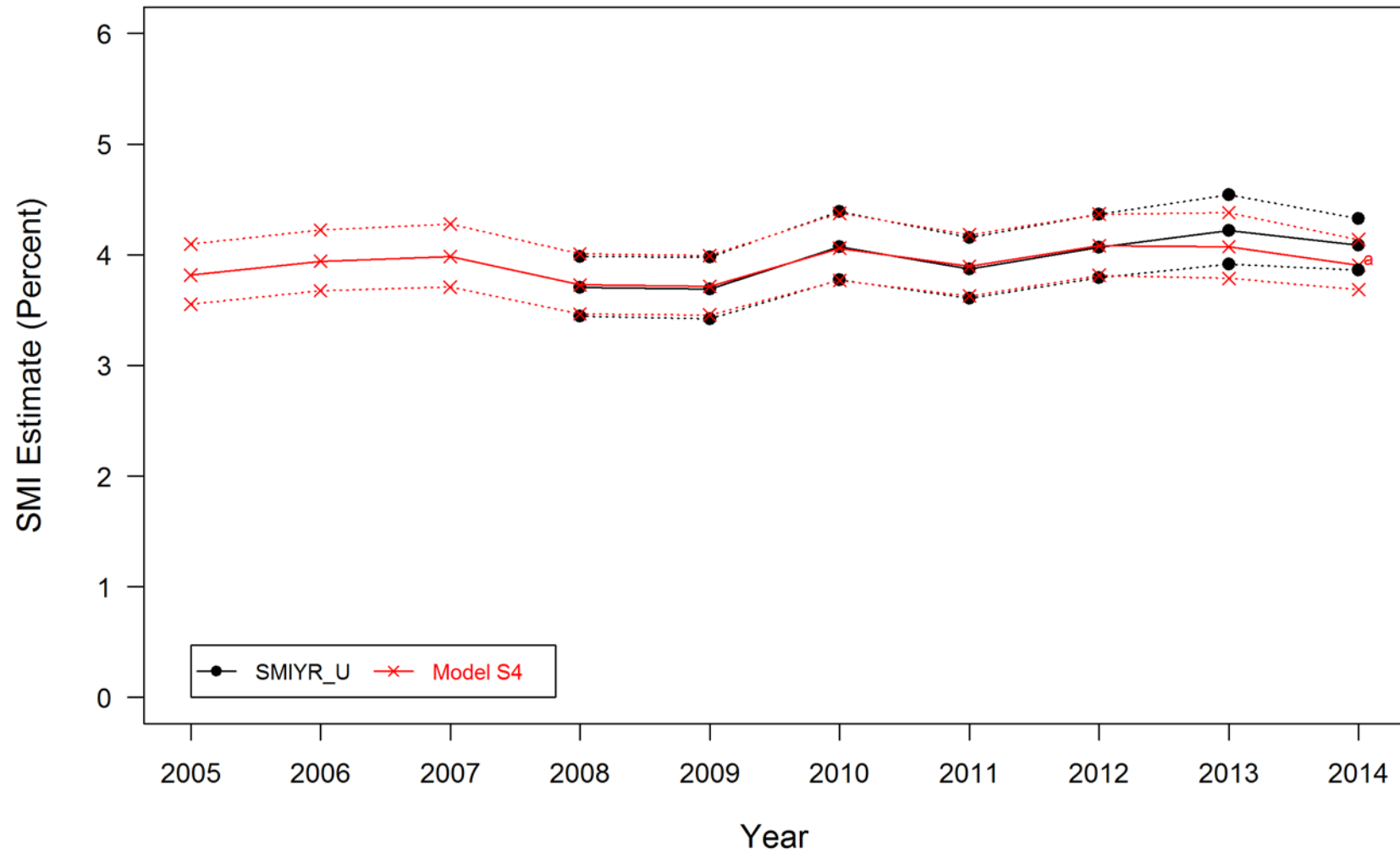
<sup>1</sup> Variable JBSTATR is not available in public use files, but JBSTATR2 is and can be used to create JBSTAT14 exactly as described by simply using it instead of JBSTATR.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2005 to 2012.

## **Appendix B: Plots of SMI Estimates Derived from Model S4 and SMIYR\_U, by Domain**

*This page intentionally left blank*

**Figure B.1 SMI Estimates Derived from Model S4 and SMIYR\_U: 2005 to 2014**



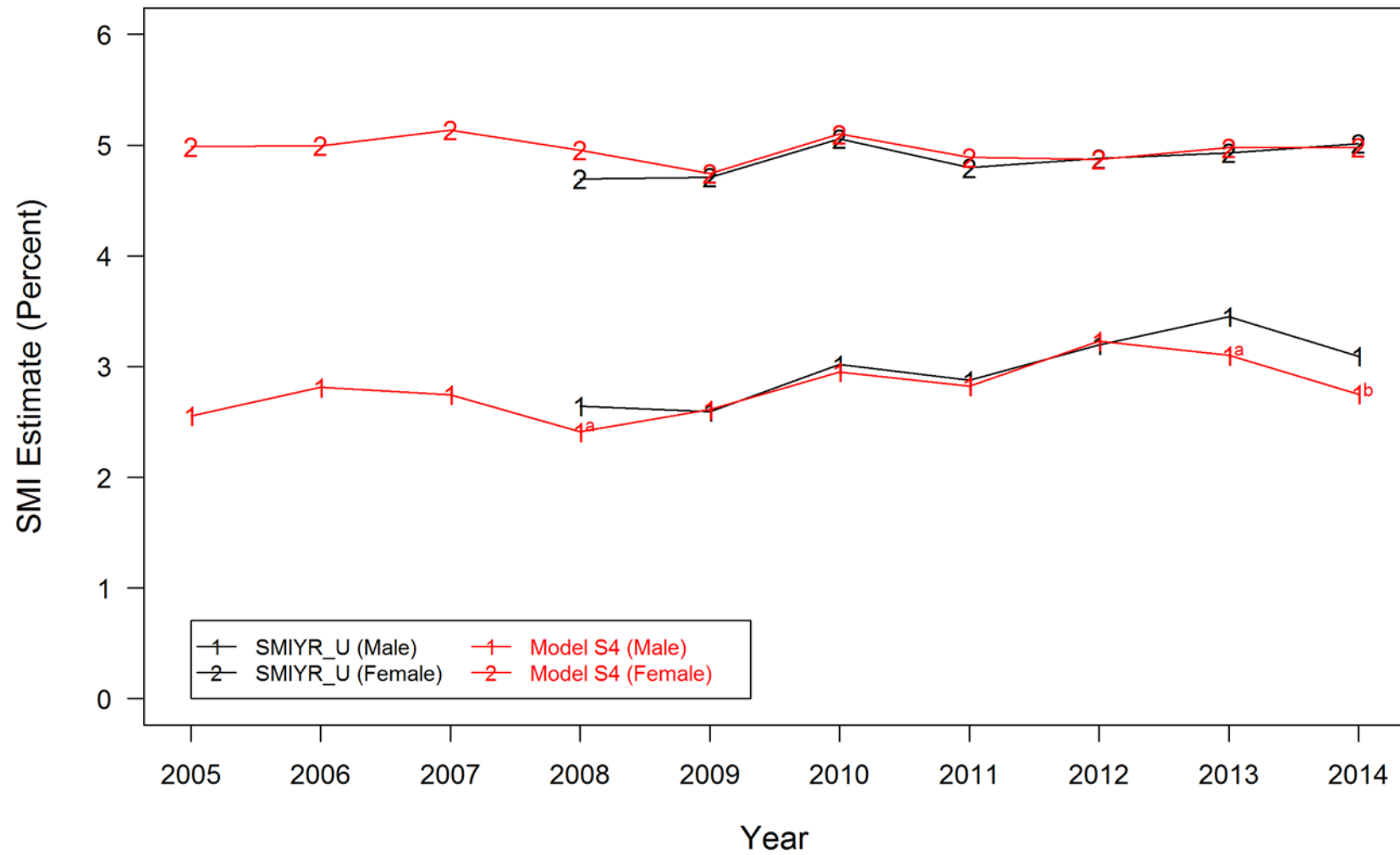
SMI = serious mental illness.

Note: Dashed lines indicate 95 percent confidence limits.

<sup>a</sup> Difference significant at .05 level.

<sup>b</sup> Difference significant at .01 level. No differences attained this level of significance in this figure.

**Figure B.2 SMI Estimates Derived from Model S4 and SMIYR\_U, by Gender: 2005 to 2014**



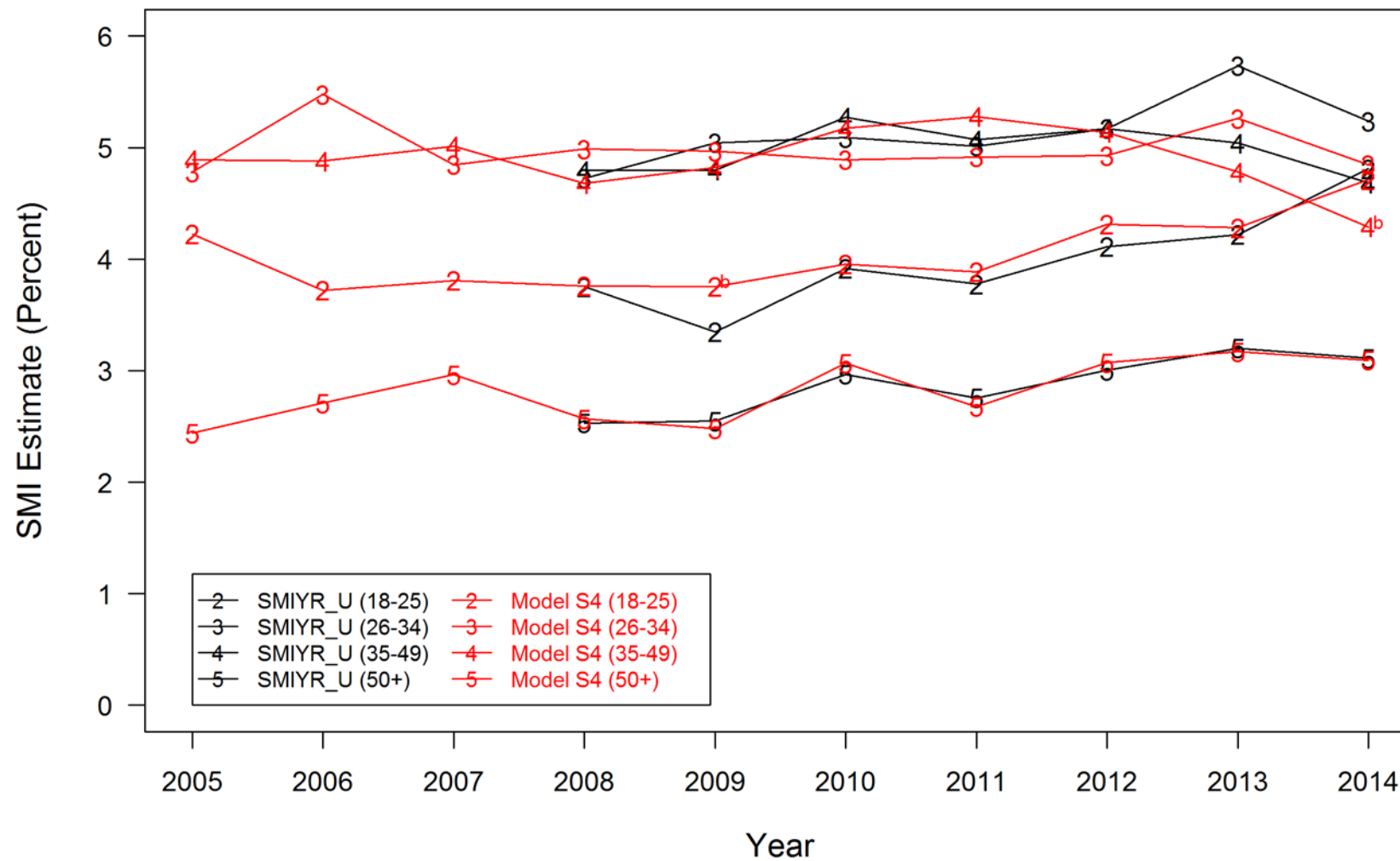
SMI = serious mental illness.

<sup>a</sup> Difference significant at .05 level.

<sup>b</sup> Difference significant at .01 level.



**Figure B.3 SMI Estimates Derived from Model S4 and SMIYR\_U, by Age Group: 2005 to 2014**

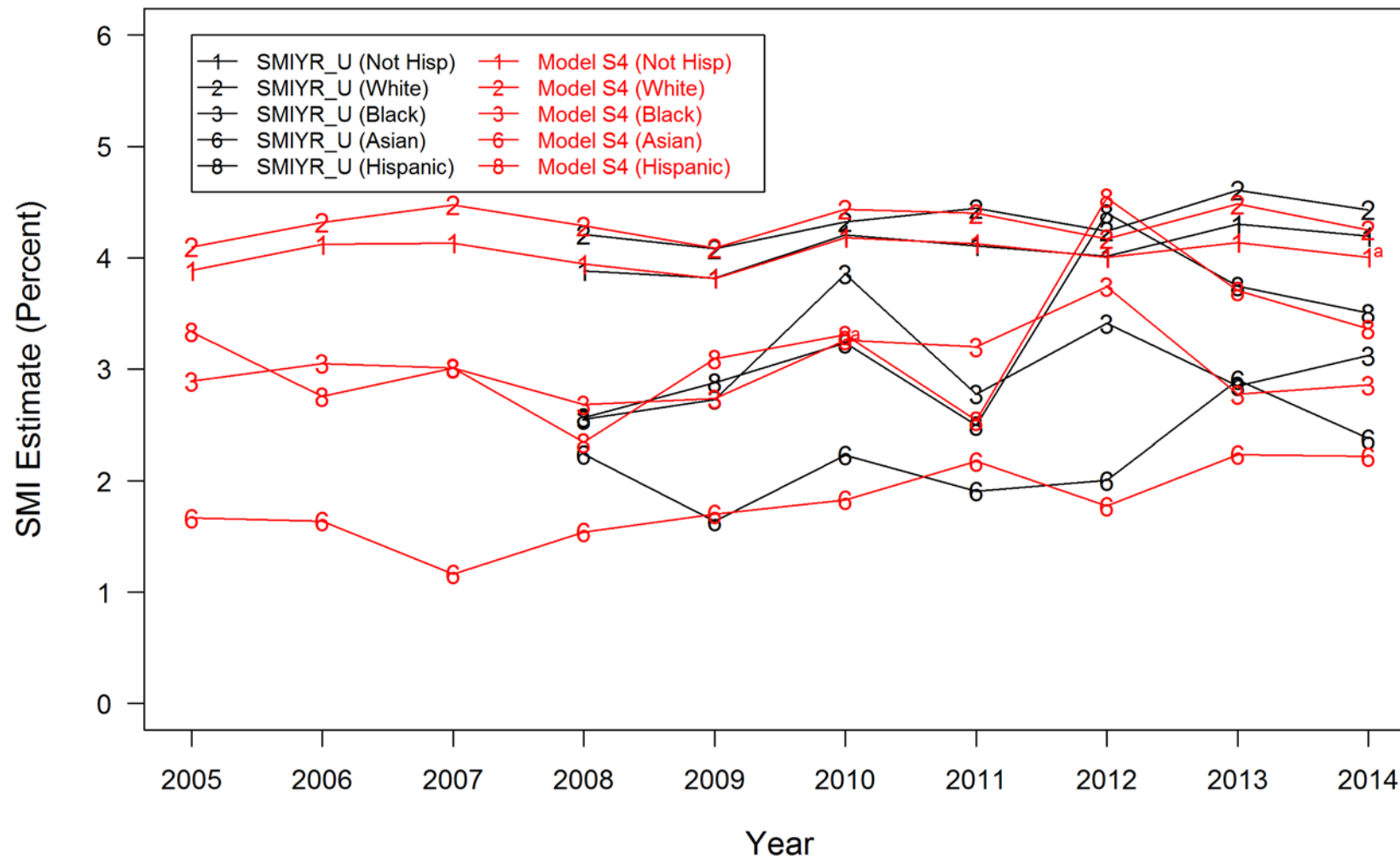


SMI = serious mental illness.

<sup>a</sup> Difference significant at .05 level. No differences attained this level of significance in this figure.

<sup>b</sup> Difference significant at .01 level.

**Figure B.4 SMI Estimates Derived from Model S4 and SMIYR\_U, by Race/Ethnicity (Five Largest Groups): 2005 to 2014**

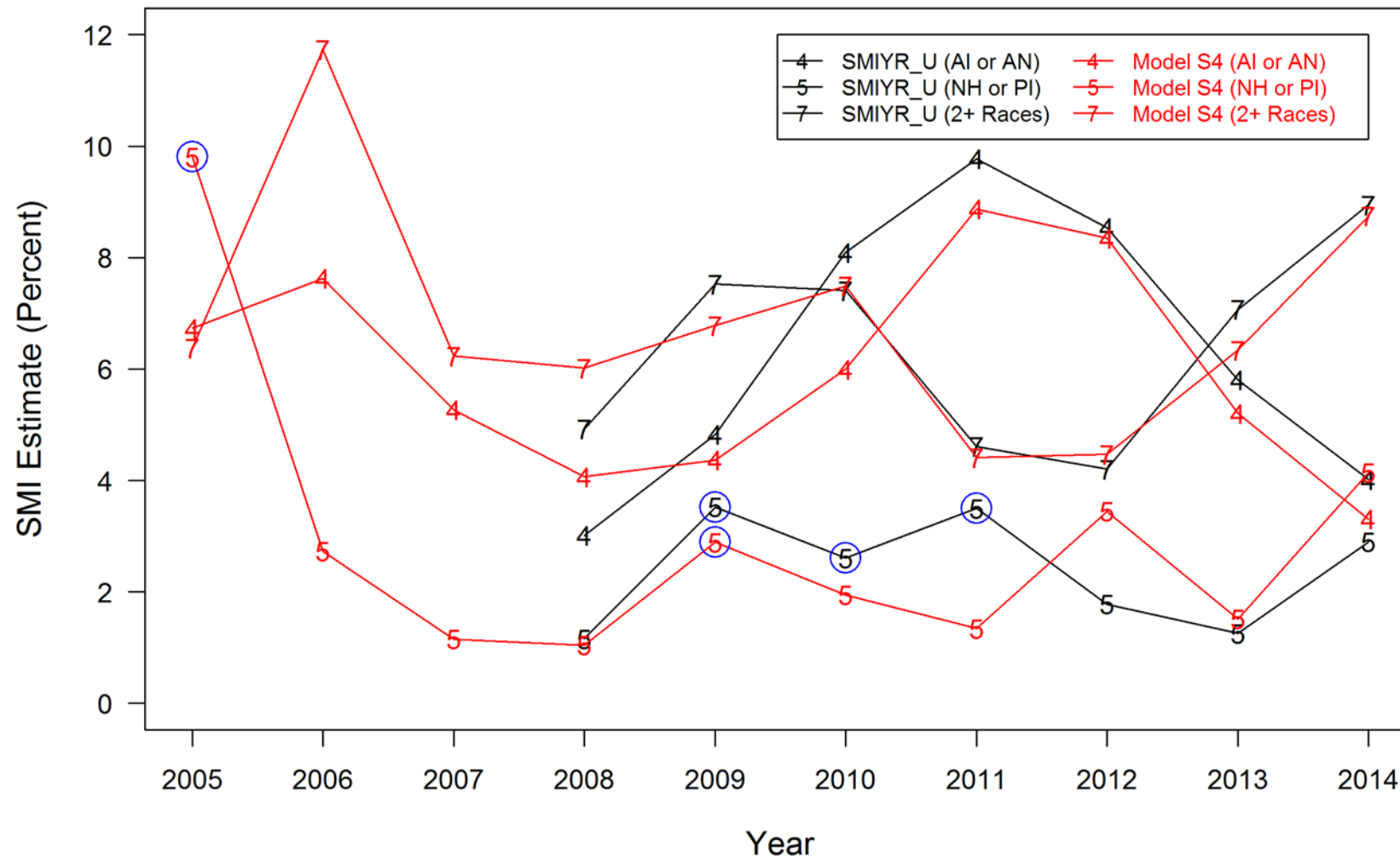


Not Hisp = not Hispanic; SMI = serious mental illness.

<sup>a</sup> Difference significant at .05 level.

<sup>b</sup> Difference significant at .01 level. No differences attained this level of significance in this figure.

**Figure B.5 SMI Estimates Derived from Model S4 and SMIYR\_U, by Race/Ethnicity (Three Smallest Groups): 2005 to 2014**



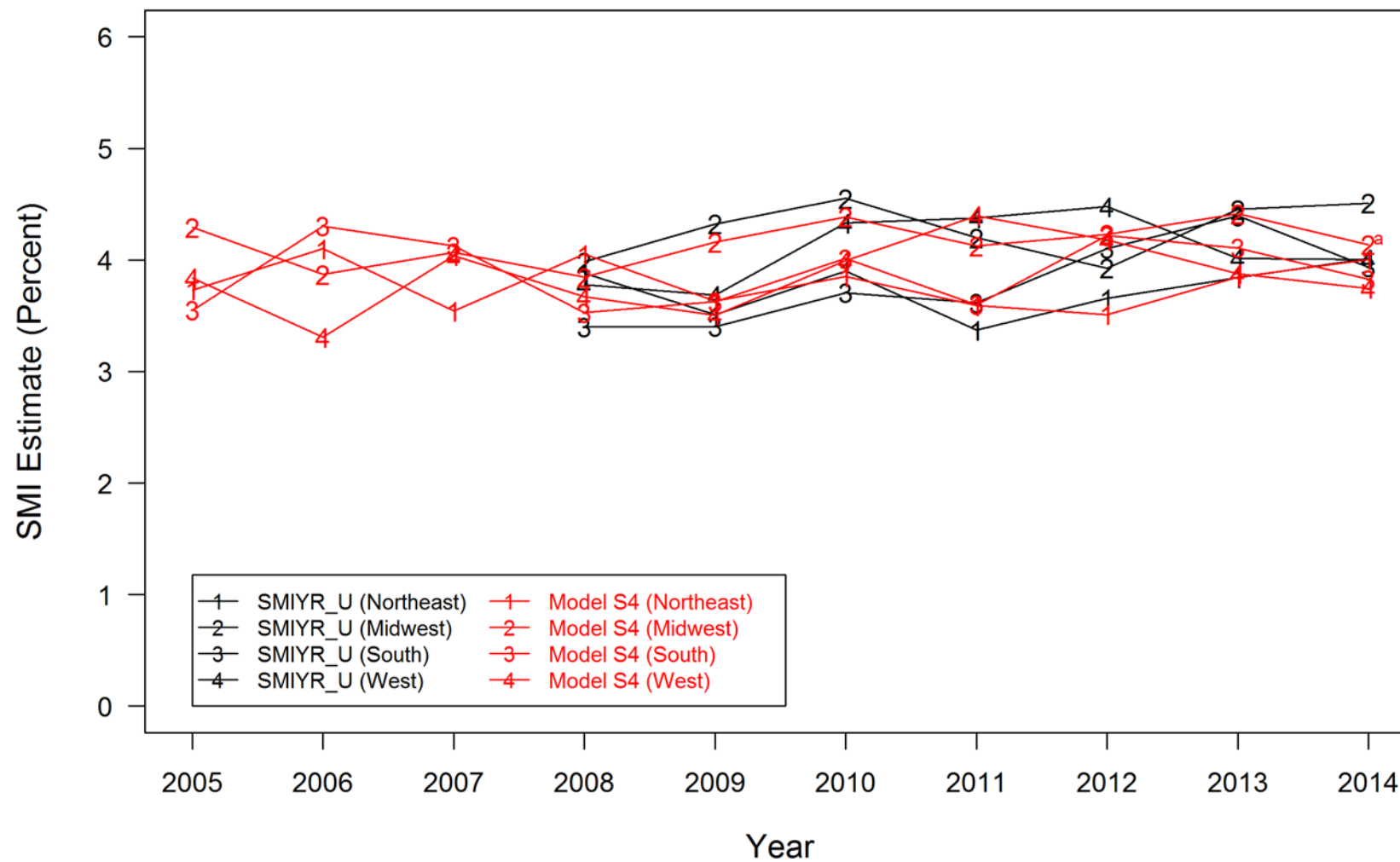
AI = American Indian; AN = Alaska Native; NH = Native Hawaiian; PI = Pacific Islander; SMI = serious mental illness.

Note: Blue-circled estimates meet the criteria for precision-based suppression.

<sup>a</sup> Difference significant at .05 level. No differences attained this level of significance in this figure.

<sup>b</sup> Difference significant at .01 level. No differences attained this level of significance in this figure.

**Figure B.6 SMI Estimates Derived from Model S4 and SMIYR\_U, by Region: 2005 to 2014**

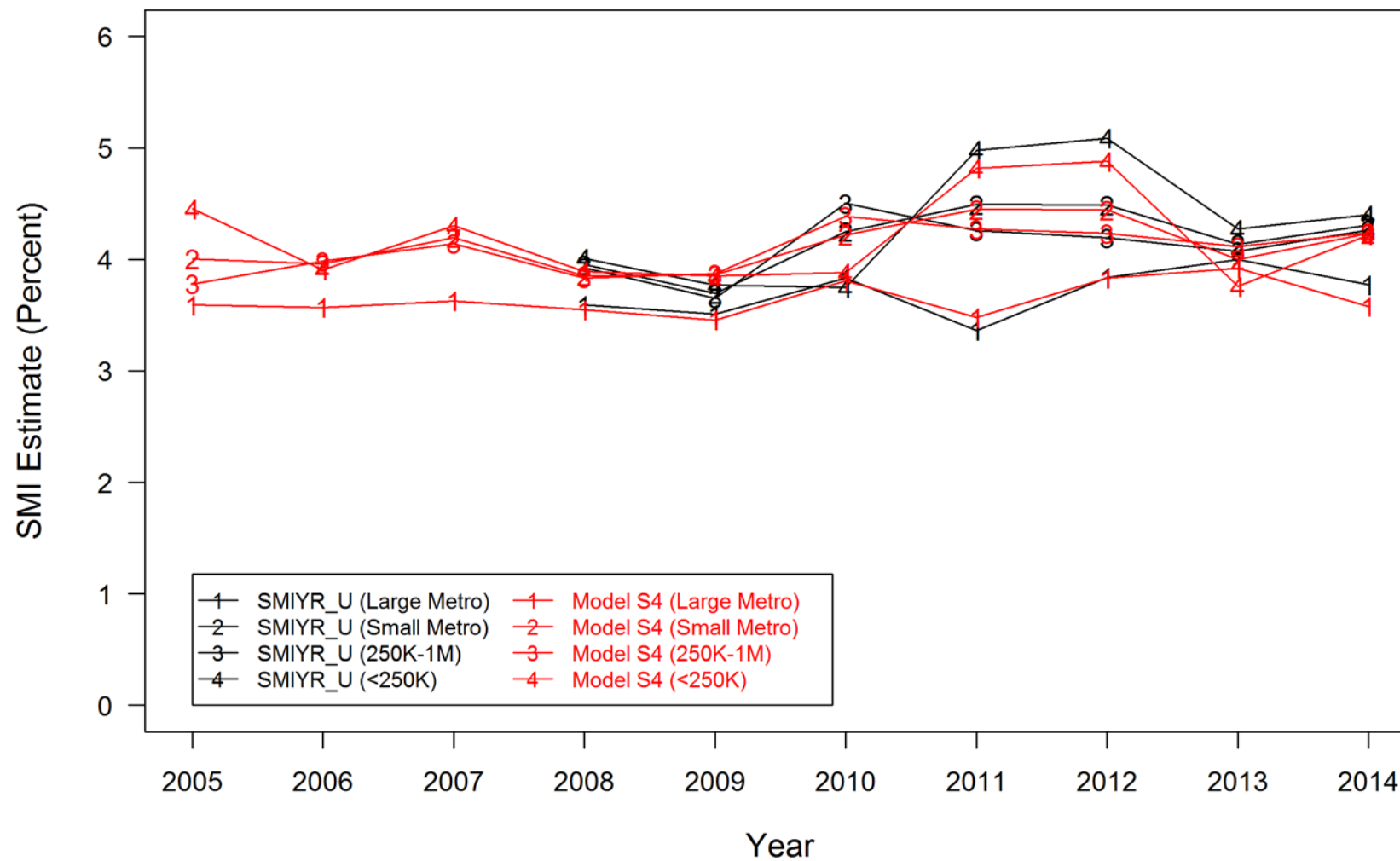


SMI = serious mental illness.

<sup>a</sup> Difference significant at .05 level.

<sup>b</sup> Difference significant at .01 level. No differences attained this level of significance in this figure.

**Figure B.7 SMI Estimates Derived from Model S4 and SMIYR\_U, by County Type (Large and Small Metro Areas): 2005 to 2014**

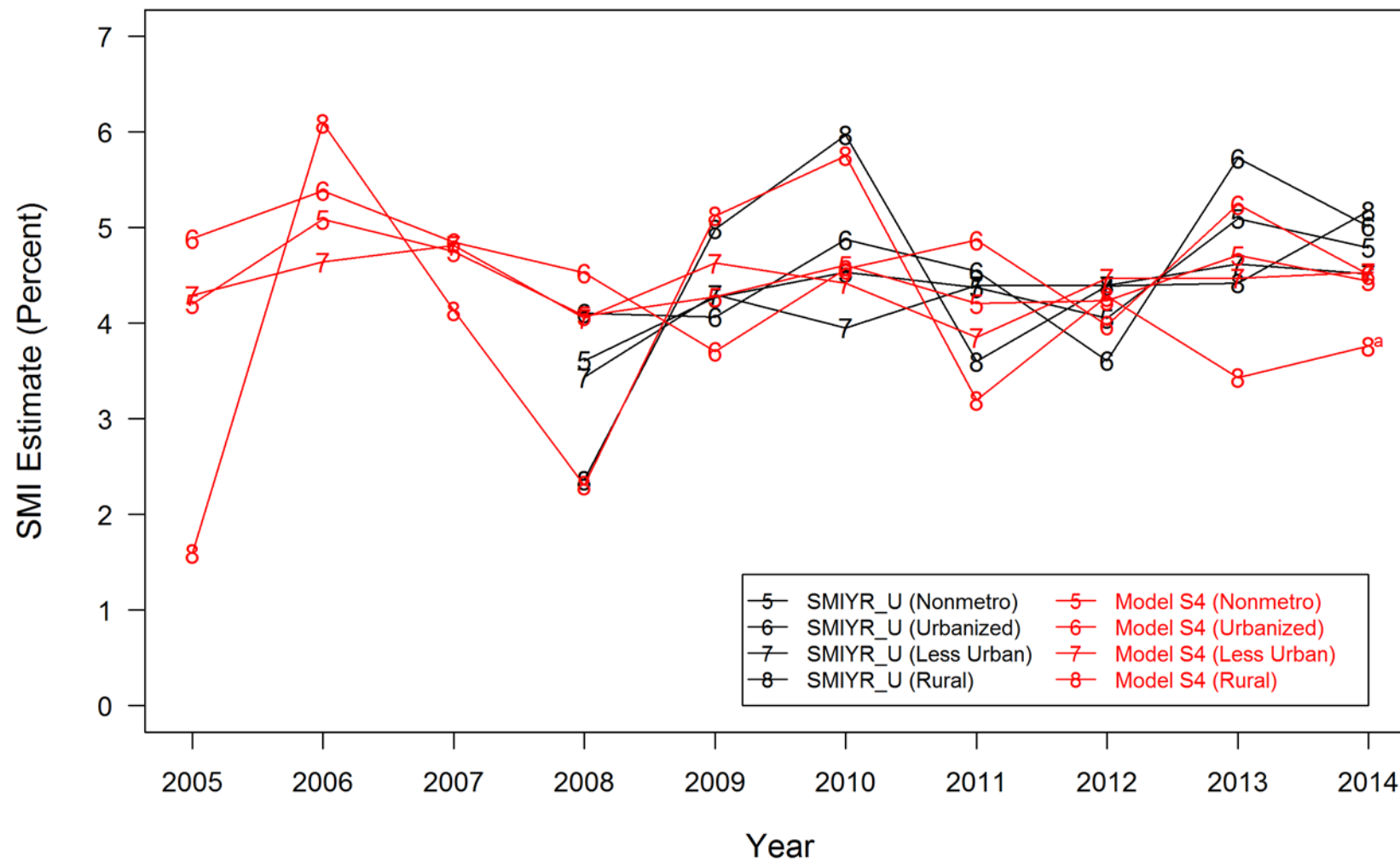


K = thousand; M = million; Metro = metropolitan; SMI = serious mental illness.

<sup>a</sup> Difference significant at .05 level. No differences attained this level of significance in this figure.

<sup>b</sup> Difference significant at .01 level. No differences attained this level of significance in this figure.

**Figure B.8 SMI Estimates Derived from Model S4 and SMIYR\_U, by County Type (Nonmetro Areas): 2005 to 2014**

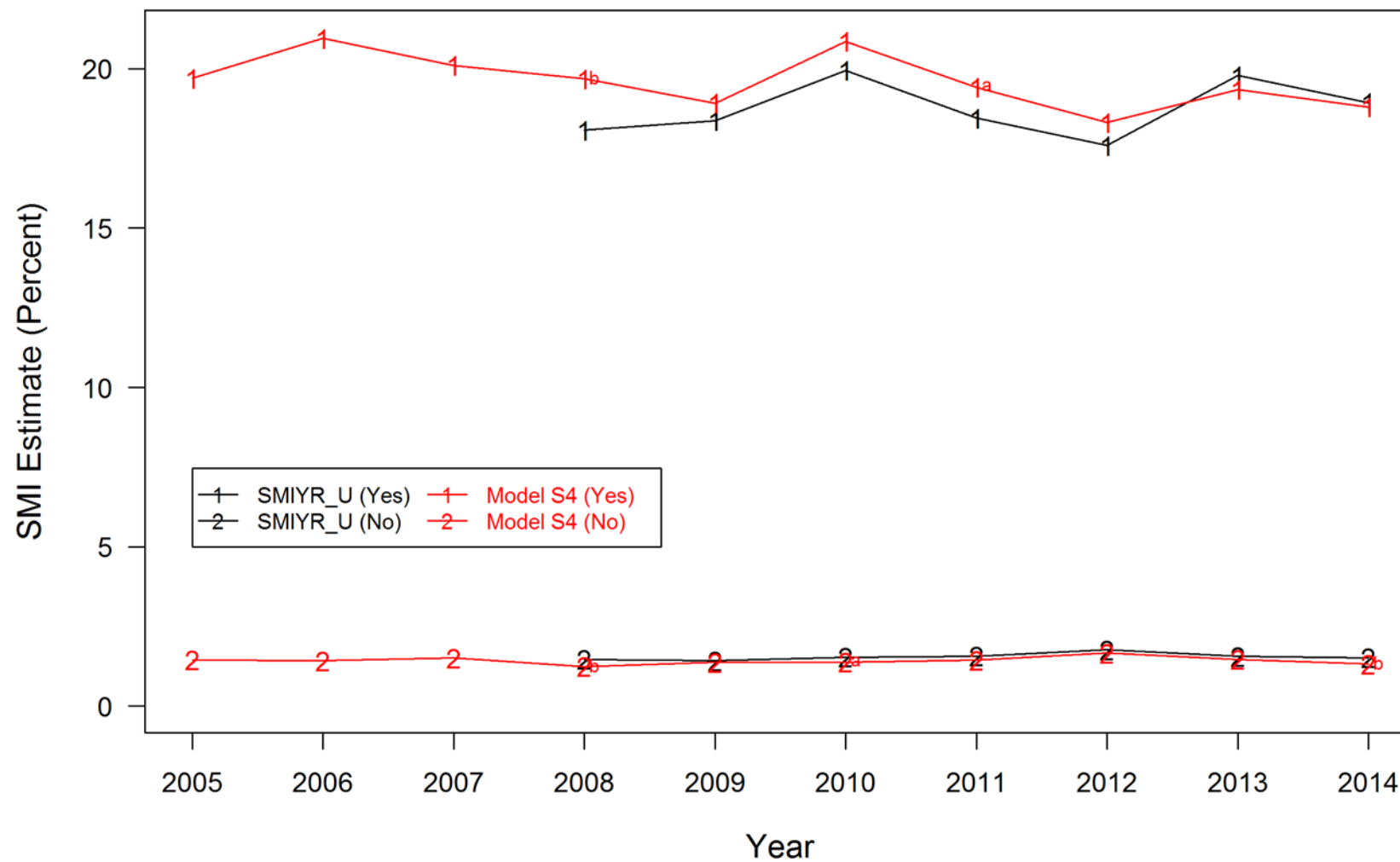


Nonmetro = nonmetropolitan; SMI = serious mental illness.

<sup>a</sup> Difference significant at .05 level.

<sup>b</sup> Difference significant at .01 level. No differences attained this level of significance in this figure.

**Figure B.9 SMI Estimates Derived from Model S4 and SMIYR\_U, by Receipt of Mental Health Services: 2005 to 2014**

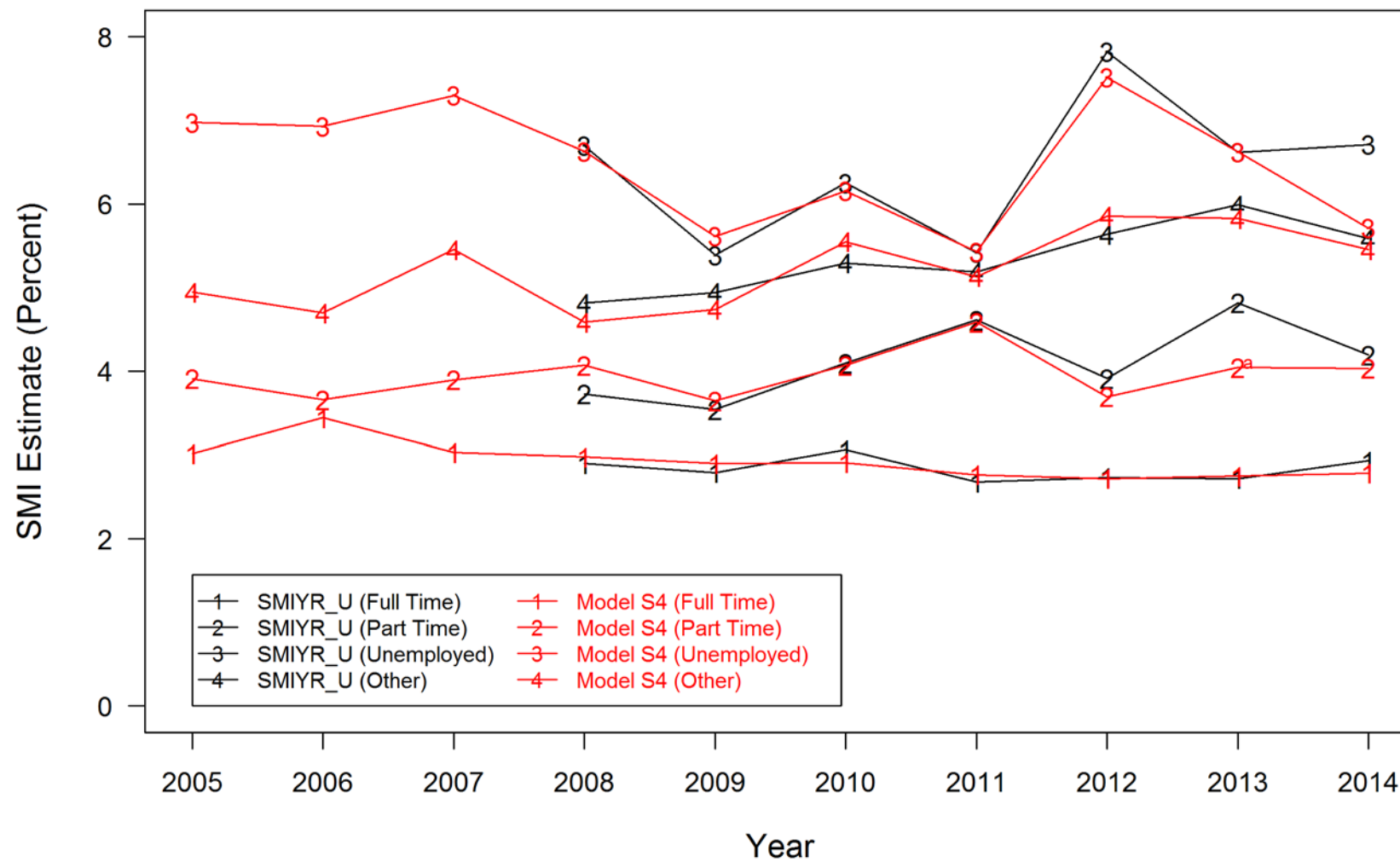


SMI = serious mental illness.

<sup>a</sup> Difference significant at .05 level.

<sup>b</sup> Difference significant at .01 level.

**Figure B.10 SMI Estimates Derived from Model S4 and SMIYR\_U, by Employment Status: 2005 to 2014**



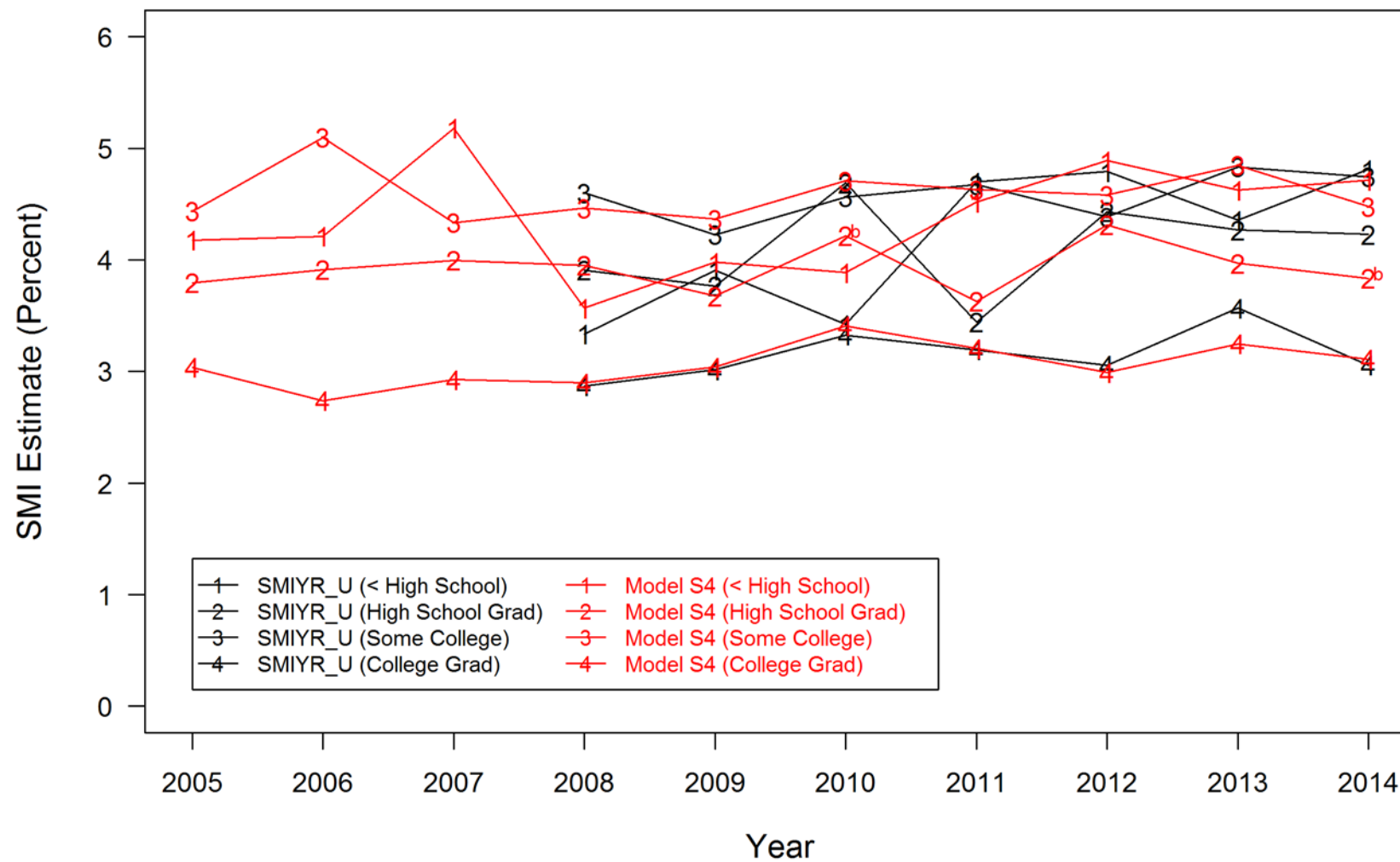
SMI = serious mental illness.

<sup>a</sup> Difference significant at .05 level.

<sup>b</sup> Difference significant at .01 level. No differences attained this level of significance in this figure.



Figure B.11 SMI Estimates Derived from Model S4 and SMIYR\_U, by Education Level: 2005 to 2014

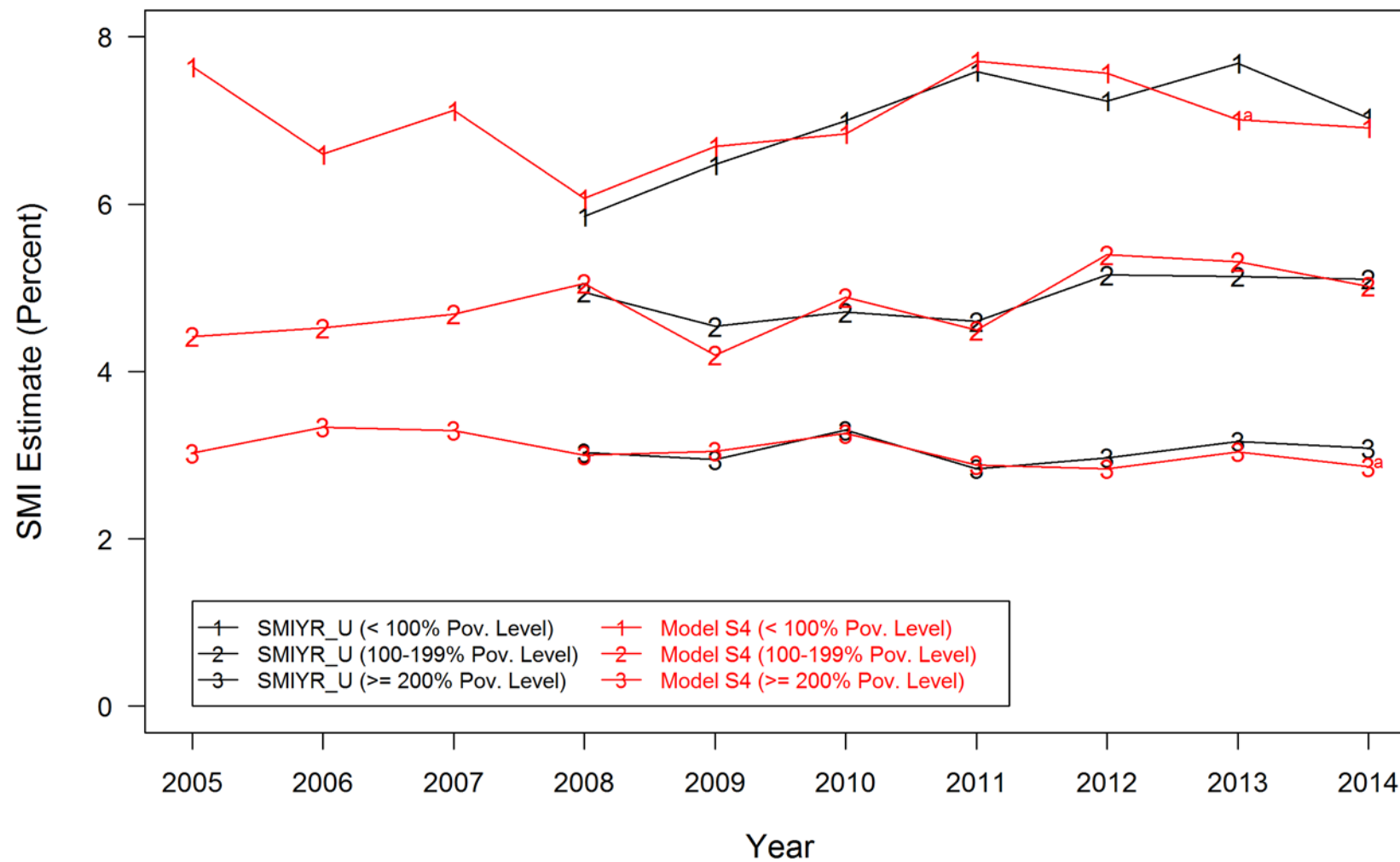


Grad = graduate; SMI = serious mental illness.

<sup>a</sup> Difference significant at .05 level. No differences attained this level of significance in this figure.

<sup>b</sup> Difference significant at .01 level.

Figure B.12 SMI Estimates Derived from Model S4 and SMIYR\_U, by Poverty Level: 2005 to 2014

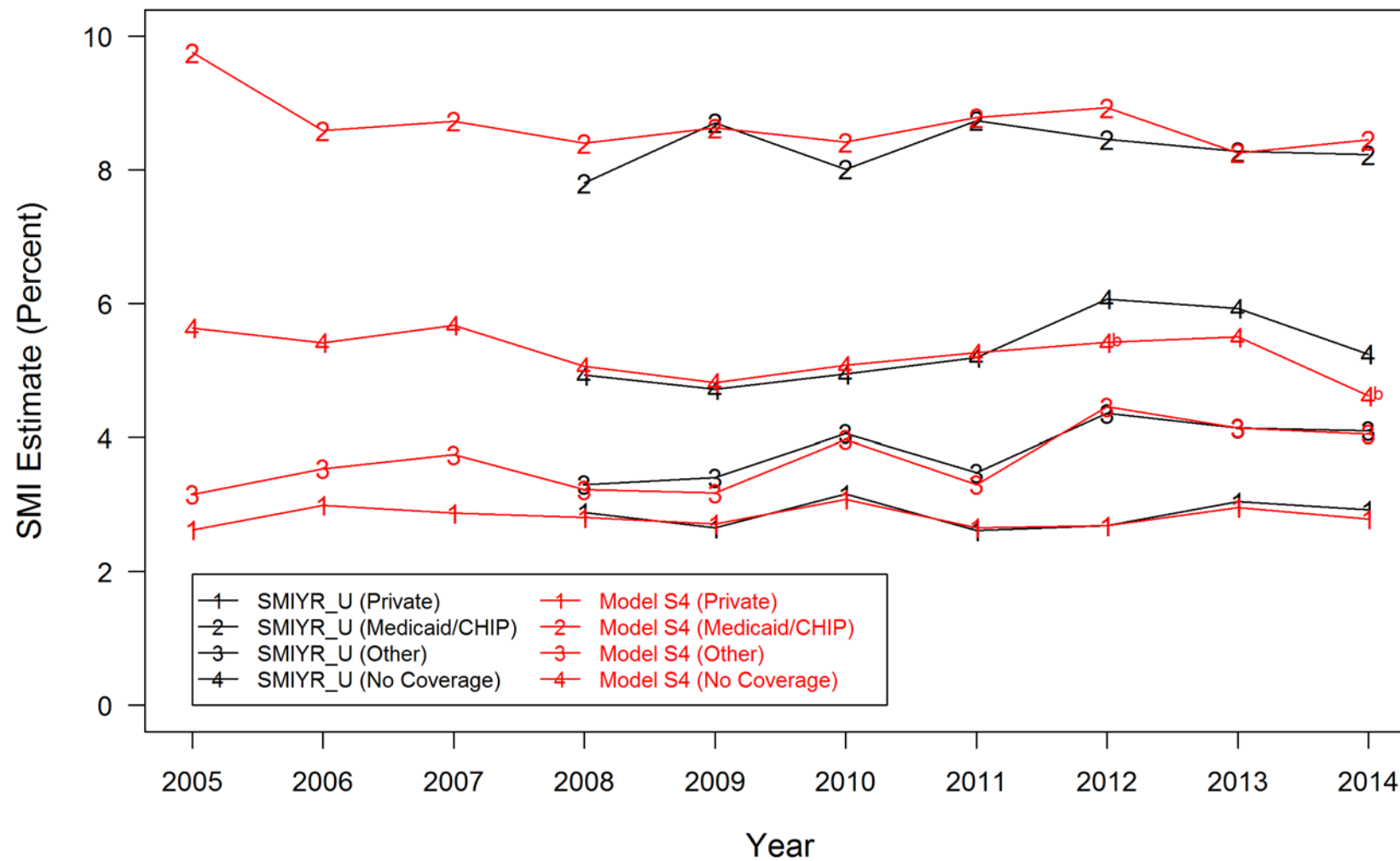


Pov. = poverty; SMI = serious mental illness.

<sup>a</sup> Difference significant at .05 level.

<sup>b</sup> Difference significant at .01 level. No differences attained this level of significance in this figure.

**Figure B.13 SMI Estimates Derived from Model S4 and SMIYR\_U, by Health Insurance Coverage: 2005 to 2014**



CHIP = Children's Health Insurance Program; SMI = serious mental illness.

<sup>a</sup> Difference significant at .05 level. No differences attained this level of significance in this figure.

<sup>b</sup> Difference significant at .01 level.

*This page intentionally left blank*