International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 19(Supplement 1): 61–87 (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/mpr.312

The National Survey on Drug Use and Health Mental Health Surveillance Study: calibration analysis

JEREMY ALDWORTH,[1] LISA J. COLPE,[2] JOSEPH C. GFROERER,[2] SCOTT P. NOVAK,[1] JAMES R. CHROMY,[1] PEGGY R. BARKER,[2] KORTNEE BARNETT-WALKER,[1] RHONDA S. KARG,[1] KATHERINE B. MORTON[1] & KATHERINE SPAGNOLA[1]

1 RTI International, Research Triangle Park, NC, USA
2 Substance Abuse and Mental Health Services Administration, Office of Applied Studies, Rockville, MD, USA

Key words
calibration, Mental Health Surveillance Study, serious mental illness

Correspondence
Lisa J. Colpe, PhD, MPH; National Institute of Mental Health, 6001 Executive Blvd, Bethesda, MD 20892, USA.
Telephone +1 301-443-3815
Email: Lisa.Colpe@NIH.gov

Received 8 March 2010; accepted 8 March 2010

Abstract

The Mental Health Surveillance Study (MHSS) is an ongoing initiative by the Substance Abuse and Mental Health Services Administration to develop and implement methods for measuring the prevalence of serious mental illness (SMI) among adults in the USA. The 2008 MHSS used data from clinical interviews administered to a sub-sample of respondents to calibrate mental health screening scale data from the National Survey on Drug Use and Health (NSDUH) for estimating the prevalence of SMI in the full NSDUH sample. The mental health scales included the K6 screening scale of psychological distress (administered to all respondents) along with two measures of functional impairment (each administered to a random half-sample of respondents): the World Health Organization Disability Assessment Schedule (WHODAS) and the Sheehan Disability Scale (SDS). The Structured Clinical Interview for DSM-IV (SCID) was administered to a sub-sample of 1506 adult NSDUH respondents within 4 weeks of completing the NSDUH interview. Results indicate that while SMI prediction accuracy of the K6 is improved by adding either the WHODAS or the SDS to the prediction equation, the models with the WHODAS are more robust. The results of the calibration study and methods used to derive prevalence estimates of SMI are presented.

Introduction

Overview

The overarching goals of the Mental Health Surveillance Study (MHSS) for the National Survey on Drug Use and Health (NSDUH) are to provide accurate estimates of the prevalence of serious mental illness (SMI) among adults (aged 18 years or older) at national and state levels and to monitor prevalence rates over time. These data are critical in determining the need for treatment and support services for this population.

On 20 May 1993, the Center for Mental Health Services of the Substance Abuse and Mental Health Services Administration (SAMHSA) published its definition of

serious mental illness in the Federal Register. Pursuant to Section 1912(c) of the Public Health Service Act, as amended by Public Law 102-321, 'adults with serious mental illness' are defined as the following:

- Persons aged 18 and over, who currently or at any time during the past year, have had diagnosable mental, behavioral, or emotional disorder of sufficient duration to meet diagnostic criteria specified within DSM-III-R [sic] that has resulted in functional impairment, which substantially interferes with or limits one or more major life activities.
- These disorders include any mental disorders (including those of biological etiology) listed in DSM-III-R or their ICD-9-CM equivalent (and subsequent revisions), with the exception of DSM-III-R 'V' codes, substance-use disorders, and developmental disorders, which are excluded unless they co-occur with other diagnosable serious mental illness.
- All of these disorders have episodic, recurrent, or persistent features; however, they vary in terms of severity or disabling effects. Functional impairment is defined as difficulties that substantially interfere with or limit role functioning in one or more major life activities including basic daily living skills (e.g. eating, bathing, dressing); instrumental living skills (e.g. maintaining a household, managing money, getting around the community, taking prescribed medication); and functioning in social, family, and vocational/educational contexts.
- Adults who would have met functional impairment criteria during the referenced year without benefit of treatment or other support services are considered to have serious mental illnesses.

In December 2006, a technical advisory group (TAG) meeting of expert consultants was convened by Center for Mental Health Services to solicit recommendations for mental health surveillance data collection strategies among the US population. The panel recommended that the NSDUH should be used to make estimates of SMI among adults and that SAMHSA should conduct methodological studies to calibrate the NSDUH's mental health items with a gold-standard clinical psychiatric interview. In response, SAMHSA's Office of Applied Studies initiated the MHSS under its NSDUH contract with RTI International (a trade name of Research Triangle Institute) to develop and implement the methods for SMI estimation. At the time, the NSDUH contained a six-item scale (K6) with five response options in each item that captured information on psychological distress (Kessler *et al.*, 2003). However, the K6 scale does not capture information on functional impairment, which is needed to define cases as meeting the SAMHSA definition for SMI. In consultation with the TAG, two candidate impairment scales were selected by SAMHSA to be added to the 2008 NSDUH to obtain such data. They are the World Health Organization Disability Assessment Schedule (WHODAS) (Rehm *et al.*, 1999) and the Sheehan Disability Scale (SDS) (Leon *et al.*, 1997). An initial step of the MHSS was to modify these scales for use in a general population survey, including changes to question wording and length (Novak *et al.*, 2010, this issue).

Primary objectives

The MHSS calibration study analysis had two primary objectives:

1 To determine which of the two disability scales, used in combination with the K6 scale, provides the more accurate prediction of SMI in the NSDUH and will therefore be administered to the entire sample of adults in the 2009 and later surveys.
2 To develop algorithms that exhibit sound psychometric properties and will accurately classify NSDUH respondents as meeting or not meeting criteria for SMI. These algorithms were to be used to produce 2008 estimates of SMI prevalence.

Materials and methods

The MHSS sample

In the 2008 NSDUH, a split-sample design was used where all adult respondents received the K6, but a random half of the sample received the WHODAS and the other half received the SDS. In addition, a sub-sample of approximately 1500 adult NSDUH participants was recruited for a follow-up clinical interview to provide data for calibration of the NSDUH full-sample interview data on mental health status. The randomization of the impairment scales was maintained within this clinical interview sub-sample, which we refer to as the MHSS calibration sample, so that about half of the MHSS calibration sample participants (approximately 750) were administered the WHODAS and the other half were administered the SDS. A diagram illustrating the structure of the MHSS sampling design is given in Figure 1.

The MHSS sample was stratified, based on respondents' K6 scores in 2008, to optimize the MHSS sample allocation for calibration modeling. Strata were constructed according to seven scoring bands described in

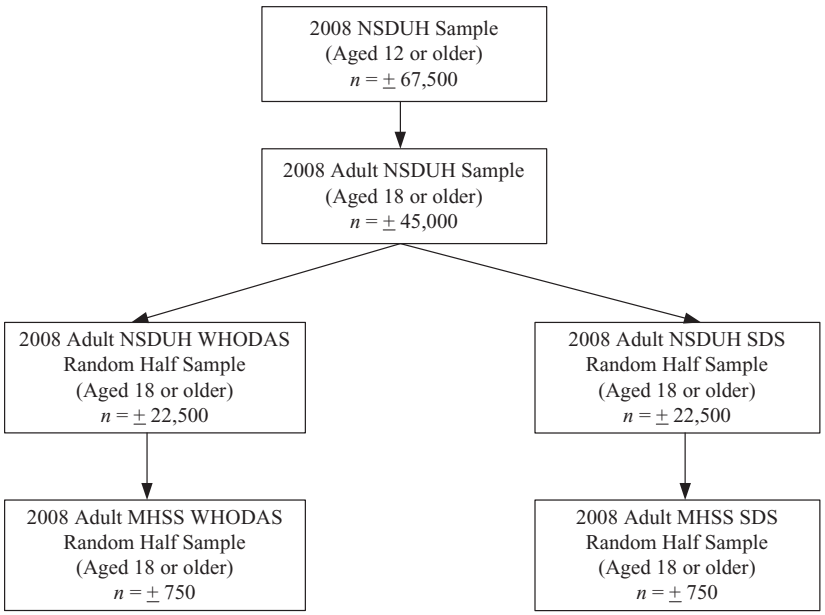


Figure 1 Structure of Mental Health Surveillance Study sampling design.

Table 1 Mental Health Surveillance Study sample allocation (N = 1500)

K6 score	Percent of population[1]	Assumed SMI rate (%)	Expected sample size	Expected SMI count	Sampling rate (%)
0 to 3	48.04	0.03	96	0	0.0084
4 to 5	13.98	0.30	88	0	0.0228
6 to 7	11.16	0.30	110	0	0.0345
8 to 9	6.95	10.00	200	20	0.1026
10 to 11	5.53	13.00	214	28	0.1190
12 to 15	8.00	40.00	450	180	0.1689
16 or higher	6.34	67.00	343	230	0.1349
TOTAL	100.00	8.95	1501	458	

K6 = six-item psychological distress scale, SMI = serious mental illness.
[1] Source: 2006 National Survey on Drug Use and Health.

Table 1. Assumed SMI rates were estimated using K6 score distribution data from the 2006 NSDUH and raw K6 score and clinical case data from the National Comorbidity Survey Replication clinical calibration study. Sampling rates for the 2008 study were substantially lower for K6 scores 0 to 7 under the assumption that fewer clinical positives would be identified in that scoring range. Table 1 shows the expected sample distribution for the 1500 clinical follow-up interviews and the expected number of those with positive SMI status. The design effect for a prevalence estimate of SMI due to this two-phase sample stratified by K6 scores is 0.2121 (i.e. the variance is reduced almost five-fold in comparison to a simple random sample). The usual design effect for adults in the main survey is approximately 3.0 (e.g. for the prevalence of serious psychological distress), so the overall design effect for the MHSS sample is estimated to be 0.6363. The effective sample size is therefore approximately 2357, and the projected standard error and relative standard error of an estimate of SMI are 0.59% and 6.57%, respectively. The overall expected proportion of positive SMI counts is 0.305.

The probability sample of 1500 clinical follow-up interviews was distributed across four calendar quarters with a slightly larger sample in the first quarter (425 follow-up interviews) and the remaining sample equally divided among the remaining quarters (approximately 358 interviews in each of Quarters 2 through 4 for a combined sample of 1075 follow-up interviews). The intention of the larger sample in Quarter 1 was to provide some cushion in case clinical interview response rates were lower than anticipated and to generate an adequate sample size for the 6-month analysis. The agreement rate for the clinical follow-up interview was projected to be 85% and the participation rate among those who agreed to complete the interview was projected to be 90%.

The unweighted and weighted response rates for each of the seven K6 score categories are given in Table 2. The unweighted response rates are fairly evenly balanced between the two half-samples, but there appear to be some unbalanced K6 score categories for the weighted response rates, particularly in the '4 to 5' and '6 to 7' categories.

The K6 scale

The K6 scale, used to capture non-specific psychological distress (Kessler *et al.*, 2003), consists of two sets of six questions that ask respondents how frequently they experienced symptoms of psychological distress during two different time periods: during the past 30 days and the 1 month in the previous 12 months when they were at their worst emotionally. Respondents were only asked about the second time period if they indicated that there was a month in the past 12 months when they felt more depressed, anxious, or emotionally stressed than they felt during the past 30 days. The six domains covered by the questions corresponded to how often the respondents felt nervous, hopeless, restless or fidgety, sad or depressed, worthless, and that everything was an effort during the target time periods. To create a score, the six items related to the first time period were coded from 0 to 4 so that 'all of the time' was coded 4, 'most of the time' 3, 'some of the time' 2, 'a little of the time' 1, and 'none of the time' 0, with 'don't know' and 'refuse' also coded 0. Summing across the six responses resulted in a total score with a range from 0 to 24. The six items related to the second time period were coded similarly, and the worst K6 total score was calculated as the higher of the total scores from the two time periods. An alternative version of the worst K6 total score was formulated as follows: worst K6 total scores less than 8 were recoded as 0, and worst K6 total scores between 8 and 24 were recoded as 1 to 17. The reason behind the alternative version was that SMI prevalence was typically extremely low for respondents with worst K6 total scores less than 8, and the prevalence rates only started increasing once total scores were 8 or greater. Therefore, a score band of 0–17 collapsed the less informative lower scores into a one-score category (0) while preserving the more informative scores at the higher end of the scale (1–17). See Appendix A for further details of the actual questions in the K6.

Table 2 Response rates (unweighted and weighted) by K6 score category

	Sample A (WHODAS)				Sample B (SDS)			
K6 score	Number selected	Number completed	URR (%)	WRR (%)	Number selected	Number completed	URR (%)	WRR (%)
0 to 3	83	51	61.5	55.2	80	46	57.5	46.9
4 to 5	77	54	70.1	62.0	69	47	68.1	78.0
6 to 7	77	49	63.6	59.3	81	59	72.8	77.4
8 to 9	161	103	64.0	61.8	163	108	66.3	53.8
10 to 11	156	106	68.0	67.1	140	102	72.9	76.3
12 to 15	331	225	68.0	64.2	341	218	63.9	60.3
16 or higher	289	173	59.9	58.0	243	161	66.3	58.0
Total	1174	761	64.8	58.5	1117	741	66.3	58.3

K6 = six-item psychological distress scale, SDS = four-item Sheehan Disability Scale, URR = unweighted response rate, WHODAS = eight-item World Health Organization Disability Assessment Schedule, WRR = weighted response rate.
Note: This table excludes four cases from the MHSS sample because of unusual weights or because all mental health item scores were missing.

Int. J. Methods Psychiatr. Res. 19(Supplement 1): 61–87 (2010). DOI: 10.1002/mpr

The impairment scales

An abbreviated version of the WHODAS (Rehm *et al.*, 1999), as described in this issue by Novak *et al.* (2010), was used to assess impairment in one half of the sample. This version consists of eight questions that ask respondents how much their emotions, nerves, or mental health caused them to have difficulties in daily activities during the 1 month in the past year when they were at their worst emotionally. The following eight domains were covered by the questions:

1 remembering to do things they needed to do
2 concentrating on doing something important when other things were going on around them
3 going out of the house and getting around on their own
4 dealing with people they did not know well
5 participating in social activities
6 taking care of household responsibilities
7 taking care of daily responsibilities at work or school
8 getting daily work done as quickly as needed.

To create a score, the eight items were coded from 0 to 3 so that 3 represented 'severe difficulty', 2 was 'moderate difficulty', 1 was 'mild difficulty', and 0 was 'no difficulty', with 'don't know' and 'refuse' also coded 0. Some items had a fifth category to deal with 'not applicable' responses. For example, the question about difficulties taking care of daily responsibilities at work or school had a fifth category, 'you didn't go to work or school'. If this category was selected, respondents were asked if their emotions, nerves, or mental health kept them from going to work or school. A 'yes' response was coded 3, and a 'no' was coded 0. Summing across the eight responses resulted in a total score with a range from 0 to 24. An alternative version of the WHODAS total score was formulated as follows: item scores less than 2 were recoded as 0, and item scores 2 to 3 were recoded as 1 and then summed for a total score ranging from 0 to 8. The alternative version of the WHODAS total score was created because of the expectation that a dichotomous measure dividing respondents who experienced moderate or severe difficulties from the remaining respondents would fit better than a linear continuous measure. See Appendix B for further details of the actual questions in the WHODAS.

The SDS (Leon *et al.*, 1997) was used to assess impairment in the other half of the sample. This scale consists of four questions that ask respondents how much their emotions, nerves, or mental health interfered with their daily activities over the past year. Four domains were covered by the questions: (1) home management, (2) work, (3) close relationships with others, and (4) social life. For each of the four items, respondents were asked to select a number from 0 to 10 on a visual analog scale, where 0 means no interference, 1 to 3 means mild interference, 4 to 6 means moderate interference, 7 to 9 means severe interference, and 10 means very severe interference. Summing across the four responses resulted in a total score with a range from 0 to 40. An alternative version of the SDS total score was formulated as follows: item scores less than 7 were recoded as 0, and item scores 7 to 10 were recoded as 1, and then summed for a total score ranging from 0 to 4. The alternative version of the SDS total score was also created because of the expectation that a dichotomous measure dividing respondents who experienced severe or very severe interference from the remaining respondents would fit better than a linear continuous measure. See Appendix C for further details of the actual questions in the SDS.

The clinical follow-up interview

Each participant in the MHSS calibration study subsample was administered standard clinical interview measures by mental health clinicians via paper-and-pencil interviewing over the telephone. The clinical interview instrument used was the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders edition 4 (DSM-IV-TR) Axis I Disorders Non-Patient Edition (SCID) (First *et al.*, 2002), adapted by its senior author to have a 12-month reference period for this study. Functional impairment ratings were assigned by clinical interviewers using the Global Assessment of Functioning (GAF) scale. A respondent was coded positive for SMI if he or she was determined to have any of the mental disorders assessed in the MHSS SCID and a GAF score of 50 or below in the past 12 months.

Analysis methods

The analysis was based on SCID data derived from the clinical interviews completed in the 2008 survey and corresponding data from the computer-assisted interviewing part of the main survey. A sample of 1506 respondents completed both the clinical interview and the corresponding data from the main survey. However, a consequence of the sample design was that respondents with low K6 total scores typically had relatively large weights, and three records with unusually large weights that had the effect of unduly influencing the ROC models were removed from the dataset. One record with missing data on all K6 items and all SDS items was also removed, leaving 1502 analyzable records in the data file used for

the calibration analysis. Data from the computer-assisted interviewing were lightly edited, except for demographic variables, which went through the thorough editing and imputation processes typically implemented at the end of a survey year. The analysis weights for the MHSS included the following weight components: main study analysis weight, inverse of probability of selection for clinical follow-up, non-response adjustment for clinical interview (the four completed cases removed from the analysis were added to the set of non-respondents), and post-stratification adjustments to the 1502 analyzable cases by gender, age, and race/ethnicity using interview data. Because the SCID dataset was much smaller than the full adult NSDUH dataset, the 2008 NSDUH stratification and clustering design variables had to be appropriately collapsed to accommodate this smaller dataset.

Descriptive analyses were conducted to examine the distribution of respondent characteristics in the MHSS to check for imbalances between the two half-samples, each of which was assigned to one of the impairment scales. Modeling analyses were conducted to develop algorithms based on the K6 scale and each of the impairment scales in turn, with the goal of identifying the best possible model for each impairment scale. This involved fitting a variety of models using alternative predictors, including different forms of the K6 and impairment variables. For each model, receiver operating characteristic (ROC) analyses were conducted to select the optimal cut-point for determining SMI status. Weighted counts were used in the ROC classifications because primary interest is in estimating SMI status in the adult US population. Models to determine SMI were compared and evaluated based on three criteria:

1 *model robustness* (e.g. preference given to parsimonious models that could be generalized to data beyond that used in the modeling process)
2 *minimization of misclassification errors in SMI prediction* (i.e. exhibiting reasonable ROC statistics, such as sensitivity and AUC, defined as the area under the ROC curve based on the optimal cut-point described above)
3 *reasonable SMI estimates based on the full dataset* (i.e. balanced across several demographic sub-groups and across the WHODAS and the SDS half-samples).

Preliminary analysis

A preliminary analysis was conducted using the first 6 months of data collected to achieve the first primary objective of the MHSS calibration; that is, to determine which of the two disability scales, used in combination with the K6 scale, provides the more accurate prediction of SMI in the NSDUH in time to be applied to the 2009 survey. The methods used in the preliminary analysis were broadly similar to those used in the final analysis, and the result was that the WHODAS was chosen in favor of the SDS to act in combination with the K6 scale to predict SMI for the 2009 NSDUH. For more details please refer to Aldworth *et al.* (2008).

Results of final analysis

Descriptive analyses

Initial descriptive analyses and statistical tests were conducted to check for imbalances in key demographic, mental health, and substance-use characteristics between the two half-samples assigned to either of the two impairment scales. Key demographic characteristics included gender, age, race/ethnicity, and education; mental health characteristics included lifetime and past-year depression, depression treatment, mental health treatment, and suicidality measures; mental health characteristics from the SCID included SMI status and diagnoses of mental health and substance use, and substance use characteristics included past-month tobacco and marijuana use and past-year alcohol abuse or dependence.

Unweighted descriptive statistics of the demographic, mental health, substance use, and SCID mental health variables were derived. These statistics showed that the selection of females for the SCID was disproportionately high when compared with the 12-month NSDUH sample; that the prevalence of mental health problems and substance use was higher in the SCID sub-sample than in the NSDUH sample, as would be expected given the SCID selection process; and that none of the demographic or mental health measures appeared to be seriously unbalanced between the two half-samples, although there was evidence of an imbalance with respect to past-month cigarette and marijuana use. Weighted descriptive statistics of the same variables are shown in Tables 3, 4, and 5. Included in the descriptive statistics are frequencies and percentages of the entire 12-month NSDUH sample, the subset of respondents selected for the SCID, and the analyzed subset of those that completed the SCID. Chi-square tests were conducted to compare the analyzable SCID cases between the two half-samples. Table 3 shows some imbalance with respect to gender between the analyzable cases in the half-samples, but it is not statistically significant. Table 5 shows statistically significant imbalances with respect to some of the substance use measures. Other measures in Tables 3, 4, and 5 indicate minor imbalances, but none is statistically significant.

Table 3 Weighted[1] descriptive statistics of demographic characteristics (percentages)

		Selected SCID cases		Analyzable SCID cases				
Variable	2008 12-month NSDUH cases[2]	Sample A (WHODAS)	Sample B (SDS)	Sample A (WHODAS)	Sample B (SDS)	Total	χ^2 (DF)	*P*-value
Gender								
Male	45.7	45.7	46.0	51.4	45.1	48.3	0.52 (1)	0.472
Female	54.3	54.3	54.0	48.6	54.9	51.7		
Race/ethnicity								
White, NH	67.5	68.0	72.7	64.6	73.1	68.8	0.87 (3)	0.459
Black, NH	11.1	10.9	8.0	12.5	10.1	11.3		
Other	6.4	9.0	9.4	9.9	3.0	6.4		
Hispanic	15.0	12.1	10.0	13.1	13.9	13.5		
Age								
18–25	14.4	14.2	14.1	14.6	14.7	14.6	0.34 (2)	0.716
26–49	45.0	43.7	45.6	41.2	47.6	44.4		
50+	40.6	42.1	40.3	44.2	37.7	41.0		
Education								
<High School	16.2	11.1	8.5	8.4	9.7	9.1	0.27 (3)	0.848
High School graduate	31.5	31.8	38.2	30.7	35.6	33.1		
Some college	25.5	29.8	22.6	30.4	25.6	28.0		
College graduate	26.8	27.3	30.7	30.5	29.0	29.8		

DF = degrees of freedom, NH = non-Hispanic, SCID = Structural Clinical Interview for DMS-IV, SDS = four-item Sheehan Disability Scale, WHODAS = eight-item World Health Organization Disability Assessment Schedule.
[1]The overall NSDUH analysis weight was used for the NSDUH cases. The overall NSDUH analysis weight multiplied by the inverse of the SCID selection probability was used for the selected SCID cases. The MHSS Combined Unadjusted Sample Weight was used for completed SCID cases. The MHSS Combined Unadjusted Sample Weight included the following weights: overall NSDUH analysis weight; inverse of the SCID selection probability; non-response adjustment for clinical interview; and post-stratification adjustments by gender, race/ethnicity, and age.
[2]This includes all cases for persons aged 18 or older.

Both unweighted and weighted distributions of worst K6 total score (i.e. maximum of past 30-day K6 total score and worst-month K6 total score) were reasonably balanced between half-samples. As a result, no post-stratification adjustments in the weights were deemed to be necessary.

Specifications of modeling analyses

The process of selecting models began by developing a series of weighted logistic regression prediction models for the K6 and each of the two impairment scales, respectively. With SMI status based on having a SCID diagnosis plus a GAF ≤ 50, we defined the response variable Y so that $Y = 1$ when an SMI diagnosis is positive; otherwise, $Y = 0$. If $\mathbf{X}$ is a vector of explanatory variables, then we can estimate the response probability $\pi = \Pr(Y = 1 \mid \mathbf{X})$ using the following logistic regression models for the WHODAS and SDS half-samples, respectively:

$$\begin{aligned}\text{logit}(\pi_w) &\equiv \log(\pi_w/(1-\pi_w)) \\ &= \beta_{w0} + \sum_{i=1}^{n_k}\beta_{ki}X_{ki} + \sum_{j=1}^{n_w}\beta_{wj}X_{wj} \end{aligned} \quad (1)$$

$$\text{logit}(\pi_s) = \beta_{s0} + \sum_{i=1}^{n_k}\beta_{ki}X_{ki} + \sum_{j=1}^{n_s}\beta_{sj}X_{sj} \quad (2)$$

where the X_{ki}, X_{wj}, and X_{sj} terms refer to K6, WHODAS, and SDS terms, respectively. All of the models tested were of this general form. Some models included 'total score' variables, which combined items from a scale into one value, while other models included 'item score' variables, where separate variables represented individual items from the scale. For example, if worst K6 total score is included in the model instead of terms for individual items, then there will be only one X_{ki} term (i.e. $n_k = 1$, and the summation reduces to $\beta_k X_k$). However, if the six individual K6 items are included in the model, then there will be six X_{ki} terms corresponding to those items. A similar

Table 4 Weighted[1] descriptive statistics of mental health characteristics (percentages)

Variable	2008 12-month NSDUH cases[2]	Selected SCID cases: Sample A (WHODAS)	Selected SCID cases: Sample B (SDS)	Analyzable SCID cases: Sample A (WHODAS)	Analyzable SCID cases: Sample B (SDS)	Analyzable SCID cases: Total	χ^2 (DF)	*P*-value
Depression								
LT, but not PY	6.5	4.8	6.0	5.6	7.2	6.4	0.26 (3)	0.854
PY, no imp	2.4	2.6	2.5	2.8	2.3	2.6		
PY, with imp	4.5	3.6	4.0	3.8	4.3	4.0		
No occurrence	86.5	89.1	87.5	87.8	86.2	87.0		
Depression TX, PY								
Tx, non-med	44.6	45.7	54.9	45.0	54.0	49.8	0.67 (1)	0.415
Tx, med	40.7	44.1	50.7	42.2	45.8	44.1	0.13 (1)	0.714
Tx, any	52.7	58.4	63.1	58.3	62.1	60.3	0.13 (1)	0.718
MH TX, PY								
Tx, outpatient	6.8	8.2	9.3	7.2	8.1	7.6	0.13 (1)	0.722
Tx, inpatient	0.8	0.6	0.8	0.3	0.2	0.3	0.22 (1)	0.640
Tx, med	11.5	11.2	13.7	11.8	10.6	11.2	0.14 (1)	0.710
Tx, any	13.5	14.0	16.5	13.7	13.2	13.4	0.02 (1)	0.883

DF = degrees of freedom, imp = impairment, LT = lifetime, med = medication, PY = past year, SCID = Structural Clinical Interview for DMS-IV, SDS = four-item Sheehan Disability Scale, TX or Tx = treatment, WHODAS = eight-item World Health Organization Disability Assessment Schedule.

[1]The overall NSDUH analysis weight was used for the NSDUH cases. The overall NSDUH analysis weight multiplied by the inverse of the SCID selection probability was used for the selected SCID cases. The MHSS Combined Unadjusted Sample Weight was used for completed SCID cases. The MHSS Combined Unadjusted Sample Weight included the following weights: overall NSDUH analysis weight; inverse of the SCID selection probability; non-response adjustment for clinical interview; and post-stratification adjustments by gender, race/ethnicity, and age.

[2]This includes all cases for persons aged 18 or older.

procedure follows for the WHODAS and SDS terms. The beta coefficients are regression coefficients corresponding to their related terms, and they are estimated in the modeling procedures.

In addition, versions of Models (1) and (2) included the demographic covariates of gender, age, race/ethnicity, and education; alternative versions of calculating total scores; squared terms of total scores; and separate models for demographic subgroups.

Each model was fitted using SUDAAN® software, with appropriate weights and design variables. The terms in the models were tested, ROC statistics were estimated, and SMI predictions based on the model were generated in the dataset of respondents aged 18 or older and who were assigned to either half-sample ($N = 46\,180$).

The weighted number of false positives and false negatives was used to identify the optimal cut-point of each model as follows. Consider the *j*th respondent in the half-sample in question (i.e. WHODAS or SDS). The SMI status based on the SCID/GAF of this respondent will serve as the gold standard, and this will be matched with his or her predicted SMI probability $\hat{\pi}_j$, based on the models described above. Then, for a particular cut-point probability π_0, this respondent will be predicted as SMI positive if $\hat{\pi}_j \geq \pi_0$; otherwise, he or she will be predicted as SMI negative. As a consequence, we now have enough information to know which cell in the ROC 2×2 contingency table represented in Table 6 the predicted SMI status of this respondent belongs to at this cut-point. Repeat this process for each respondent at this cut-point to fill out the cells in Table 6. The optimal cut-point is determined to be the one that results in the approximate equalization of the weighted number of false positives and false negatives. This cut-point was chosen because it is expected to minimize bias in the final SMI estimates generated from the full NSDUH data. For each model and its chosen cut-point, several ROC statistics were derived, including the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the area under the ROC curve (AUC) (definitions provided at the bottom of Table 7).

Table 5 Weighted[1] descriptive statistics of suicidality, substance use, and SCID variables (percentages)

		Selected SCID cases		Analyzable SCID cases				
Variable	2008 12-month NSDUH cases[2]	Sample A (WHODAS)	Sample B (SDS)	Sample A (WHODAS)	Sample B (SDS)	Total	χ^2 (DF)	*P*-value
Suicidality, PY								
Suicide thoughts	3.8	3.3	4.1	3.6	4.9	4.2	0.77 (1)	0.382
Suicide plans	1.0	1.1	1.1	1.5	1.0	1.2	0.74 (1)	0.392
Suicide attempts	0.5	0.3	0.6	0.2	0.4	0.3	0.79 (1)	0.375
Substance use								
Cigarettes, PM	25.4	32.1	22.9	36.1	18.1	27.2	5.21 (1)	0.025
Marijuana, PM	5.7	5.9	4.3	7.0	3.6	5.3	2.39 (1)	0.125
Alcohol Abuse/ Dependence, PY	7.4	7.3	6.5	4.0	8.7	6.3	4.25 (1)	0.042
SCID variables								
SMI50	N/A	5.3	4.8	4.4	4.3	4.3	0.01 (1)	0.936
SMI59	N/A	9.8	9.0	7.9	8.4	8.1	0.07 (1)	0.787
Any MH	N/A	21.6	19.0	18.0	18.5	18.2	0.02 (1)	0.897
Any SUD	N/A	6.0	6.3	5.7	6.6	6.2	0.15 (1)	0.701

DF = degrees of freedom, MH = mental health, N/A = not applicable, PM = past month, PY = past year, SCID = Structural Clinical Interview for DMS-IV, SDS = four-item Sheehan Disability Scale, SMI50 = any SCID MH diagnosis used to determine serious mental illness (SMI) status and Global Assessment of Functioning (GAF) score of 50 or below, SMI59 = any SCID MH diagnosis used to determine SMI status and GAF score of 59 or below, SUD = substance-use disorder, WHODAS = 8-item World Health Organization Disability Assessment Schedule.

[1]The overall NSDUH analysis weight was used for the NSDUH Cases. The overall NSDUH analysis weight multiplied by the inverse of the SCID selection probability was used for the Selected SCID Cases. The MHSS Combined Unadjusted Sample Weight was used for Completed SCID Cases. The MHSS Combined Unadjusted Sample Weight included the following weights: overall NSDUH analysis weight; inverse of the SCID selection probability; non-response adjustment for clinical interview; and post-stratification adjustments by gender, race/ethnicity and age.

[2]This includes all cases for persons aged 18 or older.

Table 6 Receiver operating characteristic 2 × 2 contingency table

		SMI diagnosis based on SCID/GAF	
		Positive	Negative
Predicted SMI probability $\hat{\pi} \geq$ Cut-point Probability π_0	Positive	True positive	False positive
	Negative	False negative	True negative

GAF = Global Assessment of Functioning, SCID = Structural Clinical Interview for DMS-IV, SMI = serious mental illness.

Results of modeling analyses

Models to determine SMI were selected according to three criteria:

1 model parsimony or robustness
2 minimization of misclassification errors in SMI prediction
3 national SMI estimates based on the full 12-month dataset from the main survey that are in line with estimates based on previous research.

To illustrate the process of model selection, a representative set of six WHODAS and six SDS models is presented in Table 7. Model 1 in each half-sample contained only a term for worst K6 total score. The reason for

Table 7 Receiver operating characteristic statistics of a selection of WHODAS and SDS models (weighted numbers in thousands)

Model[1]	Cut-point	P	N	Pred_P	Pred_N	TP	TN	FP	FN	Sens	Spec	AUC	PPV	NPV
WHODAS Model 1	0.28953	4977	108453	5022	108408	1928	105358	3095	3050	0.387	0.971	0.679	0.384	0.972
WHODAS Model 2	0.30246	4977	108453	4978	108452	2396	105871	2582	2581	0.481	0.976	0.729	0.481	0.976
WHODAS Model 3	0.28095	4977	108453	4998	108432	2463	105918	2535	2514	0.495	0.977	0.736	0.493	0.977
WHODAS Model 4	0.33341	4977	108453	4984	108446	2414	105883	2570	2564	0.485	0.976	0.731	0.484	0.976
WHODAS Model 5	0.26972	4977	108453	5116	108314	2516	105853	2600	2461	0.506	0.976	0.741	0.492	0.977
WHODAS Model 6	0.31233	4977	108453	5058	108372	2454	105849	2604	2523	0.493	0.976	0.735	0.485	0.977
SDS Model 1	0.27773	4744	106748	4606	106886	1328	103469	3279	3417	0.280	0.969	0.625	0.288	0.968
SDS Model 2	0.31188	4744	106748	4794	106698	2057	104011	2737	2687	0.434	0.974	0.704	0.429	0.975
SDS Model 3	0.30509	4744	106748	4979	106514	2289	104059	2689	2455	0.483	0.975	0.729	0.460	0.977
SDS Model 4	0.32753	4744	106748	4966	106527	2288	104071	2677	2456	0.482	0.975	0.729	0.461	0.977
SDS Model 5	0.26657	4744	106748	4837	106655	1782	103693	3055	2963	0.376	0.971	0.673	0.368	0.972
SDS Model 6	0.30704	4744	106748	4840	106652	2893	104801	1947	1851	0.610	0.982	0.796	0.598	0.983

AUC = area under receiver operating characteristics curve based on optimal cut-point [(sensitivity + specificity)/2], FN = number of false negatives based on prediction, FP = number of false positives based on prediction, N = number of negative SMI cases, NPV = negative predictive value (TN/Pred_N), P = number of positive SMI cases, PPV = positive predictive value (TP/Pred_P), Pred_N = number of predicted negative cases, Pred_P = number of predicted positive cases, SDS = four-item Sheehan Disability Scale, Sens = sensitivity (TP/P), Spec = specificity (TN/N), TN = number of true negatives based on prediction, TP = number of true positives based on prediction, WHODAS = eight-item World Health Organization Disability Assessment Schedule.

[1] The terms and parameter estimates of the listed models are given in Tables 12 and 13.

Int. J. Methods Psychiatr. Res. 19(Supplement 1): 61–87 (2010). DOI: 10.1002/mpr

including these 'K6 only' models was to compare them with models also containing WHODAS or SDS terms to see how much the models improved by the addition of these impairment scales. For each of these models, worst K6 total score was highly statistically significant, so this term alone clearly has some predictive power.

Model 2 in each of the half-samples evolved from K6 and WHODAS or SDS item scores. Exploratory data analyses indicated that individual item scores within a scale were all highly correlated, which gave rise to problems of collinearity in models that included all of these items. Consequently, the parameter estimates of these items were unstable (e.g. removal of one term could have a dramatic effect on the parameter estimates of the remaining terms) and some were uninterpretable (e.g. the parameter estimates were negative). Therefore, only item scores with a *P*-value less than 0.10 were retained in the models. WHODAS Model 2 consisted of two K6 item scores and two WHODAS item scores, and SDS Model 2 consisted of two K6 item scores and two SDS item scores (one of which had a negative coefficient). Clearly, beyond the problems with collinearity, another limitation of these models containing a subset of item scores is that information may be lost with the items discarded from the models. Therefore, in the following models, preference was given to total score terms that summarized information from all items.

WHODAS Model 3 consisted of worst K6 and WHODAS total scores, and all terms in this model were statistically significant. SDS Model 3 consisted of worst K6 and SDS total scores, but SDS total score was not significant ($P = 0.6012$) and the parameter estimate was negative. This provides some evidence that the SDS total score may not help much in predicting SMI.

WHODAS Model 4 added a squared worst K6 total score term to explore whether this would explain some of the non-linearity in SMI as a function of worst K6 total score. The squared term in WHODAS Model 4 was statistically significant, suggesting that this is a useful term, but it was not statistically significant in SDS Model 4.

WHODAS Model 5 contained the alternative version of worst K6 and WHODAS total scores described earlier. The squared term of the alternative version of worst K6 total score was not statistically significant, suggesting that the alternative version explains the non-linearity in a more parsimonious way than in WHODAS Model 4. The alternative version of WHODAS total score also appeared to be as predictive as the original version. SDS Model 5 contained the alternative version of worst K6 and SDS total scores. The alternative version of worst K6 total score appeared to be as predictive as the original version, and although the alternative version of SDS total score was not statistically significant, at least its coefficient was positive, suggesting that this version of SDS is a more appropriate predictor of SMI.

The five models for each of the impairment scales discussed so far included no covariates. The reason for this exclusion was the desire for parsimonious robust models that could be applied with some confidence to the full adult 12-month data. We were concerned that the relatively small sample size of the MHSS data together with some unusual weights observed therein might give rise to demographic effects peculiar to this dataset, and that models adjusting for these local effects applied to the full adult 12-month data would induce these effects into estimates based on the larger dataset. Exploratory data analyses showed that this was indeed the case, and, for illustrative purposes, the sixth model for each of the impairment scales was identical to Model 5, except that statistically significant demographic covariates were included.

WHODAS Model 6 contained the alternative version of worst K6 and WHODAS total scores and race/ethnicity. SDS Model 6 contained the alternative version of worst K6 and SDS total scores and education. These same demographic covariates were statistically significant in all of the models discussed above. The fact that the significant covariates were different for each of the half-samples suggests that these represent local effects behaving differently in each of the half-samples.

A comparison of WHODAS Models 5 and 6 indicated that the terms in Model 5 were robust to the inclusion or exclusion of race/ethnicity in the model, but a comparison between the corresponding SDS models indicated that the alternative SDS total score was not robust to the inclusion or exclusion of education in the model.

In summary, these results show that both K6 and WHODAS terms are important in the WHODAS models, but in the SDS models the evidence is less clear about the importance or meaning of SDS terms because of the lack of statistical significance or because of negative coefficients. The alternative version of worst K6 total score appears to be a parsimonious way of dealing with non-linearity, and while the WHODAS models do not appear to be affected much by the choice of original versus alternative versions of WHODAS total score, the alternative version of SDS total score does appear to be a more appropriate predictor of SMI because it has a positive coefficient. WHODAS Model 5 seems to be robust to the inclusion or exclusion of race/ethnicity in the model, but this is not the case for the corresponding SDS model with respect to education. It is also curious that while race/

ethnicity is significant in WHODAS Model 6, education is significant in SDS Model 6, suggesting that the two half-samples have different local effects with respect to these different covariates.

Next, we see how these different models performed in terms of minimizing misclassification errors in SMI prediction, where performance was measured primarily on the ROC statistics of sensitivity and AUC. Because of the relatively low SMI prevalence rate, the specificity of the models did not change much (and therefore had a limited ability to discriminate between models), but the sensitivity was quite variable. Therefore, changes in AUC were almost entirely driven by changes in sensitivity, so both of these statistics were useful in discriminating between models.

The ROC statistics of all six WHODAS models described above are given in Table 7. With the exception of Model 1 (i.e. K6-only model), sensitivity and AUC do not vary much. These results illustrate the improvement that WHODAS terms bring to the models with respect to statistics showing the extent of misclassification error. They also indicate that the models are robust to the actual expression of the K6 and WHODAS terms and whether race/ethnicity is included or not. These results suggest that the most parsimonious model among the models, excluding Model 1, might be considered as a candidate model (i.e. Model 5).

The ROC statistics of the six SDS models described above are given in Table 7, but the pattern of sensitivity and AUC statistics is not as clear as it is in the case of the WHODAS models. Similar to the WHODAS case, SDS Model 1 has the lowest sensitivity and AUC statistics, but, unlike the WHODAS case, these statistics vary somewhat among the remaining models. In particular, there is a relative decrease in the ROC statistics for Model 5, but for Model 6 (i.e. education added to the same terms as in Model 5), there is a relative increase in these statistics. While these results also illustrate that the ROC statistics are improved by including SDS terms in the models, the SDS models are not as robust to the actual expression of SDS terms and are clearly not robust to whether education is included or not.

A more detailed examination of the ROC statistics of the latter two WHODAS and SDS models are given in Tables 8 and 9, respectively, where the statistics are provided for the subgroups of the four demographic variables. Interestingly, for the WHODAS models, the one that excludes race/ethnicity as a covariate appears to provide better sensitivity and AUC measures within the race/ethnicity subgroups. Sensitivity and AUC also seem to be as good, if not better, across the other subgroups. Sensitivity and AUC measures across the subgroups of the two SDS models seem to be somewhat more variable, although sensitivity in the 'Total' row is much higher for the model with education included as a covariate, as noted above.

The third model-selection criterion was related to the behavior of SMI estimates generated by the models for the full 12-month data set. But first, unweighted and weighted gold-standard SMI estimates based on clinical interviews in the MHSS are provided in total and at the demographic subgroup level in Table 10. Weighted gender-level estimates are unbalanced with respect to the two half-samples, and particular imbalances appear in the race/ethnicity subgroups for the WHODAS half-sample and in the education subgroups for the SDS half-sample. This indicates some local effects peculiar to the two half-samples with respect to those covariates, and these have been picked up by the various models. Considering that the MHSS is a fairly small sample, and that the weights vary quite substantially, it is not surprising that local peculiarities occur in the MHSS data.

Weighted estimates of SMI prevalence rates in the full NSDUH data based on the six WHODAS and SDS models are shown in Table 11. Because the WHODAS was selected over the SDS for use in 2009 and later surveys, model selection was first applied to the WHODAS, and then an appropriate model was selected for the SDS. The SMI prevalence estimates based on the five WHODAS models that exclude race/ethnicity show reasonable balance across the subgroups of all demographic variables, including those of race/ethnicity. By contrast, estimates based on the model that includes race/ethnicity (i.e. Model 6) appear to mimic the local imbalances among the race/ethnicity subgroups in the WHODAS half-sample of the MHSS data shown in Table 10, thereby indicating that race/ethnicity would not be a useful term in a model. Therefore, among the four remaining candidate models (i.e. excluding Models 1 and 6), WHODAS Model 5 is the most parsimonious model (and hence is likely to be the most robust to external datasets); it appears to deal with the non-linearity in SMI as a function of worst K6 total scores; and its ROC statistics compare favorably with all the other WHODAS models. Consequently, a decision was made to select WHODAS Model 5 as the final WHODAS model.

The five SDS models that exclude education appear to show reasonable balance across all subgroups, whereas Model 6, which includes education, shows imbalances among the education subgroups. This suggests that education would not be a useful term in an SDS model applied to the adult 12-month data, even though Model 6

Table 8 Subgroup receiver operating characteristic statistics of two selected WHODAS models (weighted numbers in thousands)

	WHODAS Model 5 (Alternative Worst K6 Total Score and Alternative WHODAS Total Score)													
Demographic subset	Cut-point	P	N	Pred_P	Pred_N	TP	TN	FP	FN	Sens	Spec	AUC	PPV	NPV
Total	0.26972	4977	108453	5116	108314	2516	105853	2600	2461	0.506	0.976	0.741	0.492	0.977
Gender = male	0.26972	1724	56524	1759	56490	814	55579	945	911	0.472	0.983	0.728	0.463	0.984
Gender = female	0.26972	3253	51928	3358	51824	1703	50273	1655	1551	0.523	0.968	0.746	0.507	0.970
Age = 18–25	0.26972	881	15652	1466	15068	496	14682	970	386	0.562	0.938	0.750	0.338	0.974
Age = 26–49	0.26972	2375	44385	2459	44301	1162	43088	1298	1213	0.489	0.971	0.730	0.472	0.973
Age = 50+	0.26972	1721	48415	1191	48945	859	48082	333	863	0.499	0.993	0.746	0.721	0.982
Race/Ethnicity = white	0.26972	4538	68714	4384	68868	2228	66558	2156	2310	0.491	0.969	0.730	0.508	0.966
Race/Ethnicity = black	0.26972	286	13860	483	13663	230	13606	253	56	0.804	0.982	0.893	0.476	0.996
Race/Ethnicity = other	0.26972	33	11163	153	11043	23	11032	130	10	0.686	0.988	0.837	0.148	0.999
Race/Ethnicity = Hispanic	0.26972	120	14716	96	14740	35	14655	60	85	0.293	0.996	0.644	0.368	0.994
Education = < high school	0.26972	693	8876	737	8833	455	8594	282	239	0.656	0.968	0.812	0.618	0.973
Education = high school	0.26972	2028	32772	1506	33294	812	32079	694	1216	0.401	0.979	0.690	0.539	0.963
Education = some college	0.26972	1251	33258	1772	32737	651	32137	1121	600	0.520	0.966	0.743	0.367	0.982
Education = college graduate	0.26972	1005	33546	1102	33450	598	33043	504	407	0.595	0.985	0.790	0.543	0.988

Table 8 *Continued*

Demographic subset	WHODAS Model 6 (Alternative Worst K6 Total Score, Alternative WHODAS Total Score, and Race/Ethnicity)												
	Cut-point	P	N	Pred_P	Pred_N	TP	TN	FP	FN	Sens	Spec	AUC	PPV	NPV
Total	0.31233	4977	108453	5058	108372	2454	105849	2604	2523	0.493	0.976	0.735	0.485	0.977
Gender = male	0.31233	1724	56524	1607	56641	771	55688	836	953	0.447	0.985	0.716	0.480	0.983
Gender = female	0.31233	3253	51928	3451	51731	1683	50161	1768	1570	0.517	0.966	0.742	0.488	0.970
Age = 18–25	0.31233	881	15652	1223	15311	382	14812	840	499	0.434	0.946	0.690	0.313	0.967
Age = 26–49	0.31233	2375	44385	2397	44364	1213	43202	1184	1162	0.511	0.973	0.742	0.506	0.974
Age = 50+	0.31233	1721	48415	1439	48698	859	47835	580	863	0.499	0.988	0.743	0.597	0.982
Race/Ethnicity = white	0.31233	4538	68714	4744	68509	2317	66287	2427	2221	0.510	0.965	0.738	0.488	0.968
Race/Ethnicity = black	0.31233	286	13860	258	13888	112	13714	146	174	0.391	0.989	0.690	0.434	0.987
Race/Ethnicity = other	0.31233	33	11163	0	11196	0	11163	0	33	0.000	1.000	0.500	0.000	0.997
Race/Ethnicity = Hispanic	0.31233	120	14716	56	14780	26	14685	31	95	0.213	0.998	0.605	0.453	0.994
Education = < high school	0.31233	693	8876	944	8625	419	8350	526	275	0.604	0.941	0.772	0.443	0.968
Education = high school	0.31233	2028	32772	1479	33321	797	32091	682	1231	0.393	0.979	0.686	0.539	0.963
Education = some college	0.31233	1251	33258	1523	32986	569	32303	955	682	0.455	0.971	0.713	0.373	0.979
Education = college graduate	0.31233	1005	33546	1111	33441	670	33105	441	336	0.666	0.987	0.827	0.603	0.990

AUC = area under receiver operating characteristics curve based on optimal cut-point [(sensitivity + specificity)/2], FN = number of false negatives based on prediction, FP = number of false positives based on prediction, N = number of negative SMI cases, NPV = negative predictive value (TN/Pred_N), P = number of positive SMI cases, PPV = positive predictive value (TP/Pred_P), Pred_N = number of predicted negative cases, Pred_P = number of predicted positive cases, Sens = sensitivity (TP/P), Spec = specificity (TN/N), TN = number of true negatives based on prediction, TP = number of true positives based on prediction, WHODAS = eight-item World Health Organization Disability Assessment Schedule.

Int. J. Methods Psychiatr. Res. 19(Supplement 1): 61–87 (2010). DOI: 10.1002/mpr

Table 9 Receiver operating characteristic statistics of two selected SDS models (weighted numbers in thousands)

	SDS Model 5 (Alternative Worst K6 Total Score and Alternative SDS Total Score)													
Demographic subset	Cut-point	P	N	Pred_P	Pred_N	TP	TN	FP	FN	Sens	Spec	AUC	PPV	NPV
Total	0.26657	4744	106748	4837	106655	1782	103693	3055	2963	0.376	0.971	0.673	0.368	0.972
Gender = male	0.26657	2636	47669	1801	48504	895	46763	906	1741	0.340	0.981	0.660	0.497	0.964
Gender = female	0.26657	2109	59079	3036	58152	887	56930	2150	1222	0.421	0.964	0.692	0.292	0.979
Age = 18–25	0.26657	787	15618	1331	15074	596	14883	735	191	0.758	0.953	0.855	0.448	0.987
Age = 26–49	0.26657	1737	51335	2507	50565	879	49707	1628	858	0.506	0.968	0.737	0.351	0.983
Age = 50+	0.26657	2220	39795	999	41017	306	39102	693	1914	0.138	0.983	0.560	0.307	0.953
Race/Ethnicity = white	0.26657	2740	78741	2925	78556	1325	77141	1600	1415	0.484	0.980	0.732	0.453	0.982
Race/Ethnicity = black	0.26657	1373	9847	531	10688	33	9349	498	1339	0.024	0.949	0.487	0.063	0.875
Race/Ethnicity = other	0.26657	539	2753	1211	2081	394	1935	818	145	0.731	0.703	0.717	0.325	0.930
Race/Ethnicity = Hispanic	0.26657	92	15408	170	15330	30	15268	140	63	0.323	0.991	0.657	0.176	0.996
Education = < high school	0.26657	1690	9137	424	10403	197	8909	227	1493	0.116	0.975	0.546	0.464	0.856
Education = high school	0.26657	627	39117	1147	38597	430	38400	717	197	0.686	0.982	0.834	0.375	0.995
Education = some college	0.26657	1454	27081	1803	26731	527	25804	1276	927	0.363	0.953	0.658	0.292	0.965
Education = college graduate	0.26657	973	31414	1463	30924	628	30579	835	345	0.645	0.973	0.809	0.429	0.989

Table 9 *Continued*

	SDS Model 6 (Alternative Worst K6 Total Score, Alternative SDS Total Score, and Education)													
Demographic subset	Cut-point	P	N	Pred_P	Pred_N	TP	TN	FP	FN	Sens	Spec	AUC	PPV	NPV
Total	0.30704	4744	106748	4840	106652	2893	104801	1947	1851	0.610	0.982	0.796	0.598	0.983
Gender = male	0.30704	2636	47669	2533	47772	2200	47335	333	436	0.834	0.993	0.914	0.868	0.991
Gender = female	0.30704	2109	59079	2307	58881	693	57466	1614	1415	0.329	0.973	0.651	0.301	0.976
Age = 18–25	0.30704	787	15618	958	15447	389	15049	568	398	0.494	0.964	0.729	0.406	0.974
Age = 26–49	0.30704	1737	51335	2006	51066	967	50297	1039	770	0.557	0.980	0.768	0.482	0.985
Age = 50+	0.30704	2220	39795	1877	40139	1537	39455	340	683	0.692	0.991	0.842	0.819	0.983
Race/Ethnicity = white	0.30704	2740	78741	2246	79235	1201	77697	1045	1539	0.438	0.987	0.713	0.535	0.981
Race/Ethnicity = black	0.30704	1373	9847	1580	9639	1264	9531	316	109	0.921	0.968	0.944	0.800	0.989
Race/Ethnicity = other	0.30704	539	2753	842	2450	398	2308	445	141	0.738	0.838	0.788	0.472	0.942
Race/Ethnicity = Hispanic	0.30704	92	15408	172	15328	30	15266	142	63	0.323	0.991	0.657	0.174	0.996
Education = < high school	0.30704	1690	9137	2161	8666	1522	8498	638	168	0.901	0.930	0.915	0.705	0.981
Education = high school	0.30704	627	39117	424	39320	337	39029	88	291	0.537	0.998	0.767	0.793	0.993
Education = some college	0.30704	1454	27081	1460	27074	435	26055	1026	1019	0.299	0.962	0.631	0.298	0.962
Education = college graduate	0.30704	973	31414	795	31592	599	31218	196	374	0.616	0.994	0.805	0.754	0.988

AUC = area under receiver operating characteristics curve based on optimal cut-point [(sensitivity + specificity)/2], FN = number of false negatives based on prediction, FP = number of false positives based on prediction, N = number of negative SMI cases, NPV = negative predictive value (TN/Pred_N), P = number of positive SMI cases, PPV = positive predictive value (TP/Pred_P), Pred_N = number of predicted negative cases, Pred_P = number of predicted positive cases, Sens = sensitivity (TP/P), Spec = specificity (TN/N), TN = number of true negatives based on prediction, TP = number of true positives based on prediction, SDS = four-item Sheehan Disability Scale.

Int. J. Methods Psychiatr. Res. 19(Supplement 1): 61–87 (2010). DOI: 10.1002/mpr

Table 10 SMI estimates in the MHSS data, based on clinical interviews (weighted numbers in thousands)

	Unweighted SMI estimates				Weighted SMI estimates			
	WHODAS		SDS		WHODAS		SDS	
Demographic subset	N	%	N	%	N	%	N	%
Total	95	12.5	77	10.4	4977	4.4	4744	4.3
Gender = male	29	10.4	28	10.4	1724	3.0	2636	5.2
Gender = female	66	13.7	49	10.4	3253	5.9	2109	3.4
Age = 18–25	49	10.8	49	11.4	881	5.3	787	4.8
Age = 26–49	38	15.4	23	9.3	2375	5.1	1737	3.3
Age = 50+	8	13.1	5	7.5	1721	3.4	2220	5.3
Race/Ethnicity = white	76	13.9	55	10.4	4538	6.2	2740	3.4
Race/Ethnicity = black	9	9.8	8	11.3	286	2.0	1373	12.2
Race/Ethnicity = other	4	7.5	8	13.6	33	0.3	539	16.4
Race/Ethnicity = Hispanic	6	8.6	6	7.3	120	0.8	92	0.6
Education = < high school	19	18.1	16	14.5	693	7.2	1690	15.6
Education = high school	28	12.7	19	8.6	2028	5.8	627	1.6
Education = some college	29	10.9	30	12.7	1251	3.6	1454	5.1
Education = college grad	19	11.1	12	6.9	1005	2.9	973	3.0

N = frequency, SDS = four-item Sheehan Disability Scale, SMI = serious mental illness, WHODAS = eight-item World Health Organization Disability Assessment Schedule.

appears to provide better ROC statistics. Therefore, taking this into account and considering the variability in the modeling and ROC statistics of the four remaining candidate models, it seems that the most appropriate model would be the one that provides weighted SMI prevalence estimates for the adult 12-month data that are most similar to those based on WHODAS Model 5. Consequently, Model 5 was selected as the final SDS model because this model satisfies that condition.

The parameter estimates and Wald statistics of the selected WHODAS and SDS models are shown in Tables 12 and 13.

Based on these final models, Table 14 displays the SMI cut-points in terms of alternative and original worst K6 total scores for each alternative WHODAS and SDS total score, respectively. Compare the results in Table 14 with the unidimensional SMI cut-point based only on worst K6 total score from Model 1 of both half-samples. For WHODAS Model 1, the K6 cut-point is 17, and for SDS Model 1, the cut-point is 18, irrespective of the degree of impairment captured by the WHODAS and SDS terms. Table 14 suggests that for respondents with WHODAS scores of 5 or more, the K6 cut-point of 17 would be too high, and for those with WHODAS scores of 4 or less, the cut-point would be too low. Table 14 also suggests that for respondents with SDS scores of 3 or more, the K6 cut-point of 18 would be too high, and for those with an SDS score of 0 the cut-point would be too low. Finally, note that these K6 cut-points should not be compared with the historical cut-point of 13 determined by Kessler et al. (2003), because in Kessler's study, the impairment component that determined SMI was specified as GAF ≤ 59 rather than GAF ≤ 50.

Also of interest is that for the selected WHODAS model, of the (weighted) 2600000 false-positive cases, 566000 (21.8%) had a SCID diagnosis and a GAF between 51 and 59 (i.e. these cases might be classified as having 'moderate' mental illness), and 955000 (36.7%) had a SCID diagnosis and a GAF > 59 (i.e. these cases might be classified as having 'mild' mental illness). For the selected SDS model, of the 3055000 false-positive cases, 744000 (24.2%) had a SCID diagnosis and a GAF between 51 and 59, and 1091000 (35.7%) had a SCID diagnosis and a GAF greater than 59.

Conclusions

The final WHODAS and SDS models selected were very parsimonious with only two degrees of freedom (DF) each. That is, the WHODAS model has one DF for the alternative version of the worst K6 total score and one DF for the alternative version of the WHODAS total score,

Table 11 Weighted SMI prevalence estimates in the full 12-month data, based on selected models

	Model 1				Model 2				Model 3			
	WHODAS		SDS		WHODAS		SDS		WHODAS		SDS	
Demographic subset	N	%	N	%	N	%	N	%	N	%	N	%
Total	5471	4.9	4866	4.3	4753	4.3	4470	3.9	4763	4.3	3550	3.1
Gender = male	2011	3.8	1586	2.9	1591	3.0	1464	2.7	1611	3.0	1104	2.0
Gender = female	3460	6.0	3279	5.6	3163	5.5	3006	5.1	3153	5.5	2446	4.2
Age = 18–25	1549	9.5	1242	7.5	1134	7.0	1221	7.3	1244	7.6	957	5.7
Age = 26–49	2810	5.7	2594	5.2	2529	5.1	2214	4.4	2573	5.2	1857	3.7
Age = 50+	1113	2.5	1030	2.2	1091	2.4	1034	2.2	946	2.1	735	1.6
Race/Ethnicity = white	3775	4.9	3265	4.2	3618	4.7	2882	3.7	3639	4.8	2279	2.9
Race/Ethnicity = black	553	4.4	557	4.2	379	3.0	628	4.8	370	2.9	462	3.5
Race/Ethnicity = other	238	3.4	233	3.3	269	3.8	193	2.7	210	3.0	158	2.3
Race/Ethnicity = Hispanic	905	6.1	811	5.3	488	3.3	767	5.0	545	3.6	650	4.2
Education = < high school	1254	7.4	1008	5.6	799	4.7	1089	6.0	794	4.7	844	4.7
Education = high school	1802	5.2	1378	3.9	1642	4.7	1310	3.7	1660	4.8	1035	2.9
Education = some college	1494	5.2	1644	5.7	1361	4.8	1343	4.7	1396	4.9	1084	3.8
Education = college grad	922	3.0	836	2.6	951	3.1	728	2.3	912	3.0	587	1.9

	Model 4				Model 5				Model 6			
	WHODAS		SDS		WHODAS		SDS		WHODAS		SDS	
Demographic subset	N	%	N	%	N	%	N	%	N	%	N	%
Total	4358	3.9	3336	2.9	4839	4.4	4986	4.4	4308	3.9	4603	4.0
Gender = male	1456	2.7	1031	1.9	1609	3.0	1665	3.0	1429	2.7	1570	2.8
Gender = female	2901	5.0	2305	3.9	3230	5.6	3321	5.7	2878	5.0	3033	5.2
Age = 18–25	1123	6.9	927	5.6	1224	7.5	1213	7.3	1028	6.3	1024	6.1
Age = 26–49	2313	4.7	155	3.5	2559	5.2	2674	5.3	2267	4.6	2379	4.7
Age = 50+	921	2.0	654	1.4	1056	2.3	1099	2.3	1013	2.2	1200	2.6
Race/Ethnicity = white	3351	4.4	2133	2.7	3755	4.9	3440	4.4	3938	5.2	3027	3.9
Race/Ethnicity = black	362	2.9	446	3.4	369	2.9	522	4.0	137	1.1	584	4.4
Race/Ethnicity = other	208	2.9	153	2.2	212	3.0	308	4.4	18	0.3	252	3.6
Race/Ethnicity = Hispanic	437	2.9	604	3.9	504	3.4	717	4.7	215	1.4	739	4.8
Education = < high school	626	3.7	836	4.6	779	4.6	952	5.3	669	3.9	1895	10.5
Education = high school	1541	4.4	963	2.7	1686	4.9	1385	3.9	1459	4.2	487	1.4
Education = some college	1320	4.6	1044	3.6	1446	5.1	1699	5.9	1271	4.5	1485	5.2
Education = college grad	871	2.8	492	1.6	929	3.0	949	3.0	909	3.0	736	2.3

N = frequency, SDS = four-item Sheehan Disability Scale, SMI = serious mental illness, WHODAS = eight-item World Health Organization Disability Assessment Schedule.

and the SDS model has one DF for the alternative version of the worst K6 total score and one DF for the alternative version of the SDS total score. While the model fit statistics favored the addition of race/ethnicity in the selected WHODAS model, ROC statistics at the subgroup level and SMI estimates generated from the models indicated that the inclusion of this covariate would create imbalances within several demographic subgroups. Similar conclusions apply to the SDS models, even though the ROC statistics also favored the model that included education. This suggests these parsimonious models would be fairly robust to different datasets.

Table 12 Parameter estimates of selected WHODAS regression model

	WHODAS Model 5: Alternative Worst K6 and WHODAS Total Scores					
Term	Beta	Beta SE	T-statistic	*P*-value	DF	Wald *P*-value
Intercept	–4.7500	0.3522	–13.4854	0.0000		
Alt Worst K6 TS[1]	0.2098	0.0765	2.7439	0.0060	1	0.0072
Alt WHODAS TS[2]	0.3839	0.1208	3.1768	0.0024	1	0.0020

Alt = alternative, DF = degrees of freedom, K6 = six-item psychological distress scale, SE = standard error, TS = total score, WHODAS = eight-item World Health Organization Disability Assessment Schedule.
[1] Alt Worst K6 TS: worst K6 total score < 8 recoded as 0; worst K6 total score 8–24 recoded as 1–17.
[2] Alt WHODAS TS: WHODAS item scores < 2 recoded as 0; WHODAS item scores 2–3 recoded as 1, then summed for a score ranging from 0 to 8.

Table 13 Parameter estimates of selected SDS regression model

	SDS Model 5: Alternative Worst K6 and SDS Total Scores					
Term	Beta	Beta SE	T-statistic	*P*-value	DF	Wald *P*-value
Intercept	–4.4924	0.5206	–8.6299	0.0000		
Alt Worst K6 TS[1]	0.2960	0.0973	3.0435	0.0030	1	0.0030
Alt SDS TS[2]	0.2242	0.3862	0.5805	0.5629	1	0.5629

Alt = alternative, DF = degrees of freedom, K6 = six-item psychological distress scale, SE = standard error, TS = total score, SDS = four-item Sheehan Disability Scale.
[1] Alt Worst K6 TS: worst K6 total score < 8 recoded as 0; worst K6 total score 8–24 recoded as 1–17.
[2] Alt SDS TS: SDS item scores < 7 recoded as 0; SDS item scores 7–10 recoded as 1, then summed for a score ranging from 0 to 4.

The overall conclusion from this 12-month analysis and the 6-month analysis that preceded it is that when added to models with K6, the WHODAS improves the prediction of SMI. Furthermore, WHODAS is a better predictor of SMI than SDS and should be continued as the measure of impairment in future NSDUHs. Nevertheless, using the final models, SMI estimates based on the SDS in the 2008 full dataset are very similar to those based on the WHODAS, indicating that the estimates from the two half-samples could be combined to form single estimates, without adding too much error.

A final note of caution should be added. The number of false-positive and false-negative cases, even for the final models, indicates that a modeling approach to providing SMI prevalence estimates does have limitations when compared with a direct approach, based on clinical interviews. This is to be expected in an exercise that uses brief screening scales to estimate diagnoses based on in-depth semi-structured clinical interviews. Because of this, it is important to focus on aggregate prevalence estimates, where individual false positives and false negatives have a chance to cancel out, rather than on the prediction of individual-level SMI when interpreting the screening results.

Acknowledgments

This work was prepared by the Division of Population Surveys, Office of Applied Studies, Substance Abuse and Mental Health Services Administration, US Department of Health and Human Services, and by RTI International (a trade name of Research Triangle Institute). Work by RTI was performed under Contract No. 283-2004-00022.

Declaration of interest

The authors have no competing interests.

Table 14 K6 cut-points for each WHODAS and SDS total score

Alternative WHODAS total score	Alternative worst K6 SMI cut-point	Worst K6 SMI cut-point
0	17	24
1	17	24
2	15	22
3	13	20
4	11	18
5	9	16
6	7	14
7	6	13
8	4	11

Alternative SDS total score	Alternative worst K6 SMI cut-point	Worst K6 SMI cut-point
0	12	19
1	11	18
2	11	18
3	10	17
4	9	16

K6 = six-item psychological distress scale, SDS = four-item Sheehan Disability Scale, SMI = serious mental illness, WHODAS = eight-item World Health Organization Disability Assessment Schedule.

References

Aldworth J., Barnett-Walker K., Chromy J., Karg R., Morton K., Novak S.P. (2008). *Measuring serious mental illness with the NSDUH: Results of the 2008 6-month analysis*. Prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Mental Health Surveillance Survey Deliverable No. 3.1, RTI/0209009.423.006.008, Research Triangle Institute International.

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

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. *Arch Gen Psychiatry*, **60**, 184–189.

Leon A.C., Olfson M., Portera L., Farber L., Sheehan D.V. (1997). Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *Int J Psychiatr Med*, **27**, 93–105.

Novak S.P., Colpe L.J., Barker P.G., Gfroerer J.C. (2010). Development of a brief mental health impairment scale using a nationally representative sample in the United States. *Int J Methods Psychiatr Res*, **19**(Suppl. 1), 49–60.

Rehm J., Ustun 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. *Int J Methods Psychiatr Res*, **8**, 110–123, DOI: 10.1002/mpr.61

Appendix A: K6 module

[SPLIT RANDOM SAMPLE: SAMPLE A WILL RECEIVE THE WHODAS, SAMPLE B WILL RECEIVE THE SHEEHAN DISABILITY SCALE, ALL ADULTS WILL RECEIVE THE SUICIDALITY QUESTIONS]

(Questions administered only to respondents 18 years or older.)

K6 scale (Administered to Sample A and Sample B in the Random Split sample design) the K6 has been expanded to include both 30-day and past 12-month reference periods per recommendation from the Expert Consultant group)

DIINTRO [IF CURNTAGE = 18 OR OLDER] These questions ask how you have been feeling during the **past 30 days**

NERVE30 [IF CURNTAGE = 18 OR OLDER] During the past 30 days, how often did you feel nervous?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

HOPE30 [IF CURNTAGE = 18 OR OLDER] During the past 30 days, how often did you feel hopeless?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

FIDG30 [IF CURNTAGE = 18 OR OLDER] During the past 30 days, how often did you feel restless or fidgety?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

NOCHR30 [IF CURNTAGE = 18 OR OLDER] During the past 30 days, how often did you feel so sad or depressed that nothing could cheer you up?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

EFFORT30 [IF CURNTAGE = 18 OR OLDER] During the past 30 days, how often did you feel that everything was an effort?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

DOWN30 [IF CURNTAGE = 18 OR OLDER] During the past 30 days, how often did you feel down on yourself, no good or worthless?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

WORST30 The last questions asked about how you have been feeling during the past 30 days. Now think about **the past 12 months**. Was there a month in the past 12 months when you felt more depressed, anxious, or emotionally stressed than you felt during the past 30 days?

1 Yes
2 No

DSNERV1 [IF CURNTAGE = 18 OR OLDER AND WORST30 = 1] Think of one month in the past 12 months when you were the most depressed, anxious, or emotionally stressed.

During that month, how often did you feel nervous?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

DSHOPE [IF CURNTAGE = 18 OR OLDER AND WORST30 = 1] During that same month when you were at your worst emotionally . . .

how often did you feel hopeless?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

DSFIDG [IF CURNTAGE = 18 OR OLDER AND WORST30 = 1] During that same month when you were at your worst emotionally . . .

how often did you feel restless or fidgety?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

DSNOCHR [IF CURNTAGE = 18 OR OLDER AND WORST30 = 1] During that same month when you were at your worst emotionally . . .

how often did you feel so sad or depressed that nothing could cheer you up?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

Int. J. Methods Psychiatr. Res. 19(Supplement 1): 61–87 (2010). DOI: 10.1002/mpr

DSEFFORT [IF CURNTAGE = 18 OR OLDER AND WORST30 = 1] During that same month when you were at your worst emotionally . . .

how often did you feel that everything was an effort?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

DSDOWN [IF CURNTAGE = 18 OR OLDER AND WORST30 = 1] During that same month when you were at your worst emotionally . . .

how often did you feel down on yourself, no good, or worthless?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

DEFINE DISTRESS:
IF NERVE30 = 1–4 OR HOPE30 = 1–4 OR FIDG30 = 1–4, OR NOCHR30 = 1–4 OR EFFORT30 = 1–4 OR DOWN30 = 1–4, OR DSNERV1 = 1–4 OR DSHOPE = 1–4 OR DSFIDG = 1–4 OR DSNOCHR = 1–4 OR DSEFFORT = 1–4 OR DSDOWN = 1–4, THEN DISTRESS = 1 ELSE, DISTRESS = 2

Appendix B: WHODAS module

LIKERT [IF SAMPLE A AND DISTRESS = 1] The next questions are about how much your emotions, nerves, or mental health caused you to have **difficulties in daily activities** over the past 12 months.

Press [ENTER] to continue.

LIREMEM [IF SAMPLE A AND DISTRESS = 1] During that one month when your emotions, nerves or mental health interfered **most** with your daily activities . . .

how much difficulty did you have **remembering to do things you needed to do**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
DK/REF

LICONCEN [IF SAMPLE A AND DISTRESS = 1] During that one month when your emotions, nerves or mental health interfered **most** with your daily activities . . .

how much difficulty did you have **concentrating on doing something important when other things were going on around you**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
DK/REF

LIGOOUT1 [IF SAMPLE A AND DISTRESS = 1] During that one month when your emotions, nerves or mental health interfered **most** with your daily activities . . .

how much difficulty did you have **going out of the house and getting around on your own**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
5 You didn't leave the house on your own
DK/REF

LIGOOUT2 [IF LIGOOUT1 = 5] Did problems with your emotions, nerves, or mental health keep you from leaving the house on your own?

1 Yes
2 No
K/REF

LISTRAN1 [IF SAMPLE A AND DISTRESS = 1] During that one month when your emotions, nerves or mental health interfered **most** with your daily activities . . .

how much difficulty did you have **dealing with people you did not know well**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
5 You didn't deal with people you did not know well
DK/REF

LISTRAN2 [IF LISTRAN1 = 5] Did problems with your emotions, nerves, or mental health keep you from dealing with people you did not know well?

1 Yes
2 No
DK/REF

LISOC1 [IF SAMPLE A AND DISTRESS = 1] During that one month when your emotions, nerves or mental health interfered **most** with your daily activities . . .

how much difficulty did you have **participating in social activities, like visiting friends or going to parties**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
5 You didn't participate in social activities
DK/REF

LISOC2 [IF LISOC1 = 5] Did problems with your emotions, nerves, or mental health keep you from participating in social activities?

1 Yes
2 No
DK/REF

LIHHRES1 [IF SAMPLE A AND DISTRESS = 1] During that one month when your emotions, nerves or mental health interfered **most** with your daily activities . . .

how much difficulty did you have **taking care of household responsibilities**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
5 You didn't take care of household responsibilities
DK/REF

LIHHRES2 [IF LIHHRES1 = 5] Did problems with your emotions, nerves, or mental health keep you from taking care of household responsibilities?

1 Yes
2 No
DK/REF

LIWKRES1 [IF SAMPLE A AND DISTRESS = 1] During that one month when your emotions, nerves or mental health interfered **most** with your daily activities . . .

how much difficulty did you have **taking care of your daily responsibilities at work or school**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
5 You didn't work or go to school
DK/REF

LIWKRES2 [IF LIKWKRES1 = 5] Did problems with your emotions, nerves, or mental health keep you from taking care of your daily responsibilities at work or school?

1 Yes
2 No
DK/REF

LIWKQUIC [IF SAMPLE A AND DISTRESS = 1 AND LIWKRES1 ≠ 5] During that one month when your emotions, nerves or mental health interfered **most** with your daily activities . . .

how much difficulty did you have **getting your daily work done as quickly as needed**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
DK/REF

Appendix C: Sheehan disability scale

MHAD66a [IF SAMPLE B AND DISTRESS = 1] The next questions are about how much your emotions, nerves, or mental health **interfered with your daily activities** over the past 12 months. In answering, think of **one month** in the past 12 months when your emotions, nerves, or mental health interfered **most** with your daily activities.

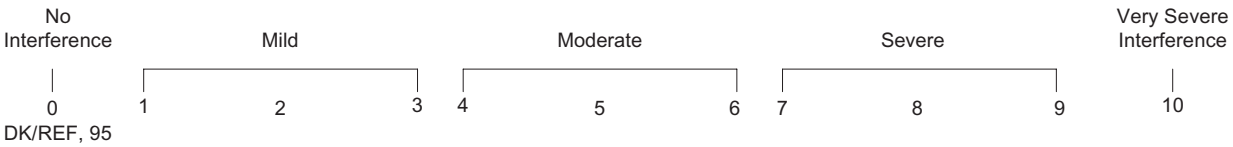
Using the 0 to 10 scale shown below, where 0 means **no** interference and 10 means very **severe** interference, select the number that describes how much **your emotions, nerves or mental health** interfered with each of the following activities during that period. You can use any number between 0 and 10 to answer. If this activity does not apply to you, type in 95.

During that month when you were at your worst emotionally, how much did your emotions interfere with your home management, like cleaning, shopping, and working around the house, apartment, or yard?

Int. J. Methods Psychiatr. Res. 19(Supplement 1): 61–87 (2010). DOI: 10.1002/mpr

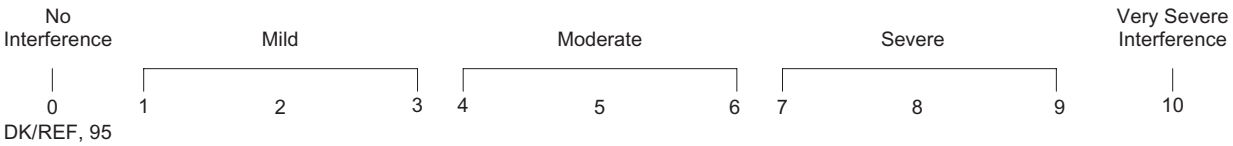
MHAD66b [IF SAMPLE B AND DISTRESS = 1] During that **month** in the past 12 months when you were at your worst emotionally how much did this interfere with your ability to work?

You can use any number between 0 and 10 to answer. If this activity does not apply to you, type in 95.



MHAD66c [IF SAMPLE B AND DISTRESS = 1] During that month when your were at your worst emotionally, how much did this interfere with your ability to form and maintain **close** relationships with other people?

You can use any number between 0 and 10 to answer. If this activity does not apply to you, type in 95.



MHAD66d [IF SAMPLE B AND DISTRESS = 1] How much did your emotions interfere with your social life during that period of time?

You can use any number between 0 and 10 to answer. If this activity does not apply to you, type in 95.

MHAD68 [IF ANY RESPONSES TO AD66a – AD66d = 1–10 OR DK/REF] About how many days out of 365 in the past 12 months were you **totally unable** to work or carry out your normal activities because of your emotions, nerves or mental health?

You can use any number between 0 and 365 to answer.

OF DAYS: _____ [RANGE: 0–365]
DK/REF, 95