ORIGINAL ARTICLE

Screening for Serious Mental Illness in the General Population

Ronald C. Kessler, PhD; Peggy R. Barker, MPH; Lisa J. Colpe, PhD, MPH; Joan F. Epstein, MS; Joseph C. Gfroerer, BA; Eva Hiripi, MS; Mary J. Howes, PhD; Sharon-Lise T. Normand, PhD; Ronald W. Manderscheid, PhD; Ellen E. Walters, MS; Alan M. Zaslavsky, PhD

Background: Public Law 102-321 established a block grant for adults with "serious mental illness" (SMI) and required the Substance Abuse and Mental Health Services Administration (SAMHSA) to develop a method to estimate the prevalence of SMI.

Methods: Three SMI screening scales were developed for possible use in the SAMHSA National Household Survey on Drug Abuse: the Composite International Diagnostic Interview Short-Form (CIDI-SF) scale, the K10/K6 nonspecific distress scales, and the World Health Organization Disability Assessment Schedule (WHO-DAS). An enriched convenience sample of 155 respondents was administered all screening scales followed by the 12-month Structured Clinical Interview for *DSM-IV* and the Global Assessment of Functioning (GAF). We defined SMI as any 12-month *DSM-IV* disorder, other than a substance use disorder, with a GAF score of less than 60.

Results: All screening scales were significantly related to SMI. However, neither the CIDI-SF nor the WHO-DAS improved prediction significantly over the K10 or K6 scales. The area under the receiver operating characteristic curve of SMI was 0.854 for K10 and 0.865 for K6. The most efficient screening scale, K6, had a sensitivity (SE) of 0.36 (0.08) and a specificity of 0.96 (0.02) in predicting SMI.

Conclusions: The brevity and accuracy of the K6 and K10 scales make them attractive screens for SMI. Routine inclusion of either scale in clinical studies would create an important, and heretofore missing, crosswalk between community and clinical epidemiology.

Arch Gen Psychiatry. 2003;60:184-189

From the Department of Health Care Policy, Harvard Medical School (Drs Kessler, Howes, Normand, and Zaslavsky and Mss Hiripi and Walters), and the Department of Biostatistics, Harvard School of Public Health (Dr Normand), Boston, Mass; the Office of Applied Studies (Mss Barker and Epstein and Mr Gfroerer) and the Division of State and Community Systems Development (Dr Manderscheid), Substance Abuse and Mental Health Services Administration, Rockville, Md; and the Division of Mental Disorders, Behavioral Research, and AIDS, National Institute of Mental Health, Bethesda, Md (Dr Colpe).

PUBLIC LAW (PL) 102-321, the Alcohol, Drug Abuse, and Mental Health Administration Reorganization Act, established a block grant for states to fund community mental health services for adults with "serious mental illness" (SMI). The law required states to include incidence and prevalence estimates in their annual applications for block grant funds. The law also required the Substance Abuse and Mental Health Services Administration (SAMHSA) to develop an operational definition of SMI and to create an estimation method based on this definition for use by the states. The definition of SMI stipulated in PL 102-321 requires the person to have at least one 12-month *DSM* disorder, other than a substance use disorder, and to have "serious impairment." Subsequently, SAMHSA decided that "serious impairment" should be defined as a Global Assessment of Functioning (GAF) score of less than 60.[1]

Although preliminary state-level estimates of SMI were based on secondary analysis of existing epidemiological surveys,[2] these estimates were recognized to be merely provisional. More precise estimates were consequently sought by developing screening scales for SMI that could be included in the annual SAMHSA National Household Survey on Drug Abuse (NHSDA). The NHSDA is a nationally representative face-to-face survey that, beginning in 1999, includes about 70000 respondents (45000 adults). The NHSDA is designed to allow direct state estimates for the 8 largest states and enough cases in the remaining states to allow small area estimation methods[3] to be used to produce indirect state-level estimates.

This report presents the results of a methodological study that was conducted to select SMI screening scales for use in the NHSDA. Three sets of screening scales were used in this research. The first set consists of a truncated version of the World Health Organization (WHO) Composite International Diagnostic Interview Short-Form (CIDI-SF) scales,[4] a series of disorder-specific scales that as-

sign predicted probabilities of meeting 12-month criteria for several *DSM-IV* anxiety and mood disorders. The CIDI-SF scales are based on analyses of the National Comorbidity Survey (NCS).[5] The NCS data were used to select the smallest set of CIDI symptom questions that could reproduce the additive association between weighted symptom counts and diagnoses for each disorder. Targeted subsample replications were used to construct the CIDI-SF scales to ensure consistent sensitivities and specificities for men and women, major racial or ethnic groups (non-Hispanic whites, non-Hispanic blacks, and Hispanics), and people who differ in age, education, and urbanicity. The CIDI-SF was modified rationally to include diagnostic stem questions and summary symptom questions. Questions to screen for nonaffective psychosis based on NCS analyses carried out after the development of the CIDI-SF scales[6] were also included in this screening battery.

The second set of screening scales consists of the K10 and K6 scales of nonspecific psychological distress.[7] These scales were developed for use in the core of the redesigned US National Health Interview Survey (NHIS). The 10 questions in the K10 scale and the subset of 6 of these questions in the K6 scale ask respondents how frequently they experienced symptoms of psychological distress (eg, feeling so sad that nothing can cheer you up) during the past 30 days. This reference period was modified to "the one month in the past 12 months when you were at your worst emotionally" for the NHSDA to match the recall period in the other scales. Responses are recorded using a 5-category scale (all of the time, most of the time, some of the time, a little of the time, and none of the time). Like other commonly used scales of nonspecific distress, the questions in the K10/K6 scales all have high loadings on a first principal factor of nonspecific distress in factor analyses carried out in general population samples.[8] This factor is indicated by a heterogeneous set of questions that define behavioral, emotional, cognitive, and psychophysiological manifestations of psychological distress. The K10/K6 scales were developed to measure this dimension by using modern item response theory methods[9] that select questions with optimal sensitivity in the 90th- to 99th-percentile range of the general population distribution of psychological distress and that have consistent item response theory sensitivities across a number of sociodemographic subsamples. The K6 scale has been shown to significantly outperform the widely used 12-question General Health Questionnaire (GHQ-12) in screening for *International Classification of Diseases, Tenth Revision (ICD-10)* disorders, even though the GHQ-12 has twice as many questions as the K6 scale.[10]

The first two sets of scales were selected to screen for *DSM* disorders, but the third scale, the WHO Disability Assessment Schedule (WHO-DAS),[11] was selected to screen for activity limitations associated with these disorders that might help detect people who have GAF scores of less than 60, a requirement for a diagnosis of SMI. The WHO-DAS was developed to operationalize the core dimensions in the *International Classification of Functioning, Disability, and Health*.[12] For this study, the WHO-DAS was streamlined to include 16 core questions that ask about level of difficulty carrying out daily activities in the domains of self-care, mobility, productive activity, social and family life, and community participation during the past 30 days (eg, difficulties dressing oneself, walking outside the home, or understanding what people are saying in social conversations). The reference period for the NHSDA version of the WHO-DAS was changed to "the one month in the past 12 months when your emotions, nerves, or mental health interfered most with your daily activities." Responses are recorded using a 4-category scale (severe, moderate, mild, or no difficulty). Exploratory factor analysis shows that all WHO-DAS items have strong loadings on a first principal factor of global impairment. This factor structure has been shown to be quite consistent across a wide range of sociodemographic groups and regions of the world. The WHO-DAS also has excellent concurrent validity in relation to a number of more extensive structured assessments of role impairment.[11]

All 4 scales were subjected to iterative rounds of cognitive laboratory testing and were modified according to the results of that testing for audio computer-assisted self-administration in the NHSDA. Audio computer-assisted self-administration involves the use of digitally recorded computerized questions that are administered to respondents through headphones attached to a laptop computer. The respondent enters responses using the computer keypad without the interviewer either hearing the questions or viewing the computer screen. Experimental research has shown that the use of audio computer-assisted self-administration significantly increases reports of potentially embarrassing or illegal behaviors, including symptoms of mental disorders,[13-15] making this data collection method of potentially great value for the assessment of serious emotional problems.

Once the audio computer-assisted self-administration version of the scales was developed, a methodological study was conducted to evaluate whether any or all of the 3 scales might be a useful screen for SMI in the NHSDA. The study involved administering all 3 sets of scales to a general population sample who were then interviewed by clinical interviewers blinded to screening scales scores and classified as having or not having SMI based on 12-month prevalences of *DSM-IV* disorders, as assessed by the Structured Clinical Interview (SCID) for *DSM-IV*[16] and scores on the GAF.[1] Logistic regression analyses were then carried out to estimate the strength of associations between the screening scales and SMI using linear and nonlinear prediction equations that assumed either additive or multiplicative associations among the different screening scales. The precision of the best-fitting equations was then evaluated using receiver operating characteristic (ROC) curve analysis.

METHODS

SAMPLE AND MEASURES

The methodological study was conducted in a 2-stage convenience sample. The first stage consisted of a brief telephone recruitment interview with a sample of 1000 respondents aged 18 years or older with listed phone numbers in the Boston, Mass, metropolitan area. First, respondents were informed that Harvard Medical School was collaborating with the federal government to develop measures for use in future government health

Table 1. Pearson Correlations and Mean (SD) Scores for the CIDI-SF, K10 and K6, and WHO-DAS Screening Scales (n = 155)

	CIDI-SF*	K10	K6	WHO-DAS
K10	0.69†			
K6	0.65†	0.97†		
WHO-DAS	0.75†	0.71†	0.67†	
Mean (SD)	0.46 (0.90)	9.20 (6.73)	5.93 (4.26)	7.69 (10.01)

Abbreviations: CIDI-SF, Comprehensive International Diagnostic Interview–Short Form; WHO-DAS, World Health Organization Disability Assessment Schedule.

* The CIDI-SF is presented as a count of screened positive diagnoses, although a number of other specifications were also considered (see text).

†$P<.05$; 2-sided test.

surveys and that we wanted to ask a few questions over the telephone to determine whether the respondent was eligible for a face-to-face in-home interview. Checklist questions about chronic physical illnesses were then asked, followed by 2 questions about perceived physical and mental health, 2 questions about role impairments because of problems with physical and mental health, and the CIDI diagnostic stem questions. A subsample of respondents was recruited for the second-stage in-home interview with a target of 100 completed interviews among those who screened positive for mental health problems and 50 among other first-stage respondents. Verbal informed consent was obtained at the beginning of the first-stage interview after respondents were informed that they might also be invited to participate in a second-stage face-to-face in-home interview. Respondents were also informed that we were offering a $25 honorarium for the in-home interview.

Second-stage assessments were carried out in the homes of 155 respondents (slightly more than the target of 150 because of a somewhat higher completion rate than anticipated among people who agreed to the interview). These assessments were administered by masters- and doctoral-level clinical psychologists. The screening scales were administered first followed by the 12-month nonpatient version of the SCID that includes a GAF rating based on SCID responses. An innovative approach was used to blind clinical interviewers to the screening scale scores before they administered the SCID. Specifically, after obtaining written informed consent, the clinical interviewer turned on a laptop computer and instructed the respondent in how to self-administer the screening scales. The respondent completed these scales without discussing the answers with the interviewer. The laptop computer was programmed in such a way that the interviewer was unable to review the respondent's answers to the self-administered questions, although some rough indication about the number of questions endorsed could be inferred from the length of time the respondent took to complete the self-administered questions. After the self-administered questions were completed, the laptop was shut off, and the interviewer administered the SCID.

The order of administration of the screening scales in the self-administered part of the in-home interview was the CIDI-SF, followed by the K10/K6 scales and the WHO-DAS. The diagnoses included in the CIDI-SF were generalized anxiety disorder, obsessive-compulsive disorder, panic disorder, phobias, posttraumatic stress disorder, dysthymia, major depression, and mania. In addition, a series of screening questions in the CIDI-SF format were developed and included for nonaffective psychosis. (The text of the screening scales is available at http://www.hcp.med.harvard.edu/ncs/relatedmaterials.htm.)

Unlike the typical administration of the WHO-DAS, in which respondents are asked about difficulties in functioning "because of problems with your health," we asked respondents about difficulties because of problems with their "emotions, nerves, or mental health." Because of this focus, respondents who failed to endorse any of the CIDI-SF stem questions or K10/K6 questions were skipped out of the WHO-DAS. The Human Subjects Committee of Harvard Medical School approved the methods and procedures of the study.

ANALYSIS PROCEDURES

We defined SMI as meeting criteria for at least one of the *DSM-IV*/SCID diagnoses, other than a substance use disorder, and having a GAF score of less than 60. Because we oversampled first-stage respondents with emotional problems, it was necessary to weight the 155 cases to have the same distribution as the 1997 NHIS sample on the cross-classification of age, sex, education, and a coarse 4-category version of responses to the K6 scale. Logistic regression analyses were then used to predict SMI from the screening scale scores. Both linear and nonlinear versions of the screening scale scores were used as predictors. We also estimated models that included multiplicative interactions between the CIDI-SF and the WHO-DAS and between the K10/K6 scales and the WHO-DAS to mimic the SMI requirement of a conjunction between disorder and serious impairment in functioning. The precision of the best-fitting prediction equations was then evaluated using ROC curve analysis; ROC curve analysis displays the relationship between the sensitivity and the additive inverse of the specificity of each value of a dimensional screening scale in predicting a dichotomous clinical outcome, in this case SMI. The area under each ROC curve (AUC) was calculated. This area can be interpreted as the probability that a randomly chosen respondent with SMI and a randomly chosen respondent without SMI would be correctly distinguished based on their screening scale scores.[17]

RESULTS

DISTRIBUTIONS AND CORRELATIONS AMONG SCREENING SCALE SCORES

Pearson correlations among the screening scales in the second-stage sample are presented in **Table 1**. The correlation between the K10 and K6 scales is almost perfect (r=0.97), and the other correlations are all very high (r=0.65-0.75). Means are considerably closer to the lower end than the higher end of the distributions of all 4 scales. As shown in **Figure 1**, this pattern indicates that the majority of people in the population report that they have not had recent episodes of the disorders assessed in the CIDI-SF, do not have significant psychological distress, and do not have functional impairment caused by emotional problems. Medians are consistently lower than means, and skew is consistently positive, indicating that a small proportion of respondents have very high scores on the scales. Cronbach α, a measure of internal consistency reliability that is appropriate for the K10 and K6 scales and the WHO-DAS, is high for all 3 of these scales (.93, .89, and .94, respectively).

LOGISTIC REGRESSION ANALYSES

The unweighted prevalence (SE) of SMI in the second-stage sample is 23.2% (3.4%), whereas the weighted preva-

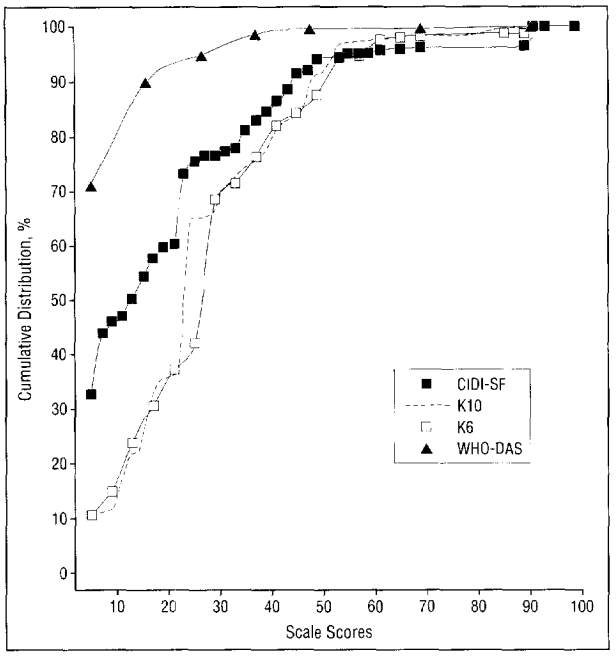

Figure 1. The weighted cumulative distributions of the Composite International Diagnostic Interview Short-Form (CIDI-SF), the K10 and K6 screening scales, and the World Health Organization Disability Assessment Schedule (WHO-DAS) (n=155). All scale ranges have been normalized to 1 to 100. Original ranges are 0 to 9 for the CIDI-SF scale, 0 to 40 for the K10 scale, 0 to 24 for the K6 scale, and 0 to 48 WHO-DAS.

lence is 7.1% (2.1%). A number of logistic regression equations were estimated to predict SMI in weighted and unweighted data. Results were similar, and only weighted estimates are reported here. The first equations were estimated separately for each of the 4 screening scales, beginning with linear associations and then adding higher-order polynomials (up to the fifth power) in an effort to test for nonlinearity. No statistically significant nonlinearity was found for any of these scales in predicting SMI (χ^2_1=0.0-3.4; P=.07-.99). Inspection of residuals and analysis of influential data points also failed to find evidence of meaningful nonlinearity for any scales other than the CIDI-SF. A series of equations was estimated, and each CIDI-SF screening diagnosis was treated as a separate dummy predictor variable. Subsequent equations deleted insignificant predictors (generalized anxiety disorder, nonaffective psychosis, obsessive-compulsive disorder, panic disorder/agoraphobia, and social phobia) and tested for the significance of differences in the slopes of the remaining predictors (major depression, mania, and posttraumatic stress disorder). No globally significant difference across disorders was found (χ^2_2=0.8; P=.66), leading to the creation of a single CIDI-SF summary count with a theoretical range of 0 to 3 for the number of screened-positive disorders.

Comparative model fit of the multivariate logistic regression equations that assessed the joint effects of the 4 scales was evaluated. Results are presented in **Table 2**. Four broad patterns were found. First, each of the scales was found to be a statistically significant predictor of SMI when considered alone (χ^2_1=3.8-12.7; P=.05-.001). Second, no significant interactions were found between any pair of scales (χ^2_1=0.7-1.9; P=.40-.17). Third, the WHO-DAS never improved on the prediction accuracy of an equation that includes either the CIDI-SF or the K10/K6 scales (χ^2_1=0.1-2.1; P=.75-.15), although both the CIDI-SF and the K10/K6 scales improve on the prediction accuracy of an equation that includes only the WHO-DAS (χ^2_1=8.9-9.4; P=.003-.002). Fourth, neither the CIDI-SF nor the K10/K6 scales improve on the prediction accuracy of an equation that includes only the other scale (χ^2_1=2.7-3.0; P=.10-.05). Taken together, these results suggest that only one of the CIDI-SF or K10/K6 scales is needed to obtain maximum prediction accuracy. However, there is one additional result that is inconsistent with this conclusion: the CIDI-SF and the K10/K6 scales both improve on the prediction accuracy of an equation that includes the other scale if the WHO-DAS is included in the equations (χ^2_1=4.2-4.7; P=.04-.03). Because the addition of the WHO-DAS and either the CIDI-SF or the K6/K10 scales to an equation that includes only the CIDI-SF or the K6/K10 scales does not improve on the prediction accuracy of the latter (χ^2_2=4.7-4.8; P=.03-.028), the models containing only a single predictor, either the CIDI-SF or the K6/K10 scales, are the best-fitting models.

Table 2. Model Fit of Logistic Regression Equations to Predict Serious Mental Illness From the CIDI-SF, K10/K6, and WHO-DAS Screening Scales (n = 155)

Parameter	*df*	χ^2 Test* K10	χ^2 Test* K6
CIDI-SF	1	12.6	12.6
K10/K6	1	12.7	12.5
WHO-DAS	1	3.8	3.8
CIDI-SF, K10/K6	2	15.3	15.5
CIDI-SF, WHO-DAS	2	12.7	12.7
K10/K6, WHO-DAS	2	13.2	12.6
CIDI-SF, K10/K6, WHO-DAS	3	17.4	17.3
ME, CIDI-SF × K10/K6†	4	18.7	17.8
ME, CIDI-SF × WHO-DAS†	4	18.1	18.1
ME, K10/K6 × WHO-DAS†	4	18.9	19.2

Abbreviations: CIDI-SF, Composite International Diagnostic Interview–Short Form; WHO-DAS, World Health Organization Disability Assessment Schedule.
*P<.05 for all analyses; 2-sided test.
†Main effects (MEs) of CIDI-SF, K10/K6, and WHO-DAS.

Questions about inpatient and outpatient treatment for emotional problems in the past year were included at the end of the screening scales in the self-administered part of the data collection. The 3 respondents who reported undergoing inpatient treatment were all classified as having SMI by the clinical interviewers. Respondent reports about outpatient treatment, in comparison, were only weakly related to SMI on their own and were not significantly related to SMI after controlling for scores on any one of the screening scales. As a result, information about outpatient treatment was not included in the final prediction equations, but inpatient treatment was taken as an indication of definite SMI.

RECEIVER OPERATING CHARACTERISTIC CURVE ANALYSES

Using the OUTROC option of the LOGISTIC procedure in SAS, version 8,[18] ROC curve analyses were used to

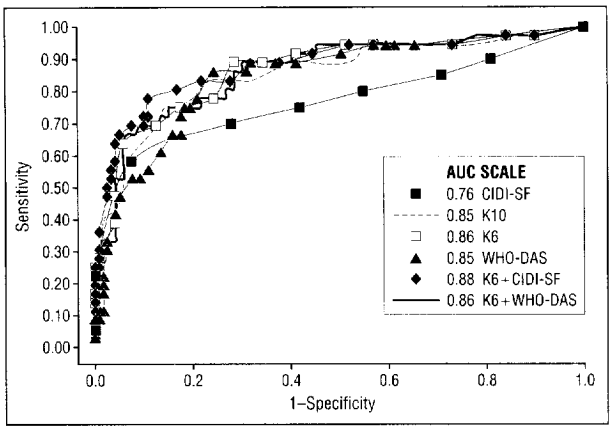

Figure 2. Receiver operating characteristic curves and the area under the curves (AUC) for the Composite International Diagnostic Interview Short-Form (CIDI-SF), the K10 and K6 screening scales, and the World Health Organization Disability Assessment Schedule (WHO-DAS) as well as for illustrative multivariate prediction equations.

evaluate the precision of the CIDI-SF and the K10/K6 scales in discriminating between respondents with SMI and those who did not meet criteria for SMI. As shown in **Figure 2**, both the K10 and the K6 scales have very good discrimination, with AUCs of 0.85 for the K10 scale and 0.86 for the K6 scale. The AUC is considerably lower for the CIDI-SF (0.76). Although the AUC is higher for predicted values of SMI based on an equation that combines the K6 scale with the CIDI-SF (0.88), the results in Table 2 show both that this is a not a statistically significant improvement and that the K6 scale is the most efficient screen for SMI. The optimal cut-point on the K6 to equalize false-positive and false-negative results in the weighted sample was 0 to 12 vs 13 or more (coding item responses 0-4 and summing items to yield a scale with a 0-24 range). At this cut point, sensitivity (SE) was 0.36 (0.08), specificity was 0.96 (0.02), and total classification accuracy was 0.92 (0.02).

COMMENT

This methodological study was designed to evaluate 4 SMI screening scales for use in the annual NHSDA. The main limitation of the study is that it was based on a relatively small local convenience sample rather than a larger nationally representative sample of the sort used in the NHSDA. The main results are that all 4 screening scales are statistically significant predictors of SMI and that the shortest of the 4 scales, the K6 scale, is the best predictor in terms of the AUC. Neither the CIDI-SF nor the WHO-DAS added significantly to the prediction accuracy of the K6 scale. These results add to a growing body of evidence that short, fully structured screening scales, when carefully constructed, can sometimes reproduce classifications based on much more lengthy clinical interviews.[19,20] They also show that symptom severity scales can identify people with serious role impairments, even though the scales do not assess impairments directly.

The item response theory analysis used to develop the K10 and K6 scales shows that the K10 has between 20% and 50% more precision than the K6 in the severity range indicative of SMI.[6] It is consequently surprising that ROC curve analysis failed to find that the K10 scale is more accurate than the K6 scale in screening for SMI. This presumably reflects the fact that the more subtle assessment of distress in the K10 is not related to the categorical distinction used to define SMI. It is also possible, however, that an investigation in a larger sample, a nationally representative sample, or a sample that had a higher concentration of people with clinically significant emotional problems (eg, a primary care sample) would show some advantage for the K10 scale. The same could be true for the CIDI-SF because the small size of the calibration sample might have compromised statistical power to detect SMI based on the CIDI-SF assessments of comparatively rare serious disorders. In this respect, it is noteworthy that the CIDI-SF screening scales for nonaffective psychosis and obsessive-compulsive disorder were both excluded in an early phase of the analysis because the small number of respondents who screened positive for these diagnoses were all found to be false-positives. Analysis in a larger or more broadly representative sample might also find that the WHO-DAS adds to the accuracy of predicting SMI. The finding that the CIDI-SF improves on the prediction accuracy of the K10/K6 scales when the WHO-DAS is also included in the equation adds to the plausibility of this possibility. Based on this finding, all 4 of these scales were included for further analysis in the 2001 and 2002 NHSDA.

In the combined 2001 and 2002 NHSDA samples, the screening scales are being administered to approximately 90 000 adults. (The SMI module of the NHSDA is available at http://www.hcp.med.harvard.edu/ncs/relatedmaterials.htm.) Given the good precision of the scales in screening for SMI, data from these 90 000 adults could generate accurate state-level estimates of SMI. Furthermore, repeating the scales periodically in subsequent years would enable the detection of changes in the prevalence of SMI over time.

In terms of estimation methods, it is possible to generate individual-level predicted probabilities of SMI from the screening scales using the ROC curve results presented here[21] and to generate state-level estimates of SMI from these transformed scores using standard small-area estimation methods.[3] However, because the ROC curve results are based on a relatively small nonprobability sample, they are only preliminary. The development of calibration rules should be based on a larger and more representative sample. An opportunity to do this is presented by the WHO World Mental Health (WMH) surveys,[22] a series of representative community surveys currently underway in 30 countries around the world with a combined sample of more than 200 000 respondents. The K10/K6 scales and the WHO-DAS are included in the WMH surveys with the explicit goal of developing calibration rules for *DSM-IV* and *ICD-10* disorders. Diagnoses of SMI in the WMH sample can be generated from data obtained in the CIDI scale, and replication analyses can be carried out in a probability subsample of WMH respondents who are also being administered the SCID and GAF. Plans exist to develop transformation rules for generating predicted probabilities of SMI from the screening scales when the data from these surveys become available. (These rules will be available at http://www.hcp.med.harvard.edu/ncs/relatedmaterials.htm.)

In addition to their use in screening for SMI, the good results reported here regarding the sensitivity of the K10 and K6 scales suggest that they would be useful broad-gauged screening scales for mental disorders in health-risk appraisal surveys and primary care screening batteries. As was noted in the introduction, the K10 scale has already been shown to be significantly better than the widely used GHQ-12 in screening for *ICD-10* disorders.[9] The fact that the K6 can easily and quickly be either self-administered or interviewer-administered in less than 2 minutes and the K10 in less than 3 minutes is an important attraction in this regard. The K10 or K6 scales might also be useful secondary outcomes in clinical studies as complements to the dimensional assessments of nonspecific impairment, such as the GAF, that are often included in such studies. Inclusion of the K10 or K6 scales in clinical studies would also provide a useful, and heretofore missing, crosswalk between community epidemiological research and clinical research by allowing a comparison of the severity distribution of nonspecific distress among community vs clinical cases.

Submitted for publication April 11, 2002; accepted June 18, 2002.

The research reported here was supported by grants RO1 MH46376, R01 MH52861, RO1 MH49098, and K05 MH00507 from the US Public Health Service; the John D. and Catherine T. MacArthur Foundation Network on Successful Midlife Development (Gilbert Brim, Director); the Pfizer Foundation; and by SAMHSA contract 283-98-9008.

Corresponding author and reprints: Ronald C. Kessler, PhD, Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Suite 215, Boston, MA 02115 (e-mail: kessler@hcp.med.harvard.edu).

REFERENCES

1. Endicott J, Spitzer RL, Fleiss JL, Cohen J. The Global Assessment Scale: a procedure for measuring overall severity of psychiatric disorders. *Arch Gen Psychiatry.* 1976;33:766-771.
2. Kessler RC, Berglund PA, Walters EE, Leaf PJ, Kouzis AC, Bruce ML, Friedman RM, Grosser RC, Kennedy C, Kuehnel TG, Laska EM, Manderscheid RW, Narrow WE, Rosenheck RA, Schneier M. Population-based analyses: a methodology for estimating the 12-month prevalence of Serious Mental Illness. In: Manderscheid RW, Henderson MJ, eds. *Mental Health, United States, 1998.* Rockville, Md: Center for Mental Health Services; 1999:99-109.
3. Schaible WL, ed. *Indirect Estimators in US Federal Programs.* New York, NY: Springer; 1996.
4. Kessler RC, Andrews G, Mroczek D, Üstün TB, Wittchen H-U. The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI-SF). *Int J Methods Psychiatr Res.* 1998;7:171-185.
5. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen H-U, Kendler KS. Lifetime and 12-month prevalence of *DSM-III-R* psychiatric disorders in the United States: results from the National Comorbidity Survey. *Arch Gen Psychiatry.* 1994;51:8-19.
6. Kendler KS, Gallagher TJ, Abelson JM, Kessler RC. Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a US community sample: the National Comorbidity Survey. *Arch Gen Psychiatry.* 1996;53:1022-1031.
7. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand S-LT, Walters EE, Zaslavsky A. Short screening scales to monitor population prevalences and trends in nonspecific psychological distress. *Psychol Med.* 2002;32:959-976.
8. Link BG, Dohrenwend BP. Formulation of hypotheses about the true relevance of demoralization in the United States. In: Dohrenwend BP, Dohrenwend BS, Gould MS, Link B, Neugebauer R, Wunsch-Hitzig R, eds. *Mental Illness in the United States: Epidemiological Estimates.* New York: Praeger Publishers; 1980: 114-132.
9. van der Linden WJ, Hambleton RK. *Handbook of Modern Item Response Theory.* New York: Springer; 1997.
10. Furukawa TA, Andrews G, Slade T, Kessler RC. The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychol Med.* In press.
11. Rehm J, Üstün TB, Saxena S, Nelson CB, Chatterji S, Ivis F, Adlaf E. On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. *Int J Methods Psychiatr Res.* 1999; 8:110-123.
12. World Health Organization. *International Classification of Functioning, Disability, and Health.* 2nd ed. Geneva, Switzerland: World Health Organization; 2001.
13. Epstein J, Barker P, Kroutil L. Mode effects in mental health self reports. *Public Opin Q.* 2001;65:529-549.
14. Substance Abuse and Mental Health Services Administration, Office of Applied Studies. *Development of Computer-Assisted Interviewing Procedures for the National Household Survey on Drug Abuse* [report online]. March 2001. Available at: http://www.samhsa.gov/oas/nhsda/compassistinterview/toc.htm. Accessed April 2002. OAS Methodology Series.
15. Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JH, Sonenstein FL. Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science.* 1998;280:867-873.
16. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for* DSM-IV *Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP).* New York: Biometrics Research, New York State Psychiatric Institute; 1997.
17. Hanley J, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143:29-36.
18. SAS Institute Inc. *SAS User's Guide, Version 8.* Carey, NC: SAS Institute Inc; 1999.
19. Tondo L, Burrai C, Scamonatti L, Weissenburger J, Rush J. Comparison between clinician-rated and self-reported depressive symptoms in Italian psychiatric patients. *Neuropsychobiology.* 1988;19:1-5.
20. Kobak KA, Greist JH, Jefferson JW, Mundt JC, Katzelnick DJ. Computerized assessment of depression and anxiety over the telephone using interactive voice response. *MD Comput.* 1999;16:64-68.
21. Peirce JC, Cornell RG. Integrating stratum-specific likelihood ratios with the analysis of ROC curves. *Med Decis Making.* 1993;13:141-151.
22. Kessler RC, Üstün TB. The World Health Organization World Mental Health 2000 Initiative. *Hospital Management International.* 2000:195-196.