

Mental Health Screening in Primary Care: A Comparison of 3 Brief Measures of Psychological Distress

Annamarie Cano, Ph.D.; Robert P. Sprafkin, Ph.D.; Douglas J. Scaturro, Ph.D.;
Larry J. Lantinga, Ph.D.; Barbara H. Fiese, Ph.D.; and Frank Brand, M.D.

Received April 5, 2001; accepted Sept. 20, 2001. From the Department of Psychology, Wayne State University, Detroit, Mich. (Dr. Cano); the Department of Veterans Affairs Medical Center at Syracuse (Drs. Sprafkin, Scaturro, Lantinga, and Brand); the Department of Psychiatry and Behavioral Sciences (Drs. Sprafkin, Scaturro, Lantinga, and Fiese) and the Department of Psychology (Drs. Sprafkin, Scaturro, Lantinga, and Fiese), Syracuse University; and the Department of Family Medicine, State University of New York Upstate Medical University (Dr. Scaturro), Syracuse, N.Y.

This research was supported in part by grant MH61569 from the National Institute of Mental Health, Bethesda, Md. (Dr. Cano).

Reprint requests to: Annmarie Cano, Ph.D., Department of Psychology, Wayne State University, 71 W. Warren Avenue, Detroit, MI 48202 (e-mail: acano@wayne.edu).

Background: The current study compared 3 brief mental health screening measures in a sample of older patients in a primary care outpatient setting. Previous mental health screening research has been conducted primarily with younger patients, often with only 1 screening measure, thereby limiting the generalizability of findings. In addition, measures have not yet been compared in terms of their ability to discriminate between cases and noncases of psychiatric disorder.

Method: One hundred thirty-four male patients attending their appointments at a primary care clinic in a Department of Veterans Affairs Medical Center participated in this study. Participants completed the General Health Questionnaire-12 (GHQ-12), the Symptom Checklist-10 (SCL-10), and the Primary Care Evaluation of Mental Disorders screening questionnaire and interview.

Results: Receiver operating characteristic analysis yielded the optimum cutoff scores on each brief mental health screening measure and showed that all 3 measures discriminated well between cases and noncases of psychiatric disorders. The 3 measures performed slightly better in terms of discriminating between cases and noncases of mood or anxiety disorders than between cases and noncases of any psychiatric disorder. There were no significant differences between the measures' abilities to accurately identify cases and noncases of disorder.

Conclusion: Primary care physicians are encouraged to use brief mental health screening measures with their patients, since many report symptoms of psychological distress and disorder. It is recommended that the SCL-10 and GHQ-12 be used to detect mood or anxiety disorders in patients such as these because of the accuracy and brevity of these measures.

(*Primary Care Companion J Clin Psychiatry* 2001;3:206-210)

The ability to conduct quick and accurate mental health assessments is becoming an important health care issue. Research has shown that primary care patients often report heightened psychological symptoms as well as psychiatric diagnoses.¹ Frequent appointments made by patients experiencing psychological symptoms may hinder timely medical care and much-needed mental health care. Therefore, it is important to identify patients with clinically significant distress and to refer these patients for a more intensive evaluation. Although some researchers have examined the clinical utility of brief mental health screenings, several issues demand further exploration. These issues include the identification of optimum cutoff scores and the examination of the sensitivity and specificity of commonly used brief screening measures in older patient samples. Furthermore, mental health screening measures need to be compared with each other to help physicians choose the most accurate brief screening instrument. The purpose of this study is to tackle these issues in a sample of older patients attending outpatient primary care appointments. Because much of the research on the efficacy of mental health screening has focused on relatively young patients,^{2,3} the results of this study will be of great use to physicians who wish to accurately identify older patients who might benefit from a more extensive psychiatric evaluation.

The sensitivities and specificities of 3 brief screening measures were examined in the current study. The measures include the General Health Questionnaire-12 (GHQ-12),³ the Symptom Checklist-10 (SCL-10),^{4(p309)} and the Primary Care Evaluation of Mental Disorders (PRIME-MD) questionnaire.⁵ Sensitivity and specificity are important factors to consider when establishing appropriate

cutoff scores for mental health screening measures. Sensitivity refers to the proportion of people who have a psychiatric disorder and who score above a cutoff on a measure of psychological symptoms. Specificity refers to the proportion of people without a psychiatric disorder who score below a cutoff on the same instrument. Choosing measures with high sensitivity as well as high specificity can help physicians identify probable psychiatric disorders while limiting the overdiagnosis of patients who are not likely to have a disorder (i.e., false-positives). Thus, physicians can refer patients while maintaining the cost-effectiveness of their clinics.

Of the 3 measures examined in this study, the GHQ-12 has received the most empirical attention. In a review of 17 published research studies on the GHQ-12, Goldberg and colleagues⁶ found that the most common cutoff score was 2/3 (a score of 2 or less indicating the absence of a mental disorder and a score of 3 or greater indicating the presence of disorder). In their own World Health Organization (WHO) investigation that included 15 centers worldwide, Goldberg and colleagues⁶ found that a cutoff score of 1/2 yielded the best sensitivity (83.5%) and specificity (75.1%) rates for identifying persons with a DSM-IV or ICD-10 diagnosis. Although they did not find age group differences on sensitivity and specificity, other researchers have found that a higher cutoff score (3/4) was more accurate in identifying older adults with psychiatric disorders in a sample of relatives of people with dementia or depression.⁷ It is possible that increased physical symptoms associated with age necessitate higher cutoff scores for older adults. Much less is known about the appropriate cutoff scores for the SCL-10 and the PRIME-MD questionnaire in older patients. To our knowledge, no published research has reported on the sensitivity and specificity of these 2 surveys at various cutoff scores.

In the current study, receiver operating characteristic (ROC) analysis^{8,9} was used to determine optimum cutoff scores associated with the highest possible sensitivity and specificity. ROC analysis is a procedure that has been used successfully in several studies that explored appropriate cutoff scores for the GHQ-12.^{2,6-8} It involves plotting the sensitivity against the false-positive fraction (i.e., 1 - specificity) for every possible cutoff point on a measure. Typically, optimum cutoff scores or thresholds are those that yield sensitivity/specificity pairs that are the highest possible while remaining balanced, with perfect specificity and sensitivity equaling 100% each. ROC analysis also yields a summary measure known as "area under the curve" (AUC) that indicates how well the measure discriminates between cases and noncases of psychiatric disorder overall, with an AUC of 1.00 indicating perfectly accurate discrimination. ROC analysis was also used in this study to determine whether any of the 3 brief mental health screening measures is significantly better

at classifying cases correctly based on diagnostic status. This is the first study to compare the GHQ-12, SCL-10, and PRIME-MD questionnaire in this manner.

In sum, the main goal of the current study was to determine the cutoff scores associated with optimum sensitivity and specificity for the GHQ-12, the SCL-10, and the PRIME-MD questionnaire in an older primary care outpatient sample. In addition, the discriminating abilities of these 3 mental health screening measures were assessed and compared with each other. The findings from this study are expected to generalize to older patients, who are a rapidly growing segment of the population.

METHOD

Participants

One hundred thirty-four male patients (mean \pm SD age = 64.07 \pm 12.74 years) attending primary care appointments at a large upstate New York Department of Veterans Affairs Medical Center participated in this study. Most of the sample was married at the time of the study (N = 80, 59.7%), 18.7% (N = 25) were separated or divorced, 11.9% (N = 16) were widowed, and 9.0% (N = 12) had never been married. One participant did not provide marital status information. Thirty-one participants (23.1%) had combat experience, and 2.2% (N = 3) had been prisoners of war.

Procedure

Mental health technicians approached patients awaiting their primary care appointments about their willingness to participate in a study designed to develop better health care assessments. Patients who agreed to participate completed a consent form and a battery of questionnaires. Some patients were then interviewed immediately after completing their questionnaires on the basis of their responses to the PRIME-MD questionnaire. The same-day administration of screenings and interview addresses the frequent problem of time lags of days or weeks between the administration of screenings and diagnostic interviews as noted by other researchers.² Since no data were collected from patients who declined to participate, no comparisons were possible between patient participants and patients who chose not to participate.

Measures

General Health Questionnaire-12. The GHQ-12 was developed to assess the presence of psychological symptoms.^{3*} Participants noted the presence or absence of 12

*GHQ: A *User's Guide* may be requested at a price of £49.50 (\$71.94) and the GHQ-12 (pack of 100) may be requested at a price of £30.50 (\$44.33) from NFER-Nelson Publishing Company Limited, Darville House, Oxford Road East, Windsor, Berkshire SL4 1DF, United Kingdom.

Table 1. Psychiatric Diagnoses of Patients Who Completed the PRIME-MD Interview (N = 133)^a

Diagnosis	N (%)
Specific diagnosis	
Major depressive disorder	15 (11.3)
Generalized anxiety disorder	8 (6.0)
Dysthymia	6 (4.5)
Anxiety not otherwise specified	8 (5.9)
Somatoform disorder not otherwise specified	6 (4.5)
Minor depressive disorder	5 (3.7)
Probable alcohol abuse	5 (3.8)
Panic disorder	3 (2.3)
Diagnostic cluster	
Mood disorders	22 (16.5)
Anxiety disorders	20 (15.0)
Somatoform disorders	10 (7.5)
Eating disorders	2 (1.5)

^aAbbreviation: PRIME-MD = Primary Care Evaluation of Mental Disorders.

different symptoms within the past few weeks. Sample items are "Have you recently been feeling unhappy and depressed?" "Have you recently lost much sleep over worry?" and "Have you recently felt you couldn't overcome your difficulties?" This scale has been used successfully in large cross-cultural studies and is correlated with psychiatric disorders in primary health care settings.^{3,10} Participants in the current study reported a mean \pm SD of 1.98 ± 3.00 symptoms. The current interitem reliability was 0.90.

Symptom Checklist-10. The SCL-10 is a 10-item instrument assessing global psychological distress derived from the SCL-90.¹¹ The SCL-10 is composed of 6 depression items (e.g., "Have you been distressed by feeling lonely?"), 2 somatization items (e.g., "Have you been distressed by feeling weak in your body?"), and 2 phobic anxiety items (e.g., "Have you been distressed by feeling tense or keyed-up?") from the SCL-90.⁴ These items were chosen on the basis of factor analyses. Participants indicated how well each item described the psychological distress they experienced within the week prior to participation using a 0 (not at all) to 5 (extremely) scale. A single global score can be used as an index of psychopathology or psychological distress.⁴ The SCL-10 had an excellent interitem reliability of 0.92 in the current study, and the mean score was 5.37 ± 7.81 .

Primary Care Evaluation of Mental Disorders. The PRIME-MD is a 2-stage psychiatric diagnostic instrument designed for use in primary care settings.⁵ First, respondents indicate the presence or absence of 10 psychological and 15 somatic symptoms within the past month on the PRIME-MD questionnaire. The PRIME-MD questionnaire is the only self-report measure in the current study that assesses a full range of somatic symptoms (e.g., headache, pain, bowel problems). Second, a trained interviewer conducts a diagnostic interview if respondents endorsed specific clusters of symptoms on the question-

Table 2. Correlations Between the SCL-10, GHQ-12, and PRIME-MD Questionnaire (N = 131)^a

Measure	SCL-10	GHQ-12
GHQ-12	0.87*	...
PRIME-MD	0.63*	0.59*

^aAbbreviations: GHQ-12 = General Health Questionnaire-12, PRIME-MD = Primary Care Evaluation of Mental Disorders, SCL-10 = Symptom Checklist-10.
* $p < .001$.

naire. The PRIME-MD has been validated in a sample of 1000 patients.⁵

In the current study, mental health technicians conducted the interviews. The technicians viewed a PRIME-MD training tape and were trained to 100% diagnostic reliability. PRIME-MD data were available for all but 1 participant. The mean number of PRIME-MD symptoms reported by participants was 5.15 ± 4.28 . The interitem reliability for self-report PRIME-MD symptoms was 0.85. Approximately 31% (N = 41) of the sample was given at least 1 PRIME-MD diagnosis. Of these 41 patients, approximately 41% (N = 17) were diagnosed with 2 or more disorders. Table 1 displays the prevalence of psychiatric diagnoses in this sample.

RESULTS

Sensitivity and Specificity of Brief Screening Measures

Three of the study participants did not complete every measure used in the correlation and ROC analyses; therefore, N = 131 for the following analyses. Pearson product moment correlations indicated that the GHQ-12, SCL-10, and PRIME-MD questionnaire scores were significantly associated with each other (Table 2).

The ROCKIT computer program⁹ was used to conduct the ROC analysis that calculates the sensitivity and specificity associated with every possible cutoff score on each of the 3 brief screening measures. PRIME-MD diagnosis obtained during the interviews was used as the diagnostic gold standard. We limited our analyses to compare (1) patients with any psychiatric diagnosis (e.g., mood, anxiety, eating, somatoform disorder, probable alcohol abuse) with those with no diagnosis and (2) those with either a mood or anxiety disorder with those with neither of those disorders. This decision was made because many physicians may be interested in identifying a patient with a probable diagnosis in a time-efficient manner rather than knowing which specific diagnosis to give a patient. In addition, we chose a combined mood and anxiety category because each group of disorders is very common and often the disorders co-occur.¹²

Table 3 shows the results of the ROC analysis on the 3 brief screening measures. Optimum cutoff scores were those that showed the best balance between sensitivity

Table 3. Best Thresholds, Sensitivity, and Specificity for the Brief Screening Measures (N = 131)^a

Measure	Best Threshold	Sensitivity (%)	Specificity (%)	AUC (95% CI)
For any psychiatric disorder				
GHQ-12	1/2	73.2	83.7	0.85 (0.74 to 0.92)
SCL-10	3/4	76.6	76.2	0.83 (0.71 to 0.91)
PRIME-MD	5/6	79.3	79.1	0.88 (0.80 to 0.93)
For a mood or anxiety disorder				
GHQ-12	1/2	87.4	82.5	0.93 (0.85 to 0.97)
SCL-10	4/5	85.6	85.1	0.93 (0.85 to 0.97)
PRIME-MD	6/7	83.0	82.3	0.91 (0.83 to 0.95)

^aAbbreviations: AUC = area under the curve, CI = confidence interval, GHQ-12 = General Health Questionnaire-12, PRIME-MD = Primary Care Evaluation of Mental Disorders, SCL-10 = Symptom Checklist-10.

and specificity. As in the work of Goldberg and colleagues,⁶ when 2 cutoff scores showed similar balance between sensitivity and specificity, we chose the cutoff score with the higher sensitivity, since the object of this investigation was to accurately identify patients with a psychiatric disorder who scored above the cutoff. As shown in Table 3, the best cutoff scores, or thresholds, for the GHQ-12, SCL-10, and PRIME-MD questionnaire were 1/2, 3/4, and 5/6, respectively, for any psychiatric diagnosis (e.g., mood, anxiety, eating, or somatoform disorder, probable alcohol abuse). All 3 measures demonstrated AUCs of 0.83 or higher, and the AUC values were not significantly different from each other. Thus, the 3 measures were equally effective discriminators between cases and noncases of psychiatric disorder. These AUC values are similar to those found in a large-scale international study investigating the GHQ-12.⁶

In addition to examining cases and noncases of any psychiatric disorder, we also examined cutoff scores for identifying cases versus noncases of mood or anxiety disorders. For a mood or anxiety disorder diagnosis, the best thresholds for the GHQ-12, SCL-10, and PRIME-MD questionnaire were 1/2, 4/5, and 6/7, respectively. The AUCs of 0.91 and higher show that these measures are slightly better at correctly discriminating between cases and noncases of mood and anxiety disorders than at discriminating between cases and noncases of any psychiatric disorder. The higher sensitivity/specificity pairs also show that all 3 measures are better at accurately identifying true-positive and true-negative cases of mood and anxiety disorders. Again, there were no significant AUC value differences among the 3 measures.

DISCUSSION

The current study used ROC analysis to determine optimum cutoff scores on the GHQ-12, SCL-10, and the PRIME-MD questionnaire that could be used to identify

cases of probable psychiatric disorder in a sample of older primary care outpatients. Given the lack of published ROC analysis research with the SCL-10 and PRIME-MD questionnaire, this study aided in the identification of appropriate cutoff scores on these measures for older outpatients. The GHQ-12 has been examined several times with ROC analysis in young and older patients. In the current study, the GHQ-12 cutoff score yielding the best sensitivity and specificity was similar to that found in a WHO study.⁶ This was surprising given other evidence suggesting that the GHQ-12 cutoff score should be raised in samples of older⁷ as well as younger² adults. The fact that the current findings are similar to those of larger scale studies conducted internationally provides further support that our findings are robust.

By comparing the diagnostic discriminative abilities of the GHQ-12, SCL-10, and PRIME-MD questionnaires in the same sample of outpatients, the current study also contributes to the existing knowledge about the relative usefulness of the 3 screening measures. Although there were no significant differences between the measures in terms of the ability to discriminate between cases and noncases of disorder, all 3 brief screening measures performed slightly better in identifying mood or anxiety disorders specifically than in identifying any psychiatric disorder more generally. The improved identification of mood and anxiety disorders most likely occurred because the items on all 3 measures focus on such symptoms. Symptoms associated with eating disorders, alcohol abuse and dependence, and the physical symptoms associated with somatoform disorders were represented only on the PRIME-MD questionnaire. Regardless, the PRIME-MD was not a significantly better discriminator of disorders than the other 2 measures, suggesting that broader item content does not necessarily lead to an increased ability of a measure to identify a broad range of disorders. It is interesting that the mood or anxiety disorder threshold was higher on the SCL-10 and the PRIME-MD questionnaire as compared with the threshold on those same measures for any disorder. These findings suggest that individuals experiencing disorders other than mood or anxiety disorders report fewer symptoms than individuals experiencing mood or anxiety disorders as assessed by these measures.

Given that the GHQ-12, SCL-10, and PRIME-MD questionnaires were equally effective at identifying older primary care outpatients with psychiatric disorders, which of these brief screening measures should be used? Since there were no significant differences between measures in their abilities to detect disorders, the GHQ-12 might be the best choice because it is one of the briefest measures and the same threshold (1/2) can be used to detect cases of any disorder (i.e., mood, anxiety, eating, somatoform, or probable alcohol abuse) as well as cases of mood or anxiety disorders specifically. However, physicians should take

note that the measures in this study are not without their flaws. For instance, although the optimum cutoff score on the GHQ-12 was similar to the optimum cutoff score found in the research, it yielded a lower sensitivity rate than found in previous studies.⁶ This means that a substantial number of older patients who can be diagnosed with a psychiatric disorder on the PRIME-MD interview may not score above the optimum threshold on the GHQ-12 (i.e., 1/2). Similar false-negative rates are present on the SCL-10 and PRIME-MD questionnaire. These patients would not be likely to receive referrals for psychiatric evaluations based on their screening result. Decreasing the cutoff to increase sensitivity is not a cost-effective solution, since there would also be an increase in false-positives. One solution is to use the GHQ-12 to detect mood and anxiety disorders only, because it appears to be slightly more accurate in detecting mood and anxiety disorders than a broader range of psychiatric disorders.

Until a more sensitive instrument is designed to identify the presence of all psychiatric disorders, this conservative approach may be the best option. In addition, future research is needed to determine whether similar cutoff scores can be used to accurately identify women with psychiatric disorders, because the current sample consisted of men only. Based on previous work,⁶ it is expected that there would be no gender differences in the ability of these measures to identify psychiatric disorders.

REFERENCES

1. McKelvey RS, Davies LC, Pfaff JJ, et al. Psychological distress and suicidal ideation among 15–24 year olds presenting to general practice: a pilot study. *Aust N Z J Psychiatry* 1998;32:344–348
2. Hardy GE, Shapiro DA, Haynes CE, et al. Validation of the General Health Questionnaire-12 using a sample of employees from England's health care services. *Psychol Assessment* 1999;11:159–165
3. Goldberg D, Williams P. *A User's Guide to the General Health Questionnaire* 1991. Windsor, United Kingdom: NFER-Nelson Publishing Company Ltd; 1991
4. Nguyen TD, Attkisson CC, Stegner BL. Assessment of patient satisfaction: development and refinement of a service evaluation questionnaire. *Eval Prog Plan* 1983;6:299–313
5. Spitzer RL, Williams JBW, Kroenke K, et al. The utility of a new procedure for diagnosing mental disorders in primary care. *JAMA* 1994;272:1749–1756
6. Goldberg DP, Gater R, Sartorius N, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med* 1997;27:191–197
7. Papassotiropoulos A, Heun R, Maier W. Age and cognitive impairment influence the performance of the General Health Questionnaire. *Compr Psychiatry* 1997;38:335–340
8. Mari JJ, Williams P. A comparison of the validity of two psychiatric screening questionnaires (GHQ-12 and SRQ-20) in Brazil, using relative operating characteristic (ROC) analysis. *Psychol Med* 1985;15:651–659
9. Metz CE. *ROCKIT Program, Version 0.9B*. Chicago, Ill: Department of Radiology, University of Chicago; 1998
10. Sartorius N, Ustun TB, Costa e Silva J, et al. An international study of psychological problems in primary care. *Arch Gen Psychiatry* 1993;50:819–824
11. Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale: preliminary report. *Psychopharmacol Bull* 1973;9:13–28
12. Clark LA, Watson D. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *J Abnorm Psychol* 1991;100:316–336