

A Brief Measure for Assessing Generalized Anxiety Disorder

The GAD-7

Robert L. Spitzer, MD; Kurt Kroenke, MD; Janet B. W. Williams, DSW; Bernd Löwe, MD, PhD

Background: Generalized anxiety disorder (GAD) is one of the most common mental disorders; however, there is no brief clinical measure for assessing GAD. The objective of this study was to develop a brief self-report scale to identify probable cases of GAD and evaluate its reliability and validity.

Methods: A criterion-standard study was performed in 15 primary care clinics in the United States from November 2004 through June 2005. Of a total of 2740 adult patients completing a study questionnaire, 965 patients had a telephone interview with a mental health professional within 1 week. For criterion and construct validity, GAD self-report scale diagnoses were compared with independent diagnoses made by mental health professionals; functional status measures; disability days; and health care use.

Results: A 7-item anxiety scale (GAD-7) had good re-

liability, as well as criterion, construct, factorial, and procedural validity. A cut point was identified that optimized sensitivity (89%) and specificity (82%). Increasing scores on the scale were strongly associated with multiple domains of functional impairment (all 6 Medical Outcomes Study Short-Form General Health Survey scales and disability days). Although GAD and depression symptoms frequently co-occurred, factor analysis confirmed them as distinct dimensions. Moreover, GAD and depression symptoms had differing but independent effects on functional impairment and disability. There was good agreement between self-report and interviewer-administered versions of the scale.

Conclusion: The GAD-7 is a valid and efficient tool for screening for GAD and assessing its severity in clinical practice and research.

Arch Intern Med. 2006;166:1092-1097

Author Affiliations: Biometrics Research Department, New York State Psychiatric Institute and Department of Psychiatry, Columbia University, New York (Drs Spitzer and Williams); Regenrief Institute for Health Care and Department of Medicine, Indiana University, Indianapolis (Dr Kroenke); and Department of Psychosomatic and General Internal Medicine, University of Heidelberg, Heidelberg, Germany (Dr Löwe).

ONE OF THE MOST COMMON anxiety disorders seen in general medical practice and in the general population is generalized anxiety disorder (GAD). The disorder has an estimated current prevalence in general medical practice of 2.8% to 8.5%¹⁻³ and in the general population of 1.6% to 5.0%.⁴⁻⁶ Whereas depression in clinical settings has generated substantial research, there have been far fewer studies of anxiety. In part, this may be because of the paucity of brief validated measures for anxiety compared with the numerous measures for depression,^{7,8} such as the Primary Care Evaluation of Mental Disorders 9-item Patient Health Questionnaire (PHQ).⁹⁻¹¹ This situation is unfortunate, given the high prevalence of anxiety disorders, as well as their associated disability and the availability of effective treatments, both pharmacological and nonpharmacological.^{12,13}

Measures of anxiety are seldom used in clinical practice because of their length,

proprietary nature, lack of usefulness as a diagnostic and severity measure,¹⁴⁻¹⁷ and requirement of clinician administration rather than patient self-report.^{18,19} The goal of this study was to develop a brief scale to identify probable cases of GAD and to assess symptom severity. We conducted a study in multiple primary care sites to select the items for the final scale and to evaluate its reliability and validity.

METHODS

GAD SCALE DEVELOPMENT

We first selected potential items for a brief GAD scale. The initial item pool consisted of 9 items that reflected all of the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV) symptom criteria for GAD and 4 items on the basis of review of existing anxiety scales. A 13-item questionnaire was developed that asked patients how often, during the last 2 weeks, they were bothered by each symptom. Response options were "not at all," "sev-

eral days," "more than half the days," and "nearly every day," scored as 0, 1, 2, and 3, respectively. In addition, an item to assess duration of anxiety symptoms was included. Our goal was to determine the number of items necessary to achieve good reliability and procedural, construct, and diagnostic criterion validity.

PATIENT SAMPLE

Patients were enrolled from November 2004 through June 2005 from a research network of 15 primary care sites located in 12 states (13 family practice, 2 internal medicine) administered centrally by Clinvest, Inc (Springfield, Mo). The purpose of the project's first phase (n=2149) was to select the scale items and cutoff scores to be used for making a GAD diagnosis. The purpose of the second phase (n=591) was to determine the scale's test-retest reliability. In all, 2982 subjects were approached and 2739 (91.9%) completed the study questionnaire with no or minimal missing data. To minimize sampling bias, we approached consecutive patients at each site in clinic sessions until the target quota for that week was achieved.

In the first phase, 1654 subjects also agreed to a telephone interview, and of these, a random sample of 965 were interviewed within 1 week of their clinic visit by 1 of 2 mental health professionals (MHPs)—a PhD clinical psychologist and a senior psychiatric social worker. In the study's second phase, 591 subjects who had completed the research questionnaire were sent a 1-page questionnaire that consisted of the 13 potential GAD scale items. Of these, 236 subjects returned the completed 1-page questionnaire with no or minimal missing data within a week of completing the research questionnaire at the clinic. The mean GAD scale score of subjects returning the questionnaire did not differ from that of subjects who did not return the questionnaire. The study was approved by the Sterling Institutional Review Board, Springfield, Mo.

SELF-REPORT RESEARCH QUESTIONNAIRE

Before seeing their physicians, patients completed a 4-page questionnaire that included the 13 items being tested for use in the GAD scale, as well as questions about age, sex, education, ethnicity, and marital status; the Medical Outcomes Study Short-Form General Health Survey (SF-20),^{20,21} which measures functional status in 6 dimensions; and either the 12-item anxiety subscale from the Symptom Checklist-90¹⁶ (first study phase only) or the Beck Anxiety Inventory¹⁴ (second study phase only). Depression was assessed with the PHQ-8, which includes all items from the PHQ-9 except for the item about suicidal ideation; PHQ-8 and PHQ-9 scores are highly correlated and have nearly identical operating characteristics.²² Finally, patients completed items regarding physician visits and disability days during the previous 3 months.

MHP INTERVIEW

The 2 MHPs conducted structured psychiatric interviews by telephone, blinded to the results of the self-report research questionnaire. The interview consisted of the GAD section of the Structured Clinical Interview for DSM-IV,²³ modified with several additional questions to assess in greater detail some of the GAD diagnostic criteria of DSM-IV. The resulting DSM-IV GAD diagnosis, with the DSM-IV 6-month duration criterion, was used as the criterion standard for assessing the validity of the new scale. The interview also included the 13 potential GAD scale items to test agreement between self-report and clinician administration (ie, procedural validity).²⁴

DATA ANALYSIS

The best items for the GAD scale were selected by rank ordering the correlation of each item with the total 13-item scale score in the sample of 1184 patients who did not undergo the MHP interview. Item-total score correlations were reexamined in 2 independent subsamples of the study population: the 965 patients who underwent the MHP interview and the 591 patients in the second phase of the study. In addition, we conducted receiver operating characteristic analyses with varying numbers of items in these 965 patients by using an MHP diagnosis of GAD as the criterion standard. Divergent validity of each item was assessed by calculating the difference between the item correlations with the 13-item anxiety score and the PHQ-8 depression score. Convergent validity was assessed by examining correlations of the final version of the GAD scale with the Beck Anxiety Inventory and the anxiety subscale of the Symptom Checklist-90, even though neither scale is specific for GAD.

To assess construct validity, we used analysis of covariance to examine associations between anxiety severity on the final GAD scale and SF-20 functional status scales, self-reported disability days, and physician visits, controlling for demographic variables. For criterion validity, we investigated sensitivity, specificity, predictive values, and likelihood ratios for a range of cutoff scores of the final scale with respect to the MHP diagnosis. To investigate whether anxiety as measured by the GAD-7 and depression as measured by the PHQ-8 reflect distinct dimensions, we assessed factorial validity by using confirmatory factor analyses. Finally, procedural validity and test-retest reliability were assessed by means of intraclass correlation.²⁵

RESULTS

DESCRIPTION OF PATIENTS

The mean (SD) age of the patients was 47.4 (15.5) years (range, 18-95 years). Most (65%) were female; 80% were white non-Hispanic, 8% were African American, and 9% were Hispanic; 64% were married, 13% were divorced, and 15% were never married; and 31% had a high school degree or equivalent, whereas 62% had attended some college.

ITEM SELECTION FOR THE GAD SCALES

The GAD-7 (**Figure 1**) consists of the 7 items with the highest correlation with the total 13-item scale score ($r=0.75-0.85$). Receiver operating characteristic analysis with this set of items showed an area under the curve (0.906) as good as scales with as much as the full 13-item set. These 7 items also had the highest rank correlations in the developmental sample (n=1184) and the 2 replication samples (n=965 and n=591). The 2 core criteria (A and B) of the DSM-IV definition of GAD are captured by the first 3 items of the scale.²⁶ Of note, 6 of the 7 items had the greatest divergent validity (ie, the highest difference between the item-total scale score correlation and item-PHQ-8 depression score correlation [$\Delta r=0.16-0.21$]). Because each of the 7 items is scored from 0 to 3, the GAD-7 scale score ranges from 0 to 21.

GAD-7				
Over the last 2 weeks, how often have you been bothered by the following problems?	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid as if something awful might happen	0	1	2	3

Total Score = Add Columns + + + +

If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?

Not difficult at all	Somewhat difficult	Very difficult	Extremely difficult
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 1. The generalized anxiety disorder 7-item (GAD-7) scale.

RELIABILITY AND PROCEDURAL VALIDITY

The internal consistency of the GAD-7 was excellent (Cronbach $\alpha = .92$). Test-retest reliability was also good (intraclass correlation = 0.83). Comparison of scores derived from the self-report scales with those derived from the MHP-administered versions of the same scales yielded similar results (intraclass correlation = 0.83), indicating good procedural validity.

DIAGNOSTIC CRITERION VALIDITY AND SCALE OPERATING CHARACTERISTICS

Table 1 summarizes the operating characteristics of the GAD-7 at various cut points. As expected, as the cut point increases, sensitivity decreases and specificity increases in a continuous fashion. At a cut point of 10 or greater, sensitivity and specificity exceed 0.80, and sensitivity is nearly maximized. Results were similar for men and women and for those aged less and those aged more than the mean age of 47 years. The proportion of primary care patients who score at this level is high (23%). A cut point of 15 or greater maximizes specificity and approximates a prevalence (9%) more in line with current epidemiologic estimates of GAD prevalence in primary care. However, sensitivity at this high cut point is low (48%). Most patients (89%) with GAD had GAD-7 scores of 10 or greater, whereas most patients (82%) without GAD had scores less than 10.

The mean (SD) GAD-7 score was 14.4 (4.7) in the 73 patients with GAD diagnosed according to the MHP and 4.9 (4.8) in the 892 patients without GAD. The prevalence of GAD according to the MHP interview was 9% in women and 4% in men. In the entire sample of 2739 patients, the mean GAD-7 score was 6.1 in women and 4.6 in men.

Although the GAD-7 scale inquires about symptoms in the past 2 weeks, the criterion-standard MHP interview required at least a 6-month duration of symptoms consistent with DSM-IV diagnostic criteria for GAD. Nonetheless, the operating characteristics of the scale were good because most patients with high symptom scores had

chronic symptoms. Of the 433 patients with GAD-7 scores of 10 or greater, 96% had symptoms for 1 month or more, and 67% had symptoms for 6 months or more.

CONSTRUCT VALIDITY

There was a strong association between increasing GAD-7 severity scores and worsening function on all 6 SF-20 scales (**Table 2**). As GAD-7 scores went from mild to moderate to severe, there was a substantial stepwise decline in functioning in all 6 SF-20 domains. Most pairwise comparisons within each SF-20 scale between successive GAD-7 severity levels were significant. The relationship between GAD severity and functional impairment was similar in men and women.

Figure 2 illustrates graphically the relationship between increasing GAD-7 scale scores and worsening functional status. Decrements in SF-20 scores are shown in terms of effect size (ie, the difference in mean SF-20 scores, expressed as the number of SDs, between each GAD-7 interval subgroup and the reference group). The reference group is the group with the lowest GAD-7 scores (ie, 0-4), and the SD used is that of the entire sample. Effect sizes of 0.5 and 0.8 are typically considered moderate and large between-group differences, respectively.²⁷

When the GAD-7 was examined as a continuous variable, its strength of association with the SF-20 scales was concordant with the pattern seen in Figure 2. The GAD-7 correlated most strongly with mental health (0.75), followed by social functioning (0.46), general health perceptions (0.44), bodily pain (0.36), role functioning (0.33), and physical functioning (0.30).

Table 3 shows the association between GAD-7 severity levels and 3 other measures of construct validity: self-reported disability days, clinic visits, and the general amount of difficulty patients attribute to their symptoms. Greater levels of anxiety severity were associated with a monotonic increase in disability days, health care use, and symptom-related difficulty in activities and relationships. When the GAD-7 was examined as a continuous variable, its correlation was 0.27 with disability days, 0.22 with physician visits, and 0.63 with symptom-related difficulty.

Convergent validity of the GAD-7 was good, as demonstrated by its correlations with 2 anxiety scales: the Beck Anxiety Inventory ($r = 0.72$) and the anxiety subscale of the Symptom Checklist-90 ($r = 0.74$). Consistent with results of previous studies of anxiety and depression,^{4,28} the GAD-7 and Symptom Checklist-90 anxiety scales also strongly correlated with our depression measure, the PHQ-8 ($r = 0.75$ and $r = 0.74$, respectively). Nonetheless, measuring anxiety and depression was complementary rather than duplicative. We determined the prevalence of high anxiety and high depression symptom severity in our sample, defined as severe scores (≥ 15) on the GAD-7 and PHQ-8 depression scales, respectively. In the 2114 patients who completed the GAD-7 and the PHQ-8, there were 1877 (88.8%) patients with neither high anxiety nor high depression, 99 (4.68%) with high anxiety only, 68 (3.2%) with high depression only, and 70 (3.31%) with high anxiety and high depression. Thus, more than half (99/169) of patients with high anxiety scores did not

Table 1. Operating Characteristics of GAD-7 at Different Cutoffs*

GAD-7 Scale Score†	Sensitivity, %	Specificity, %	PPV, %	NPV, %	LR+	Prevalence, %
8	92	76	24	99	3.8	29
9	90	79	26	99	4.3	26
10	89	82	29	99	5.1	23
11	82	85	31	98	5.5	20
12	73	89	35	98	6.5	16
13	66	91	38	97	7.7	13
14	56	92	37	96	7.2	12
15	48	95	42	96	8.7	9

Abbreviations: GAD-7, generalized anxiety disorder 7-item scale; LR+, likelihood ratio for a positive test; NPV, negative predictive value; PPV, positive predictive value.

*In 965 patients who underwent structured psychiatric interview by a mental health professional to determine the presence of generalized anxiety disorder by using *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* diagnostic criteria.

†The actual score is greater than or equal to the score shown.

Table 2. Relationship Between GAD-7 Severity Score and SF-20 Health-Related Quality of Life Scales*

Level of Anxiety Severity GAD-7 Scale Score	Mean (95% Confidence Interval) SF-20 Scale Score					
	Mental	Social	Role	General	Pain	Physical
Minimal 0-4 (n = 1182)	82 (81-83)	91 (89-92)	84 (82-86)	68 (67-69)	71 (70-72)	84 (82-85)
Mild 5-9 (n = 511)	65 (64-66)	79 (77-81)	69 (66-73)	52 (50-54)	56 (54-58)	74 (72-76)
Moderate 10-14 (n = 264)	54 (52-55)	69 (66-71)	59 (54-63)	43† (40-45)	51† (48-54)	66† (63-69)
Severe 15-21 (n = 171)	41 (39-43)	55 (52-59)	46 (40-52)	39† (36-43)	47† (43-50)	61† (58-65)

Abbreviations: GAD-7, generalized anxiety disorder 7-item scale; SF-20, Medical Outcomes Study Short-Form General Health Survey.

*SF-20 scores are adjusted for age, sex, race, education, and study site. Point estimates for the mean and 95% confidence intervals ($\pm 1.96 \times$ standard error of the mean) are displayed. Number of patients adds to 2128 because of missing data. Missing data for any subscale of SF-20 was less than 5%.

†Pairwise comparisons within each scale that are not significant from one another. However, most pairwise comparisons of mean SF-20 scores with each GAD-7 scale level within each scale are significant at $P < .05$ by using a Bonferroni correction for multiple comparisons.

have high depression scores. Also, when patients had high anxiety and high depression scores, there was an additive effect on the SF-20 mental health and social functioning scales, as well as self-reported disability days and health care use.

FACTORIAL VALIDITY

Principal component analysis of a set of 15 items that includes the 8 depression items of the PHQ-8 and the 7 anxiety items of the GAD-7 indicated that the first 2 emergent factors had an eigenvalue greater than 1. Sixty-three percent of the total variance was explained by the first 2 factors. The varimax-rotated component-matrix clearly confirmed the original allocation of the items to the PHQ scales, with all depression items having the highest factor loadings on 1 factor (0.58-0.75) and all anxiety items having the highest factor loadings on the second factor (0.69-0.81).

COMMENT

This study has several major findings. First, a 7-item anxiety scale—the GAD-7—is a useful tool with strong cri-

terion validity for identifying probable cases of GAD. Second, the scale is also an excellent severity measure as demonstrated by the fact that increasing scores on the GAD-7 are strongly associated with multiple domains of functional impairment and disability days. Third, although many patients had anxiety and depressive symptoms, factor analysis confirms GAD and depression as distinct dimensions.

This study reports the development and validation of a measure for evaluating the presence and severity of GAD in clinical practice, the GAD-7, one of the few GAD measures that is also specifically linked to the *DSM-IV (Text Revision)* criteria.^{19,26} A score of 10 or greater on the GAD-7 represents a reasonable cut point for identifying cases of GAD. Cut points of 5, 10, and 15 might be interpreted as representing mild, moderate, and severe levels of anxiety on the GAD-7, similar to levels of depression on the PHQ-9.¹⁰ The GAD-7 may be particularly useful in assessing symptom severity and monitoring change across time, although its responsiveness to change remains to be tested in treatment studies.

Construct validity was demonstrated by the fact that increasing scores on the GAD-7 scale were strongly associated with multiple domains of functional impair-

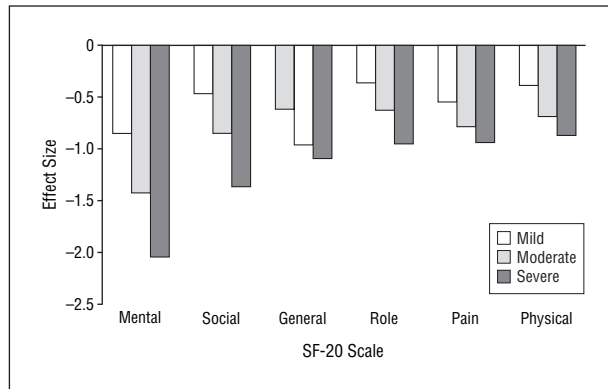


Figure 2. Relationship between anxiety severity as measured with the generalized anxiety disorder 7-item (GAD-7) scale and decline in functional status as measured with the 6 subscales of the Medical Outcomes Study Short-Form General Health Survey (SF-20). The decrement in SF-20 scores is shown as the difference between each GAD-7 scale severity group and the reference group (ie, those with GAD-7 scale scores of 0 to 4). Effect size is the difference in group means divided by the SD.

ment. Furthermore, there was a strong association with self-reported disability days and a modest association with increased health care use.

To facilitate assessment of change in severity of anxiety symptoms, the GAD-7 asks about recent symptoms (ie, in the past 2 weeks). However, most patients with high scores had chronic symptoms, which is why the operating characteristics proved good with use of our criterion-standard MHP interviews based on the conventional GAD duration criterion of 6 months. However, the National Comorbidity Survey showed that patients with episodes of 1 to 5 months do not differ greatly from those with episodes of 6 months or more in onset, persistence, impairment, comorbidity, parental GAD, or sociodemographic correlates.⁵ Kessler et al⁵ conclude that there is little basis for excluding these people from a diagnosis. Notably, 96% of patients with GAD-7 scores of 10 or greater in our primary care sample had symptoms of a month or more, whereas 67% had symptoms of 6 months or longer. It may be that in treatment trials in which response to therapy is evaluated, assessing GAD symptom change during a shorter time (eg, the past week) may be desirable.

The high comorbidity of anxiety and depressive disorders and the high correlation between depressive and anxiety measures is well known.^{17,29} Not surprisingly, our depression measure, the PHQ-8, strongly correlated with the GAD-7 and the Symptom Checklist-90 anxiety scales. Nonetheless, factor analysis confirmed the value of assessing anxiety and depression as 2 separate dimensions. In addition, a number of patients with high anxiety symptoms according to the GAD-7 did not have high depression symptom severity, and patients with increasing severity of anxiety symptoms had corresponding greater impairment in multiple domains of functional status. Together, these findings indicate that using only a depression measure to identify depressed patients who may benefit from treatment will miss a clinically important part of the patient population with disabling anxiety who also would benefit from treatment.

Table 3. Relationship Between GAD-7 Anxiety Severity Score and Disability Days, Symptom-Related Difficulty, and Clinic Visits*

Level of Anxiety Severity GAD-7 Scale Score	Mean No. of Disability Days (95% CI)†	Mean No. of Physician Visits (95% CI)†	Percentage of Symptom-Related Difficulty‡
Minimal 0-4 (n = 1182)	3.9 (3.0-4.7)	1.2 (1.1-1.3)	15.0
Mild 5-9 (n = 511)	7.5 (6.2-8.7)	1.7 (1.5-1.9)	5.5
Moderate 10-14 (n = 264)	10.7 (8.9-12.4)	2.2 (1.9-2.5)	13.7
Severe 15-21 (n = 171)	16.8 (14.6-19.1)	2.4 (2.0-2.8)	47.4

Abbreviations: GAD-7, generalized anxiety disorder 7-item scale; CI, confidence interval.

*All pairwise comparisons between each GAD-7 scale severity level are significant at $P < .05$ by using a Bonferroni correction for multiple comparisons, with the exception of mean physician visits at moderate vs severe GAD-7 score severity levels.

†Disability days refers to number of days in the past 3 months that the patients' symptoms interfered with their usual activities. Physician visits also refers to the past 3 months. Both are self-reported, and means are adjusted for age, sex, race, education, study site, and number of physical disorders. Number of patients adds to 2128 because of missing data. Missing data were less than 5%.

‡Response to single question: "How difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?" The 4 response categories are "not difficult at all," "somewhat difficult," "very difficult," and "extremely difficult." Symptom-related difficulty in this table refers to patients reporting "very" or "extremely" difficult.

Several limitations of our study should be noted. First, the GAD-7 scale focuses on only 1 anxiety disorder, although there are many patients with other anxiety disorders, such as social phobia and posttraumatic stress disorder, who need clinical attention. However, GAD is one of the most common mental disorders seen in outpatient practice. Second, the GAD-7 provides only probable diagnoses that should be confirmed by further evaluation. Third, because our study was cross-sectional, prospective observational and treatment studies are needed to determine the responsiveness of the GAD-7 in assessing change across time. Because there is already evidence for the responsiveness of the PHQ-9 and PHQ-2 depression scales,^{30,31} future research also likely will demonstrate that the GAD-7 scale is useful in assessing changes in the severity of anxiety over time.

This study has a number of strengths, including its large sample size, diverse clinical settings, and its generalizability to primary care, where most patients with anxiety and depression are treated.² Also, the GAD-7 is efficient in that it is brief and can be completed entirely by the patient. This latter feature is particularly important, given the time constraints and competing demands for busy clinicians.³² Although the GAD-7 was developed and validated in primary care, we expect that, like the PHQ-9 depression measure, the GAD-7 will have considerable utility in busy mental health settings and clinical research, which is especially important given the high prevalence and substantial disability associated with GAD.

Accepted for Publication: January 2, 2006.

Correspondence: Robert L. Spitzer, MD, Department of Psychiatry, New York State Psychiatric Institute, Unit 60, 1051 Riverside Dr, New York, NY 10032 (RLS8@columbia.edu).

Funding/Support: The development of the GAD-7 scale was underwritten by an unrestricted educational grant from Pfizer Inc (New York, NY). Dr Spitzer had full access to the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Acknowledgment: Mark Davies, MS, assisted in the statistical analysis. Jeffrey G. Johnson, PhD, assisted in data collection and commented on early drafts. Diane Engel, MSW, also assisted in data collection.

REFERENCES

1. Olfson M, Fireman B, Weissman MM, Leon AC, Sheehan DV, Kathol RG. Mental disorders and disability among patients in a primary care group practice. *Am J Psychiatry*. 1997;154:1734-1740.
2. Roy-Byrne PP, Wagner A. Primary care perspectives on generalized anxiety disorder. *J Clin Psychiatry*. 2004;65(suppl 13):20-26.
3. Leon AC, Olfson M, Broadhead WE, et al. Prevalence of mental disorders in primary care: implications for screening. *Arch Fam Med*. 1995;4:857-861.
4. Wittchen HU, Zhao S, Kessler RC, Eaton WW. *DSM-III-R* generalized anxiety disorder in the National Comorbidity Survey. *Arch Gen Psychiatry*. 1994;51:355-364.
5. Kessler RC, Brandenburg N, Lane M, et al. Rethinking the duration requirement for generalized anxiety disorder: evidence from the National Comorbidity Survey Replication. *Psychol Med*. 2005;35:1073-1082.
6. Kessler RC, Keller MB, Wittchen HU. The epidemiology of generalized anxiety disorder. *Psychiatr Clin North Am*. 2001;24:19-39.
7. Williams JW, Pignone M, Ramirez G, Perez SC. Identifying depression in primary care: a literature synthesis of case-finding instruments. *Gen Hosp Psychiatry*. 2002;24:225-237.
8. Löwe B, Spitzer RL, Gräfe K, et al. Comparative validity of three screening questionnaires for *DSM-IV* depressive disorders and physicians' diagnoses. *J Affect Disord*. 2004;78:131-140.
9. Spitzer RL, Kroenke K, Williams JB; Patient Health Questionnaire Primary Care Study Group. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA*. 1999;282:1737-1744.
10. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16:606-613.
11. Löwe B, Gräfe K, Zipfel S, Witte S, Loerch B, Herzog W. Diagnosing *ICD-10* depressive episodes: superior criterion validity of the Patient Health Questionnaire. *Psychother Psychosom*. 2004;73:386-390.
12. Fricchione G. Clinical practice: generalized anxiety disorder. *N Engl J Med*. 2004;351:675-682.
13. Westen D, Morrison K. A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: an empirical examination of the status of empirically supported therapies. *J Consult Clin Psychol*. 2001;69:875-899.
14. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol*. 1988;56:893-897.
15. Zimmerman M, Mattia JI. A self-report scale to help make psychiatric diagnoses: the Psychiatric Diagnostic Screening Questionnaire. *Arch Gen Psychiatry*. 2001;58:787-794.
16. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a measure of primary symptom dimensions. *Mod Probl Pharmacopsychiatry*. 1974;7:79-110.
17. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale: an updated literature review. *J Psychosom Res*. 2002;52:69-77.
18. Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol*. 1959;32:50-55.
19. Antony MM, Orsillo SM, Roemer L, eds. *Practitioner's Guide to Empirically Based Measures of Anxiety*. New York, NY: Kluwer Academic/Plenum Publishers; 2001.
20. Stewart AL, Hays RD, Ware JE Jr. The MOS short-form general health survey: reliability and validity in a patient population. *Med Care*. 1988;26:724-735.
21. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36), I: conceptual framework and item selection. *Med Care*. 1992;30:473-483.
22. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann*. 2002;9:1-7.
23. First MB, Spitzer RL, Williams JBW, Gibbon M. *Structured Clinical Interview for DSM-IV (SCID)*. Washington, DC: American Psychiatric Association; 1995.
24. Pinto-Meza A, Serrano-Blanco A, Penarrubia MT, Blanco E, Haro JM. Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? *J Gen Intern Med*. 2005;20:738-742.
25. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures: statistics and strategies for evaluation. *Control Clin Trials*. 1991;12:142S-158S.
26. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR (Text Revision)*. 4th ed. Washington, DC: American Psychiatric Association; 2000.
27. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care*. 1989;27:S178-S189.
28. Carter RM, Wittchen HU, Pfister H, Kessler RC. One-year prevalence of sub-threshold and threshold *DSM-IV* generalized anxiety disorder in a nationally representative sample. *Depress Anxiety*. 2001;13:78-88.
29. Clark DA, Steer RA, Beck AT. Common and specific dimensions of self-reported anxiety and depression: implications for the cognitive and tripartite models. *J Abnorm Psychol*. 1994;103:645-654.
30. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. *Med Care*. 2004;42:1194-1201.
31. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a 2-item questionnaire (PHQ-2). *J Psychosom Res*. 2005;58:163-171.
32. Klinkman MS. Competing demands in psychosocial care: a model for the identification and treatment of depressive disorders in primary care. *Gen Hosp Psychiatry*. 1997;19:98-111.