An Ultra-Brief Screening Scale for Anxiety and Depression: The PHQ–4

Kurt Kroenke, M.D., Robert L. Spitzer, M.D.
Janet B.W. Williams, D.S.W., Bernd Löwe, M.D., Ph.D.

Background: *The most common mental disorders in both outpatient settings and the general population are depression and anxiety, which frequently coexist. Both of these disorders are associated with considerable disability.* **Objective:** *When the disorders co-occur, the disability is even greater. Authors sought to test an ultra-brief screening tool for both.* **Method:** *Validated two-item ultra-brief screeners for depression and anxiety were combined to constitute the Patient Health Questionnaire for Depression and Anxiety (the PHQ–4). Data were analyzed from 2,149 patients drawn from 15 primary-care clinics in the United States.* **Results:** *Factor analysis confirmed two discrete factors (Depression and Anxiety) that explained 84% of the total variance. Increasing PHQ–4 scores were strongly associated with functional impairment, disability days, and healthcare use. Anxiety had a substantial effect on functional status that was independent of depression.* **Conclusion:** *The PHQ–4 is a valid ultra-brief tool for detecting both anxiety and depressive disorders.* (Psychosomatics 2009; 50:613–621)

The most common mental disorders in both outpatient settings and the general population are depression and anxiety,[1–5] which frequently coexist.[6,7] Both of these disorders are associated with considerable disability, and when they co-occur, the disability is even greater.[8–10] Moreover, the first-line treatment (e.g., antidepressants, cognitive-behavioral therapy) is often similar for both depression and anxiety.

A brief, nine-item measure (the PHQ–9) has been shown to have excellent reliability, as well as criterion, construct, and procedural validity for depressive disorders.[11] Similar reliability and validity have been established for a brief, seven-item measure (the GAD–7) in assessing the most common anxiety disorders found in clinical practice.[12,13] An even shorter, two-item measure (the PHQ–2), consisting of core criteria for depression, as well as a two-item measure for anxiety (the GAD–2), have also been shown to be excellent screening tools.[13,14] For example, a cutpoint of ≥3 on the 0-to-6–point PHQ–2 scale has a sensitivity of 83% and specificity of 90% for major depressive disorder.[14] Similarly, a cutpoint of ≥3 on the 0-to-6–point GAD–2 scale has reasonable sensitivity for generalized anxiety disorder (88%), panic disorder (76%), and social anxiety disorder (70%), and moderate sensitivity for posttraumatic stress disorder (59%), and good specificity (81%–83%) for all four disorders.[13]

This article analyzes data from a large sample of primary-care patients (N=2,149), to examine attributes of a four-item composite measure (the PHQ–4) that combines these two-item scales: the PHQ–2 and the GAD–2. A recent metaanalysis of ultra-brief screeners for depression found that 2- to 3-item screeners performed better

Received October 7, 2007; accepted October 16, 2007. From Regenstrief Institute, Inc. and the Dept. of Medicine, IN University, Indianapolis, IN (KK); the Biometrics Research Dept., NY State Psychiatric Institute and Dept. of Psychiatry, Columbia University, NY, NY (RLS, JBWW); and the Dept. of Psychosomatic Medicine and Psychotherapy, University of Hamburg, Germany (BL). Send correspondence and reprint requests to Kurt Kroenke, M.D., Regenstrief Institute, 5th Floor, 1050 Wishard Blvd., Indianapolis, IN 46202. e-mail: kkroenke@regenstrief.org

than single-item screeners.[15] Of note, this metaanalysis revealed that the ultra-brief depression measure for which there are the most published data are the PHQ–2, the Depression subscale of the PHQ–4. The present article addresses several important questions:

1) How does the ultra-brief PHQ–4 compare with longer depression and anxiety measures in terms of important measures of construct validity; that is, health-related quality of life, disability days, and healthcare utilization?
2) What is the prevalence of anxiety and depression in primary care that is detected by the PHQ–4 and what are their individual and combined effects on impairment?

Answers to these questions will help to determine the potential value of an ultra-brief anxiety/depressive screener for the busy clinician contending with abbreviated outpatient visits and many competing demands.

METHOD

Patient Sample

The Patient Health Questionnaire (PHQ) Anxiety Study[12] was conducted to develop a short measure to assess generalized anxiety disorder. Patients were enrolled from a research network of 15 primary-care sites (13 family practice and 2 internal medicine sites) located in 12 different states administered centrally by Clinvest, Inc., Springfield, MO, from November 2004 to June 2005. The study was approved by the Sterling Institutional Review Board.

The Generalized Anxiety Disorder–7 scale (GAD–7) was developed and validated in 2,149 patients. In the original study, 2,982 subjects were invited to participate, and 2,740 (92%) completed the four-page questionnaire and had little-or-no missing data. To minimize sampling bias, consecutive patients were approached at each site in clinic sessions until the target quota for that week was achieved. Of the 2,740 participants, the first 2,149 were used for development and validation of the Generalized Anxiety Disorder–7 scale (GAD–7) and constitute the sample for this article.

The Self-Report Research Questionnaire

Before seeing their doctors, patients completed a four-page questionnaire. Anxiety was assessed with the GAD–7, a seven-item measure that has recently been validated as an assessment tool for the four most common anxiety disorders in clinical practice: generalized anxiety, panic, social anxiety, and posttraumatic stress disorders.[12,13] Depression was assessed with the PHQ–8, which includes all items from the PHQ–9 except the item on suicidal ideation; PHQ–8 and PHQ–9 scores are highly correlated and have nearly identical operating characteristics.[16] Also, the questionnaire included questions about age, gender, education, ethnicity, and marital status, and the Medical Outcomes Study Short-Form General Health Survey (SF–20), which measures functional status on six dimensions.[17] Finally, patients completed items regarding physician visits and disability days during the previous 3 months.

Structured Psychiatric Interview

In a random sample of 965 subjects, two mental health professionals conducted structured psychiatric interviews by telephone, blinded to the results of the self-report research questionnaire, to establish independent, criteria-based diagnoses according to the Diagnostic and Statistical Manual of Mental Disorders, 4th Revision (DSM–IV).[18] Details of the sampling and structured interview have been previously described.[12] In this study, the criterion standard mental health professional interview included diagnostic modules for anxiety, but not depressive, disorders. This was partly because the principal study aim was to validate a new anxiety measure (the GAD–7) and also because the operating characteristics of the PHQ–2 for detecting depressive disorders have been well established in previous studies.[14,19–21]

Composite Depression/Anxiety Screening Scale (the PHQ–4)

Both the first two items of the PHQ–8, known as the PHQ–2, as well as the first two items of the GAD–7, known as the GAD–2, have been shown to have excellent operating characteristics for assessing depressive and anxiety disorders.[13,14] The PHQ–2 consists of the two core criteria for depressive disorders. The GAD–2 consists of the two core criteria for generalized anxiety disorder that have also been shown to be good screening items for panic, social anxiety, and posttraumatic stress disorders.[13] Therefore, we combined these two-item measures into a composite four-item scale called the PHQ–4 (see Appendix 1). As with the parent scales, the PHQ–4 begins with the stem question: "Over the last 2 weeks, how often have

APPENDIX 1. The Four-Item Patient Health Questionnaire (PHQ–4) for Anxiety and Depression

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
Feeling nervous, anxious, or on edge	0	1	2	3
Not being able to stop or control worrying	0	1	2	3
Feeling down, depressed, or hopeless	0	1	2	3
Little interest or pleasure in doing things	0	1	2	3

you been bothered by the following problems?" Responses are scored as 0 ("not at all"), 1 ("several days"), 2 ("more than half the days"), or 3 ("nearly every day"). Therefore, the total score on this composite measure ranges from 0 to 12.

Data Analysis

Construct validity was assessed by analysis of covariance to examine associations between symptom severity on the PHQ–4 scale and SF–20 functional status scales, self-reported disability days, and physician visits, controlling for age, race, sex, educational level, and study site. For some analyses, PHQ–4 scores were operationally categorized as normal (0–2), mild (3–5), moderate (6–8), and severe (9–12). To determine whether the construct validity of the PHQ–4 is comparable to its longer parent measures, the GAD–7 and PHQ–8, we examined the correlations of each of these scales with the six domains of the SF–20. To investigate whether the anxiety and depression scales of the PHQ–4 reflect distinct dimensions, factorial validity was assessed with confirmatory factor analyses. To determine whether the anxiety (GAD–2) and depression (PHQ–2) subscales of the PHQ–4 have different operating characteristics in detecting anxiety disorders, receiving operating characteristic (ROC) analyses were conducted to determine the area under the curve (AUC) for each of the four anxiety disorders. Calculations of AUCs and statistical comparisons were performed with a nonparametric procedure that accounted for the correlated nature of the data (i.e., GAD–2 and PHQ–2 AUCs were, appropriately, calculated on the same persons).[22]

Finally, to examine the independent as well as comorbid effects of anxiety and depression, we compared the SF–20 scores in patients with anxiety-only, depression-only, neither, and both.

RESULTS

Description of Patients

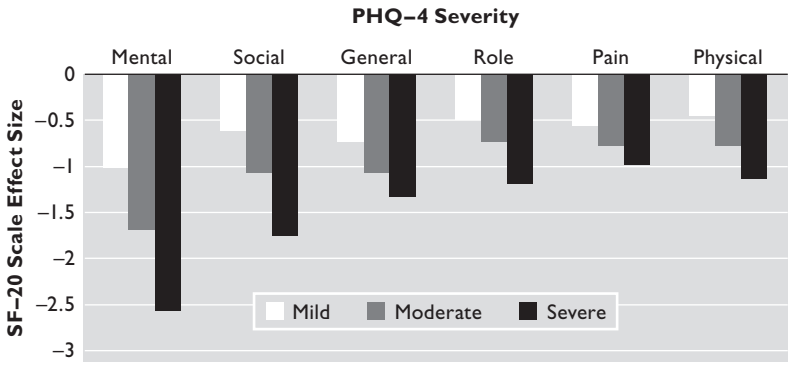
The mean age of the patients was 47.2 (standard deviation [SD]: 15.4) years, with a range of 18–95; 66% were women; 81% were non-Hispanic white, 8% were African American, 8% were Hispanic; 64% were married; 7% had less than a high school education, 31% had a high school degree or equivalent, 62% had some college. The mean PHQ–8 depression score in the sample was 5.1 (SD: 5.2), and the mean GAD–7 anxiety score was 5.3 (SD: 5.3). The proportion of patients with at least moderate depression or anxiety symptoms (defined as a score $\geq$10 on either scale) was 19% and 20%, respectively, and the proportion with moderately severe depression or anxiety symptoms (defined as a score $\geq$15 on either scale) was 7% and 8%, respectively. The mean PHQ–4 score was 2.5 (SD: 2.8), with 61.5% having normal (0–2) scores, 23.3% having mild (3–5) elevations, 10.6% having moderate (6–8) elevations, and 4.6% having severe (9–12) elevations.

Construct Validity of the PHQ–4

Figure 1 illustrates graphically the relationship between increasing PHQ–4 scores and worsening functional status. Decrements in SF–20 scores are shown in terms of the effect size, which is the difference in mean SF–20 scores expressed as the number of SDs between each PHQ–4 interval subgroup and the reference group. The reference group is the group with the lowest PHQ–4 scores (i.e., 0–2), and the SD used is that of the entire sample. Most pairwise comparisons within each SF–20 scale between successive PHQ–4 levels are highly significant ($p<0.001$). Effect sizes of 0.5 and 0.8 represent moderate and large between-group differences, respectively.[23] As PHQ–4 scores go from normal (0–2) to mild (3–5) to moderate (6–8) to severe (9–12), there is a substantial stepwise decline in functioning in all six SF–20 domains.

When the PHQ–4 was examined as a continuous variable, its strength of association with the SF–20 scales was concordant with the pattern seen in Figure 1. As shown in Table 1, the PHQ–4 correlated most strongly with mental health (0.80), followed by social functioning (0.52), gen-

FIGURE 1. Relationship Between Anxiety and Depression Symptom Severity, as Measured by the PHQ–4, and Decline in Functional Status, as Measured by the Six Subscales of the SF–20



The decrement in SF-20 scores is shown as the difference between each PHQ-4 severity group and the reference group (i.e., those with PHQ-4 scores of 0 to 2). Mild, moderate, and severe PHQ-4 categories represent scores of 3-5, 6-8, and 9-12, respectively. Effect size is the difference in group means divided by the standard deviation.

TABLE 1. Comparison of Correlations Among PHQ Anxiety/Depression Scales and SF–20 Scales in 2,149 Patients

	Anxiety		Depression		Composite	
	7 Items	2 Items	8 Items	2 Items	15 Items	4 Items
SF–20 Functional Status Scale	GAD–7	GAD–2	PHQ–8	PHQ–2	Combined GAD–7 and PHQ–8	PHQ–4
Possible scale range	0–21	0–6	0–24	0–6	0–45	0–12
Mean (SD) in sample	5.3 (5.3)	1.4 (1.7)	5.1 (5.2)	1.0 (1.4)	10.5 (9.8)	2.5 (2.8)
Cronbach's alpha	0.92	0.82	0.90	0.81	0.94	0.85
Correlations with SF–20 scales						
Mental Health	0.75	0.72	0.74	0.72	0.80	0.80
Social Functioning	0.46	0.44	0.56	0.50	0.55	0.52
General Health Perceptions	0.44	0.40	0.55	0.47	0.53	0.48
Role Functioning	0.33	0.29	0.45	0.37	0.42	0.37
Bodily Pain	0.36	0.32	0.43	0.33	0.43	0.36
Physical Functioning	0.30	0.28	0.43	0.36	0.39	0.36

Amount of missing data for any given SF–20 domain was less than 5%.

PHQ: Patient Health Questionnaire; SF–20: Medical Outcomes Study Short-Form General Health Survey; GAD: Generalized Anxiety Disorder; SD: standard deviation.

eral health perceptions (0.48), role functioning (0.37), bodily pain (0.36), and physical functioning (0.36).

Notably, the PHQ–4 composite scale correlates as well as or better than its two components (GAD–2 and PHQ–2) with the SF–20 scales (Table 1). The 15-item composite scale correlated as well as or better than its component anxiety scale (the GAD–7) and similar to its component depression scale (the PHQ–8) with the SF–20 scales. As expected, the longer anxiety, depression, and composite scales had slightly higher correlations than their ultra-brief versions, although the absolute differences were small. The internal reliability (Cronbach α) was good (>0.80) for all scales.

Table 2 demonstrates the strong incremental relationship between increasing symptom severity on the PHQ–4 scales and self-reported disability days. Also, increasing symptom severity was moderately associated with increased healthcare use. Importantly, increasing scores on the PHQ–4 have a similarly strong dose–response relationship to disability days and physician visits as the longer anxiety (GAD–7) and depression (PHQ–8) scales.

Factorial Validity of the PHQ–4

Principal-component analysis of a set of four items (i.e., PHQ–4) that includes the two depression items of the

TABLE 2. Relationship Among Anxiety/Depression Scores (GAD–7, PHQ–4, and PHQ–9) and Measures of Construct Validity (Self-Report Disability Days and Physician Visits)

Level of Severity (Score Range)			Mean Disability Days (95% CI)			Mean Physician Visits (95% CI)		
GAD–7	**PHQ–4**	**PHQ–8**	**GAD–7**	**PHQ–4**	**PHQ–8**	**GAD–7**	**PHQ–4**	**PHQ–8**
None-to-Minimal			3.9	3.9	4.9[a]	1.2	1.2	1.4[a]
(0–4)	(0–2)	(0–4)	(3.0–4.7)	(3.1–4.7)	(3.6–6.2)	(1.1–1.3)	(1.1–1.4)	(1.2–1.7)
Mild			7.5	8.2	5.6[a]	1.7	1.9[a]	1.7[a]
(5–9)	(3–5)	(5–9)	(6.2–8.7)	(6.9–9.4)	(3.7–7.5)	(1.5–1.9)	(1.7–2.1)	(1.3–2.0)
Moderate			10.7	13.0	12.0	2.2[a]	2.1[a,b]	2.5[b]
(10–14)	(6–8)	(10–14)	(8.9–12.4)	(11.1–14.9)	(9.4–14.5)	(1.9–2.5)	(1.8–2.4)	(2.1–2.3)
Severe			16.8	21.6	26.3	2.4[a]	2.8[b]	3.4[b]
(15–21)	(9–12)	(15–24)	(14.6–19.1)	(18.6–24.6)	(22.6–29.7)	(2.0–2.8)	(2.3–3.3)	(2.8–4.1)

GAD: Generalized Anxiety Disorder; PHQ: Patient Health Questionnaire; CI: confidence interval.

Disability days refers to number of days in past 3 months that patients' symptoms interfered with their usual activities. "Physician visits" also refers to past 3 months. Both are by self-report. Means are also adjusted for age, sex, race, education, and study site. Most pairwise comparisons between each severity level within a particular scale (GAD–7, PHQ–4, or PHQ–8). Comparisons for a given variable (i.e., disability days or physician visits) are significant at $p < 0.05$ with Bonferroni correction for multiple comparisons. Only those pairwise comparisons that share a common superscript letter are not significant.

PHQ–2 and the two anxiety items of the GAD–2 indicated that 84% of the total variance was explained by the first two factors. The varimax-rotated component-matrix clearly confirmed the original allocation of the items to the PHQ scales, with the two anxiety items having the highest factor loadings on Factor 1 and the two depression items having the highest factor loadings on Factor 2 (Table 3).

Operating Characteristics of GAD–2 Versus the PHQ–2 for Detecting Anxiety Disorders

Because of the overlap in depression and anxiety in many patients, some have wondered whether depression screening alone might adequately capture patients with clinically significant anxiety as well. Therefore, we examined the differential performance of the anxiety (GAD–2) and depression (PHQ–2) subscales of the PHQ–4. As shown in Table 4, the GAD–2 is significantly better than the PHQ–2 for detecting three of the four most common anxiety disorders in clinical practice, as well as any anxiety disorder (defined as one or more of the four anxiety disorders).

TABLE 3. Factorial Validity of the PHQ–4 in 2,103 Patients

	Factor Loadings	
	Factor 1 (Anxiety)	**Factor 2 (Depression)**
Depression Items (PHQ–2)		
Little interest or pleasure in doing things	0.25	**0.90**
Feeling down, depressed, or hopeless	0.40	**0.82**
Anxiety Items (GAD–2)		
Feeling nervous, anxious, or on edge	**0.87**	0.29
Not being able to stop or control worrying	**0.86**	0.33

Principal-component analysis with varimax rotation. For each item, higher factor loadings are printed in bold.

Anxiety and Depression Comorbidity

The relative effects of anxiety and depression, both individually and together, are illustrated in Figure 2. For this comparison, anxiety was defined as a score of ≥3 on the Anxiety subscale of the PHQ–4, and depression was defined as a score of ≥3 on the Depression subscale of the PHQ–4, with optimal cutpoints for the two scales based on previous research.[13,14] In the sample of patients who completed both scales, there were 1,577 (75%) patients with neither anxiety nor depression, 255 (12.1%) with anxiety only, 93 (4.4%) with depression only, and 178 (8.5%) with both anxiety and depression. Thus, more than half (255/433, or 59%) of patients with high anxiety scores did not have high depression scores. The impairment in mental health associated with anxiety is identical to that associated with depression. The impairment in social functioning and general health perceptions is only slightly less for anxiety than for depression. Moreover, there was an additive effect of anxiety and depression on a number of the SF–20 domains.

DISCUSSION

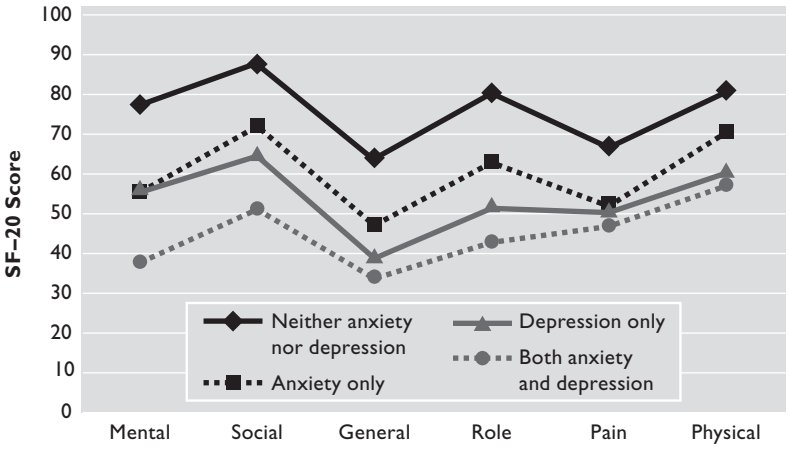
This study has several major findings. First: a four-item scale, the PHQ–4, consisting of two core anxiety items

TABLE 4. Comparison of GAD–2 and PHQ–2 Scales by Receiver Operating Curve Analysis for Four Common Anxiety Disorders

Anxiety Disorder	Area Under the Curve (95% CI)		Comparison of Area Under the Curve	
	GAD–2	PHQ–2	χ^2	p
Generalized anxiety disorder	0.908 (0.876–0.940)	0.813 (0.765–0.862)	17.7	0.0001
Panic disorder	0.848 (0.805–0.891)	0.750 (0.691–0.809)	13.67	0.0002
Social anxiety disorder	0.827 (0.773–0.881)	0.781 (0.723–0.839)	4.64	0.031
Posttraumatic stress disorder	0.802 (0.759–0.844)	0.767 (0.711–0.811)	1.47	0.226
Any anxiety disorder	0.853 (0.823–0.883)	0.780 (0.743–0.817)	16.04	0.0001

Comparisons were made using a nonparametric approach, taking into account the correlated nature of the data.[22]
GAD: Generalized Anxiety Disorder; PHQ: Patient Health Questionnaire; CI: confidence interval.

FIGURE 2. Relationships Among Anxiety, Depression, and Functional Status



and two core depression items, is an efficient ultra-brief tool for identifying individuals who may be suffering from one or both of these common mental disorders. Second: increasing PHQ–4 scores are strongly associated with decrements in multiple domains of functional impairment. Third: the anxiety and depression subscales make unique overall contributions to the PHQ–4, both in terms of factorial and criterion validity. Fourth, and perhaps most importantly: the results indicate that anxiety has a substantial independent effect on functioning, and even more so when comorbid with depression. Therefore, screening for both anxiety and depression, rather than either alone, is advisable.

The PHQ–4 may be particularly useful as a brief screening scale for busy clinicians interested in identifying potential cases of depression and anxiety. The value of this scale is supported by the internal reliability, construct validity, and factorial validity established in this study, as well as the reliability, and criterion, construct, and procedural validity of its two subscales, the PHQ–2 and GAD–2, established in previous studies.[12–14,20] Those with elevated PHQ–4 scores can be further assessed to determine whether full diagnostic criteria are met for either disorder or whether some intervention is indicated.

Construct validity was demonstrated by the fact that increasing PHQ–4 scores were strongly associated with multiple domains of functional impairment. The effect of increasing PHQ–4 symptom severity was greatest for mental health, intermediate for social and general health perceptions, and least for bodily pain and physical functioning, which is the expected rank ordering of functional impairment associated with mental disorders.[24–26] Furthermore, there was a strong association with self-reported disability days and a modest association with increased healthcare utilization.

Despite the high correlation between depression and anxiety measures,[27,28] factor analysis confirmed that the Anxiety and Depression subscales of the PHQ–4 reflect two separate dimensions. Also, ROC analysis demonstrated that the Anxiety subscale is superior to the Depression subscale in detecting anxiety disorders. Therefore, the scores on the two subscales, the PHQ–2 and GAD–2, help

indicate whether the patient is predominantly affected by anxiety or depressive symptoms. Previous research has established that a score of 3-or-greater on the Depression subscale represents a reasonable cut-point for identifying potential cases of major depression or other depressive disorders.[14,20] Likewise, a score of 3-or-greater on the Anxiety subscale represents a reasonable cut-point for generalized anxiety, panic, social anxiety, and posttraumatic stress disorders.[13] The total PHQ–4 score complements the subscale scores as an overall measure of symptom burden, as well as impairment and disability. At the same time, the PHQ–4 is only a screener, and, as with other case-finding measures,[29] the positive predictive value for a DSM–IV disorder, even for individuals with depression or anxiety subscale scores of 3-or-over is limited.[13,14] Thus, an elevated PHQ–4 score is not diagnostic, but is, instead, an indicator for further inquiry to establish the presence or absence of a clinical disorder warranting treatment.

Although depression and anxiety are frequently comorbid,[6,7] assessing for both conditions appears warranted. First, one-quarter of our primary-care sample had anxiety and/or depression, and, of these patients, half had anxiety-only. In fact, had we screened only for depression, 50% of these anxious patients may not have been identified. Second, anxiety and depression both have substantial and independent effects on functional status. Third, although both anxiety and depression often respond to similar first-line treatments such as antidepressants and cognitive-behavioral therapy,[30] psychotherapy, especially cognitive-behavioral therapy, is considered a first-line treatment for anxiety disorders by some experts.[31] Therefore, knowing whether an individual patient has depression, anxiety, or both may be useful in selecting and monitoring treatment. In short, researchers and clinicians should consider screening for both conditions, rather than either one alone. It must be emphasized, however, that to improve patient outcomes, screening by itself is insufficient and that systems of care must be in place to monitor treatment response and adherence, adjust therapy, and collaborate with mental health professionals as needed.[32–35]

Since previous research has established the validity and reliability of the PHQ–2 and GAD–2, what is the added value of coupling them into the composite PHQ–4? First, the construct validity data (Table 2, Table 3, Figure 2) show that the overall PHQ–4 score serves as a general marker of psychological distress, and such measures of potential psychiatric "caseness" (e.g., the GHQ–12,[36,37] the MHI–5,[38] the WHO–5,[39] and the K10/K6[40]) can be valuable in primary-care and population screening. Second, composite measures consisting of Anxiety and Depression subscales, such as the longer, 14-item Hospital Anxiety and Depression Scale, perform well in assessing the symptom severity and caseness of patients with anxiety and depressive disorders in medical, psychiatric, and general population samples.[28] Third, although first-line treatment of anxiety or depression may be similar, comorbidity may adversely affect remission rates. Thus, assessment of both anxiety and depression may be helpful in identifying high-scoring patients with both conditions who may require closer monitoring and/or earlier referral for more specialized treatments, particularly if they fail to improve.

Limitations of this study should be noted. First, although the PHQ–4 is a good measure of caseness, it does not evaluate other important conditions that should be evaluated before treatment, such as suicidality and bipolarity, or the possibility that there is an underlying general-medical condition that accounts for the mood disturbance. Second, in this study we had a criterion standard interview for anxiety but not depressive disorders. However, a large body of work has documented the operating characteristics of the PHQ–2 for depressive disorders.[14,19–21] Third, because this is a cross-sectional study, the responsiveness of an ultra-brief measure like the PHQ–4 for monitoring treatment needs to be determined in prospective trials. Fourth, although our prevalence and impairment data suggest that anxiety and depression both warrant detection, prospective trials focused on joint screening and systematic treatment are needed to determine the added value of screening for both disorders, as opposed to the current emphasis on depression screening only.

The major strengths of this study include its large sample size, diverse clinical settings, and its generalizability to primary care, where the majority of patients with anxiety and depression are treated.[41,42] Because of its excellent operating characteristics, the PHQ–4 may well substitute for its parent scales (the GAD–7 and PHQ–9) as well as even longer depression and anxiety measures for screening purposes. To our knowledge, this ultra-brief, four-item scale is the shortest validated composite measure currently available for assessing depressive and anxiety disorders, and could be very useful in busy clinical settings where these mental disorders commonly coexist.

We thank Jeffrey G. Johnson, Ph.D., and Diane Engel, M.S.W., who assisted in data collection, and Dieter Schellberg, Ph.D., Dept. of Psychosomatic and General

Internal Medicine, University of Heidelberg, Germany, who assisted in some of the data analyses.

Data collection was supported by a grant from Pfizer, Inc. All investigators had complete access to the data, and no funding was provided to conduct the data analysis. The funding source had no role in the design, conduct, or reporting of the study or in the decision to submit the article for publication.

References

1. Spitzer RL, Williams JB, Kroenke K, et al: Utility of a new procedure for diagnosing mental disorders in primary care: The PRIME–MD 1000 Study JAMA 1994; 272:1749–1756
2. 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
3. Demyttenaere K, Bruffaerts R, Posada-Villa J, et al: Prevalence, severity, and unmet need for treatment of mental disorders in The World Health Organization World Mental Health surveys. JAMA 2004; 291:2581–2590
4. Ansseau M, Dierick M, Buntinkx F, et al: High prevalence of mental disorders in primary care. J Affect Disord 2004; 78: 49–55
5. Kessler RC, McGonagle KA, Zhao S, et al: 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. Olfson M, Shea S, Feder A, et al: Prevalence of anxiety, depression, and substance use disorders in an urban general medicine practice. Arch Fam Med 2000; 9:876–883
7. Schonfeld WH, Verboncoeur CJ, Fifer SK, et al: The functioning and well-being of patients with unrecognized anxiety disorders and major depressive disorder. J Affect Disord 1997; 43:105–119
8. Olfson M, Fireman B, Weissman MM, et al: Mental disorders and disability among patients in a primary care group practice. Am J Psychiatry 1997; 154:1734–1740
9. Andrews G, Henderson S, Hall W: Prevalence, comorbidity, disability, and service utilisation: overview of the Australian National Mental Health Survey. Br J Psychiatry 2001; 178:145–153
10. Alonso J, Angermeyer MC, Bernert S, et al. Disability and quality-of-life impact of mental disorders in Europe: results from The European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. Acta Psychiatr Scand 2004; (suppl.):38–46
11. Kroenke K, Spitzer RL, Williams JBW: The PHQ–9: validity of a brief depression severity measure. J Gen Intern Med 2001; 16:606–613
12. Spitzer RL, Kroenke K, Williams JB, et al: A brief measure for assessing generalized anxiety disorder: the GAD–7. Arch Intern Med 2006; 166:1092–1097
13. Kroenke K, Spitzer RL, Williams JBW, et al: Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007; 146:317–325
14. Kroenke K, Spitzer RL, Williams JB: The Patient Health Questionnaire–2: validity of a two-item depression screener. Med Care 2003; 41:1284–1292
15. Mitchell AJ, Coyne JC: Do ultra-short screening instruments accurately detect depression in primary care? a pooled analysis and meta-analysis of 22 studies. Br J Gen Pract 2007; 57:144–151
16. Kroenke K, Spitzer RL: The PHQ–9: A new depression and diagnostic severity measure. Psychiatr Ann 2002; 32:509–521
17. Stewart AL, Hays RD, Ware JE: The MOS Short-Form General Health Survey: reliability and validity in a patient population. Med Care 1988; 26:724–732
18. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition. Washington, DC, American Psychiatric Publishing, 1994
19. Corson K, Gerrity MS, Dobscha SK: Screening for depression and suicidality in a VA primary care setting: two items are better than one item. Am J Manag Care 2004; 10:839–845
20. Lowe B, Kroenke K, Grafe K: Detecting and monitoring depression with a two-item questionnaire (PHQ–2). J Psychosom Res 2005; 58:163–171
21. Li CY, Friedman B, Conwell Y, et al: Validity of The Patient Health Questionnaire–2 (PHQ–2) in identifying major depression in older people. J Am Geriatr Soc 2007; 55:596–602
22. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988; 44: 837–845
23. Kazis LE, Anderson JJ, Meenan RF: Effect sizes for interpreting changes in health status. Med Care 1989; 27:S178–S189
24. 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
25. Pasacreta JV: Measuring depression, in Instruments for Clinical Health-Care Research. Edited by Frank-Stromborg M, Olsen SJ. Sudbury, MA, John & Bartlett Publishers, 1997, pp 342–360
26. Spitzer RL, Kroenke K, Linzer M, et al: Health-related quality of life in primary care patients with mental disorders: results from The Prime-MD 1000 Study. JAMA 1995; 274:1511–1517
27. 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
28. Bjelland I, Dahl AA, Haug TT, et al: The validity of The Hospital Anxiety and Depression Scale: an updated literature review. J Psychosom Res 2002; 52:69–77
29. Williams JW Jr, Pignone M, Ramirez G, et al: Identifying depression in primary care: a literature synthesis of case-finding instruments. Gen Hosp Psychiatry 2002; 24:225–237
30. Barlow D, Allen L, Choate M: Toward a unified treatment for emotional disorders. Behav Ther 2004; 35:205–230
31. McIntosh A, Cohen A, Turnbull N, et al. Clinical Guidelines and Evidence Review for Panic Disorder and Generalised Anxiety Disorder. London, UK, University of Sheffield/London: National Collaborating Centre for Primary Care. NICE Clinical Guideline 22, National Collaborating Centre for Primary Care, 2004
32. Kroenke K: Depression screening is not enough. Ann Intern Med 2001; 134:418–420
33. Pignone MP, Gaynes BN, Rushton JL, et al: Screening for depression in adults: a summary of the evidence for the U.S.

Preventive Services Task Force. Ann Intern Med 2002; 136:765–776
34. Gilbody S, Whitty P, Grimshaw J, et al: Educational and organizational interventions to improve the management of depression in primary care: a systematic review. JAMA 2003; 289: 3145–3151
35. Gilbody S, Bower P, Fletcher J, et al: Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. Arch Intern Med 2006; 166:2314–2321
36. Schmitz N, Kruse J, Tress W: Improving screening for mental disorders in the primary-care setting by combining the GHQ–12 and SCL–90-R subscales. Compr Psychiatry 2001; 42:166–173
37. Henkel V, Mergl R, Kohnen R, et al: Identifying depression in primary care: a comparison of different methods in a prospective cohort study. BMJ 2003; 326:200–201
38. Berwick DM, Murphy JM, Goldman PA, et al: Performance of a five-item mental health screening test. Med Care 1991; 29:169–176
39. Henkel V, Mergl R, Kohnen R, et al: Use of brief depression screening tools in primary care: consideration of heterogeneity in performance in different patient groups. Gen Hosp Psychiatry 2004; 26:190–198
40. Kessler RC, Andrews G, Colpe LJ, et al: Short screening scales to monitor population prevalences and trends in nonspecific psychological distress. Psychol Med 2002; 32:959–976
41. Regier DA, Narrow WE, Rae DS, et al: The de facto U.S. mental and addictive disorders service system: Epidemiologic Catchment Area prospective 1-year prevalence rates of disorders and services. Arch Gen Psychiatry 1993; 50:85–94
42. Young AS, Klap R, Sherbourne CD, et al: The quality of care for depressive and anxiety disorders in the United States. Arch Gen Psychiatry 2001; 58:55–61