Limitations of the Patient Health Questionnaire in Identifying Anxiety and Depression: Many Cases Are Undetected

Shaun M. Eack, M.S.W.[*],
6001 Saint Marie St., Apt. #122, Pittsburgh, PA 15206

Catherine G. Greeno, Ph.D., and
Western Psychiatric Institute and Clinic, University of Pittsburgh, 3811 O'Hara Street, Pittsburgh, PA 15213

Bong-Jae Lee, Ph.D.
Western Psychiatric Institute and Clinic, University of Pittsburgh, 3811 O'Hara Street, Pittsburgh, PA 15213

Abstract

Objective—To determine the concordance between the Structured Clinical Interview for DSM-IV (SCID) and the Patient Health Questionnaire (PHQ) in diagnosing anxiety and depressive disorders.

Method—Fifty women seeking psychiatric services for their children at two mental health centers in Western Pennsylvania were assessed for anxiety and depressive disorders using the SCID and the PHQ.

Results—Twenty-five women met SCID criteria for at least one anxiety disorder, 11 (44%) of whom the PHQ failed to identify. The PHQ was particularly limited in identifying individuals with anxiety disorders other than panic disorder. Seventeen women met SCID criteria for at least one major depressive disorder, 6 (35%) of whom the PHQ failed to identify. The PHQ was particularly limited in identifying depressed individuals with dysthymia.

Conclusions—Caution should be used when screening for anxiety and depression with the PHQ. Implications for improving diagnostic accuracy in social work practice are discussed.

Accurate diagnosis is often regarded as the first and most important step in treating any medical condition. Indeed, social workers, who provide the majority of direct mental health services in the United States (Substance Abuse and Mental Health Services Administration, 2001), rely increasingly on accurate diagnostic information to guide their treatment efforts. Yet, within the field of mental health, clinical diagnoses have been historically unreliable (Matarazzo, 1983). With the advent of the DSM-III, the reliability of clinical diagnoses has improved (Spitzer, Forman, & Nee, 1979), and the introduction of structured assessments has continued to improve diagnostic reliability (e.g., Skre, Onstad, Torgersen, & Kringlen, 1991; Riskind, Beck, Berchick, & Brown, 1987), leading some to argue for the routine use of standardized methods of diagnosing mental disorders (Endicott & Spitzer, 1972; Shear et al., 2000).

Researchers routinely use two forms of structured assessments. Structured interviews, such as the Structured Clinical Interview for DSM-IV (SCID), include a question for every symptom for each disorder in the DSM, and are often considered the "gold standard" in

[*]Correspondence concerning this article should be addressed to Shaun M. Eack, 6001 Saint Marie St., Apt. #122, Pittsburgh, PA 15206. Phone: (412) 596-7657. sme12@pitt.edu.

psychiatric assessment (Basco et al., 2000; Kranzler, Kadden, Burleson, & Babor, 1995; Shear et al., 2000). On the other hand, rapid assessment instruments, such as the Patient Health Questionnaire (PHQ; Spitzer, Kroenke, & Williams, 1999), which is under study here, the Center for Epidemiologic Studies-Depression Scale (CES-D; Radloff, 1977), and the Beck Depression and Anxiety Inventories (Beck, Epstein, Brown, & Steer, 1988; Beck, Steer, & Brown, 1996) consist of a relatively small number of items that attempt to identify the most common psychiatric symptoms and disorders. While using the gold standard is ideal, these instruments are lengthy and require substantial training to administer, which hampers their usefulness for social workers practicing in community settings. Furthermore, financial exigencies in community mental health settings cannot currently permit diagnostic procedures that monopolize clinician time or require royalties to research publishers. As such, some have hoped that rapid assessment instruments could be comparable in accuracy to structured interviews, and could serve as low-cost substitutes or at least excellent screening tools.

The PHQ is a screening instrument used for identifying anxiety and depression, and while it is commonly used throughout research settings, its ability to adequately identify these disorders compared to an established criterion has not been thoroughly investigated. To be an effective screen for anxiety and depression, the PHQ must have a high rate of "true positives", or be able to identify the majority of those cases a gold standard indicates are afflicted. Likewise, having a high rate of "false negatives", or not identifying those cases a gold standard indicates are afflicted, would be the most severe limitation of the PHQ, because many true cases would not be identified for treatment. Other limitations in criterion validity, such as over-identification of diagnostic prevalences or high rates of "false positives", are somewhat less problematic for screening instruments, such as the PHQ, because diagnosis can always be performed in a tiered fashion. For example, individuals who screen positive for depression on the PHQ can always be followed up with a more comprehensive diagnostic assessment. However, if a screening instrument cannot detect a problematic case to begin with, nothing can be done to ensure that the individual suffering from the ailment being screened gets the services he/she needs.

Unfortunately, few studies have investigated the criterion validity of the PHQ. Some studies have indicated that the PHQ converges with clinician diagnosis and other screening instruments for anxiety and depression (Löwe et al., 2003; Löwe et al., 2004; Spitzer, Kroenke, & Williams, 1999). However, clinician diagnosis unaided by structured interviews can have questionable validity (Basco et al., 2000; Kranzler, Kadden, Burleson, & Babor, 1995; Shear et al., 2000), and many of the screening instruments the PHQ has been compared to have unknown criterion validity as well. As such, to date, no known studies in the United States have examined diagnostic accuracy of the PHQ compared a known gold standard, such as the SCID. However, in studies translating the PHQ to other languages, concordance with the SCID has suggested that the PHQ may be an effective screen for anxiety and depressive disorders. In two large studies of German translations of the PHQ, Löwe and colleagues found the PHQ had adequate concordance with the SCID for diagnosing major depressive and panic disorders in primary care settings (Löwe et al., 2003; Löwe et al., 2004). Becker, Zaid, and Faris (2002) found, in an Arabic translation study, adequate concordance between the PHQ and SCID for diagnosing depressive disorders, but poor concordance for diagnosing panic and generalized anxiety disorders. Specifically, they found that the PHQ was insensitive to detecting these disorders, incorrectly classifying much of their anxiety disordered sample as healthy. Such findings suggest that the PHQ may be an adequate screen for depression, but less accurate when screening for anxiety disorders. Surprisingly, the concordance between the original English version of PHQ and the SCID has yet to be explored, and no studies have examined the adequacy of using the PHQ to diagnose anxiety and depressive disorders in community mental health settings. As such, the

effectiveness of the PHQ for diagnosing anxiety and depression in community settings continues to be largely unknown.

The purpose of this study was to determine the accuracy of the PHQ in diagnosing anxiety and depressive disorders in a community mental health setting, by comparing its assessment with the current state-of-the-art method for obtaining psychiatric diagnosis. At a time when enough evidence has accumulated indicating that standardized instruments help improve the reliability of clinical diagnoses, but managed care regulations continue to reward quantity rather than quality of care, it is increasingly important to identify brief, yet valid standardized methods of assessing for psychiatric diagnoses. The PHQ may provide such a method for identifying two of the most common mental disorders, anxiety and depression, and therefore, be of substantial use to social workers in community clinics and the many patients who receive mental health services in community settings.

Method

Participants

Participants consisted of 50 women seeking psychiatric treatment for their children at two community mental health centers in Western Pennsylvania. Participants were predominantly White ($n = 25$) or African American ($n = 22$), with three participants identifying their race as "other". Participants' ages ranged from 23 to 60 ($M = 39.20$, $SD = 9.63$). The majority of participants had graduated from high-school ($n = 37$), and over half had completed some college or formal vocational training ($n = 27$).

Measures

Patient Health Questionnaire—The PHQ is a brief self-report screening utility for identifying common DSM-IV Axis I disorders. For screening anxiety disorders, the PHQ contains a 15-item checklist of DSM-IV symptoms of panic disorder, and a 7-item checklist of other anxiety symptoms for detecting anxiety disorders other than panic. For depressive disorders, the PHQ contains a 9-item checklist of DSM-IV symptoms of major depression and major depressive episodes. Each symptom checklist, whether for anxiety or depression, has a criterion that needs to be met before a diagnosis can be made. Once this criterion is evaluated, the PHQ yields two different dichotomous diagnostic scores. For anxiety, these scores indicate the presence or absence of panic disorder or any anxiety disorder other than panic. For depression, these scores indicate the presence or absence of MDD or a depressive disorder other than MDD. As indicated above, the PHQ has been shown to converge with clinician diagnosis, other screening instruments, and its interview-based predecessor (Becker, Zaid, & Faris, 2002; Kobak et al., 1997; Löwe et al., 2003; Löwe et al., 2004; Spitzer, Kroenke, & Williams, 1999). For example, Spitzer, Kroenke, and Williams (1999) found the overall agreement between the PHQ and its clinician-rated, interviewbased predecessor, the Primary Care Evaluation of Mental Disorders (PRIME-MD; Spitzer et al., 1994), to be within acceptable ranges ($\kappa = .65$); and Löwe et al. (2004) found that the depression portion of the PHQ had good internal consistency ($\alpha = .88$) and converged significantly with the Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983).

Structured Clinical Interview For DSM-IV—The Structured Clinical Interview for DSM-IV Axis I Disorders - Non-Patient Edition (First, Spitzer, Gibbon, Williams, 1996) was used as the criterion, or "gold standard", by which to compare diagnoses made by the PHQ. The SCID is a lengthy structured interview that assesses individuals for DSM-IV Axis I diagnoses. For each Axis I disorder, the SCID yields a score that indicates the presence or absence of the disorder, or whether the person meets some criteria for the disorder but not enough for a full diagnosis. Since this study concerns the presence or absence of a

psychiatric diagnosis, rather than degrees of symptomatology, any individual whom the SCID indicated did not meet full criteria for a psychiatric diagnosis was considered to be healthy. Within the research community, the SCID is widely hailed as the most accurate and reliable diagnostic assessment instrument (Basco et al., 2000; Kranzler, Kadden, Burleson, & Babor, 1995; Shear et al., 2000). For example, Shear et al. (2000) found that the SCID was able to identify almost 50% more diagnoses of depressive disorders that would have otherwise been undetected or misdiagnosed by standard clinician diagnoses in a community mental health setting. Other empirical investigations of the SCID also indicate that it has high temporal consistency and inter-rater reliability, when administered by trained clinicians (see Segal, Hersen, & van Hasselt, 1994 for a review; Zanarini et al., 2000; Zanarini & Frankenburg, 2001). For example, Zanarini and Frankenburg (2001) found the inter-rater (range of $\kappa = .69$ to 1.00) and test-retest (range of $r = .53$ to 1.00) reliability of the SCID to be quite good for most Axis I diagnoses.

Procedure

Women were recruited to participate in a study of their own psychiatric needs, as we have previously found that this population is at high risk for psychiatric disorders (Swartz et al., 2005). Upon consent, participants were interviewed using the SCID by a master's-level clinician certified in its administration. SCID training was conducted in accordance with standards of the Biometrics Division of the New York State Psychiatric Institute, requiring that interviewers achieve 100% concordance with a certified SCID rater for primary and, if present, comorbid diagnoses, on four consecutive interviews. Participants also completed a set of assessments which included the PHQ. A randomly selected subset of assessments were reviewed for accuracy and data entry errors, and revealed no significant errors in the data collection or entry process. This study was approved by the University of Pittsburgh Institutional Review Board, and all participants gave written, informed consent prior to participation.

Results

Diagnostic Prevalences

Twenty-five of the 50 (50%) women met SCID criteria for at least one anxiety disorder. Thirteen (52%) of these women also met SCID criteria for multiple anxiety disorders. Prevalences for panic and all other anxiety disorders combined can be seen in Table 1. Prevalences are reported based on all 50 cases, except for PTSD, as one woman declined to participate in that series of questions. Women meeting criteria for "other anxiety" were diagnostically heterogeneous, with ten (20%) meeting criteria for anxiety not otherwise specified (NOS), seven (14.3%) for post-traumatic stress disorder, seven (14%) for generalized anxiety disorder, four (8%) for specific phobia, two (4%) for agoraphobia without panic disorder, two (4%) for social phobia, and one (2%) for obsessive-compulsive disorder.

Seventeen of the 50 (34%) women met SCID criteria for at least one depressive disorder, six (35%) of whom met SCID criteria for more than one depressive disorder. Prevalences of MDD and "other depression" can be seen in Table 2. Of the individuals who had a depressive disorder other than or in addition to MDD, the majority (71%) met SCID criteria for dysthymic disorder and the rest (29%) met SCID criteria for depressive disorder NOS. Prevalences reported for dysthymic disorder are based on 48 cases, as two women declined to participate in that series of questions.

Accuracy Of The PHQ In Screening Anxiety

Diagnostic concordance of the PHQ and the SCID was calculated separately for panic, anxiety disorders other than panic, and all anxiety disorders combined. As can be seen in Table 1, the PHQ failed to identify a large number of women meeting SCID criteria for an anxiety disorder. Forty percent of women who met SCID criteria for panic were not identified by the PHQ's panic module, and 58% who met SCID criteria for anxiety disorders other than panic were not identified by the PHQ's "other anxiety" module. This large number of false negatives resulted in a quite low percentage of correctly classified true positive cases, or sensitivity, when screening for panic, other, or any anxiety disorder (sensitivity = .60, .42, and .56, respectively). The percentage of individuals correctly classified as healthy, or specificity of the PHQ, was much higher for panic, other, and any anxiety disorder diagnosis (specificity = .95, .85, and .84, respectively), however, this does little to ameliorate problems caused by high false negative rates. In an attempt to account for this alarming rate of false negatives, several post-hoc analyses were conducted, which indicated that false negative rates were not higher for any specific anxiety diagnosis, $\chi^2(7, N = 24) = 9.05$, *ns*, did not differ for individuals with multiple anxiety diagnoses, $\chi^2(1, N = 25) = .34$, *ns*, and were not related to age, $\chi^2(1, N = 25) = 2.10$, *ns*, or race, $\chi^2(1, N = 25) = .06$, *ns*. However, there was a trend for individuals with comorbid depression to be slightly less likely to be undetected by the PHQ, $\chi^2(1, N = 24) = 4.03$, $p = .05$, *OR* = .17.

Accuracy Of The PHQ In Screening Depression

As with anxiety disorders, the concordance between the PHQ and the SCID for diagnosing depression was calculated separately for any depressive disorder, MDD, and depressive disorders other than MDD. As can be seen in Table 2, the PHQ failed to detect a substantial portion of individuals with depressive disorders. The PHQ was severely limited in detecting cases other than MDD. Of the seven individuals who had a depressive disorder other than MDD, the PHQ detected only one. This resulted in a very low level of sensitivity for detecting depressive disorders other than MDD (sensitivity = .14). However, the sensitivity of the PHQ in detecting MDD and any depressive disorder was fair (sensitivity = .65 and .64, respectively). Several post-hoc analyses were conducted in an attempt to elucidate the reasons for the PHQ's high false negative rate in detecting depressive disorders other than MDD. In total, six individuals who were diagnosed with a depressive disorder other than MDD were not detected by the PHQ. This small subsample precluded our use of formal statistical tests. However, visual inspection of the characteristics of this sample indicated that the majority of individuals who were missed by the PHQ had dysthymic disorder ($n = 5$), at least one anxiety disorder ($n = 5$), and comorbid MDD ($n = 4$). No discernible patterns were found that would suggest differences in false negative rates surrounding age or race.

Discussion and Applications to Social Work Practice

The diagnosis of psychiatric disorders has long been problematic. In the absence of clear pathophysiological diagnostic markers for the majority of mental disorders, it has become important for social workers to rely on standardized procedures for arriving at a diagnosis, as these procedures offer the promise of consistency across clinicians and clinics. Unfortunately, current "gold standard" procedures are not widely applicable outside of research laboratories, and the accuracy of more "community friendly" procedures is largely unknown. To the best of our knowledge, this is the first study to examine the concordance between the PHQ, a common screen for anxiety and depression, in its native language, and the SCID, considered by many to be the current gold standard in psychiatric assessment.

Previous research has suggested that the PHQ may be adequate for diagnosing anxiety and depressive disorders (Becker, Zaid, & Faris, 2002; Löwe et al., 2003; Löwe et al., 2004),

however all of these studies have been conducted in primary care settings using either German or Arabic translations of the SCID and the PHQ. This is problematic for at least two reasons. First, findings collected among primary care patients are not readily generalizable to community mental health populations, where psychiatric comorbidity and chronicity is the rule rather than the exception. Studies have shown that screening instruments like the PHQ can be less accurate in samples with high psychiatric comorbidity (e.g., Leon et al., 1999), and while anxiety and depressive disorders tend to be comorbid in non-clinical samples as well (Kessler, Chiu, Demler, & Walters, 2005), their rate of comorbidity is even higher in clinical samples (Shear et al., 2000), which could erode the accuracy of the PHQ in community mental health settings. Further, the prevalence of mental disorders invariably influences the accuracy of any brief psychiatric screen, since the ability to detect a person suffering from a mental disorder is directly related the the number of people in the population who have the disorder (Kraemer, 1992). If there are few cases of the disorder in the sample being screened, then the screen will be searching for a needle in a haystack, and thus have a high rate of false positives. Conversely, the more cases that have the disorder, the more false negatives the test will accrue, because, if the sensitivity of the test remains the same, the proportion of those misdiagnosed as healthy who are actually sick will increase linearly as the prevalence increases.

The second limitation of previous research evaluating the diagnostic efficacy of the PHQ stems from its multilingual application. The psychometric properties of a given instrument can change when translated to other languages, since lexical translations are not one-to-one. For example, some English words do not necessarily have a meaningful equivalent in German. While researchers are usually concerned with the degradation of the psychometric properties of a test that coincides with its translation, when both the test and its criterion have been translated, there is no reason to assume that the properties of a test could not improve. What can be assumed is that these properties may change, for better or worse. As such, there is a need to examine the criterion validity of the original English version of PHQ among community mental health patients, in order to determine its viability as a screen for anxiety and depression in community settings.

This study suggests that the PHQ may not adequately identify anxiety and depressive disorders. Fifty percent of our sample had at least one anxiety disorder diagnosis, nearly half of whom were classified by the PHQ as healthy. A third of the cases of MDD were also missed by the PHQ, and, of the seven individuals in our sample who had a depressive disorder other than MDD, the PHQ missed all but one. This study is limited by its modest sample size, nonetheless, these findings strongly suggest that further research validating the PHQ should be conducted if it is to be considered for use as a diagnostic, or even a screening, tool for anxiety and depressive disorders in community mental health settings.

We acknowledge that the PHQ is not the only short screening instrument for detecting anxiety and depression. Indeed, many other screens for these disorders do exist, such as the CES-D, and are widely employed in both research and clinical settings. Unfortunately, our findings concerning the diagnostic limitations of the PHQ (i.e., high false-negative rates) are not unique, but add to a growing body of literature indicating that psychiatric screening measures are less accurate than structured interviews (see Coyne, 1994 for a review). What is unique about the PHQ, is that it was developed by the same individuals who were largely responsible for developing the SCID. In fact, unlike other screening instruments which explicitly state that they yield only indicators of symptom severity or distress, not psychiatric diagnoses (e.g., the Beck Depression Inventory [Beck, Steer, & Brown, 1996]), the PHQ was a specific attempt to develop a short, self-report screen aimed at identifying psychiatric diagnoses (Spitzer, Kroenke, & Williams, 1999). As such, one might expect the PHQ to have superior concordance with the SCID than simple screens of symptom severity.

Unfortunately, these findings suggest that this is not the case, and provide a particular challenge to social workers in community settings needing to find efficient, yet valid methods of standardizing their diagnostic assessments.

We can see few immediate solutions to this problem. The old adage applied so widely in business, "Cost, efficiency, quality … pick two", appears to apply here. Unfortunately, the most reliable and trusted methods of psychiatric diagnosis, such as the SCID, are highly lacking in both cost and efficiency. However, as this research indicates, the less costly and more efficient methods of diagnostic assessment, such as the PHQ, may be inaccurate. Where is the balance? Within psychiatric assessment, this is the question that needs to be answered, and where social work, with its unique knowledge of and commitment to community-based treatments, could make a substantial contribution. Good diagnostic assessments need to be constructed not only with quality in mind, but also with cost and efficiency. There has been an tendency for diagnostic tools to be developed by researchers who have sometimes had little knowledge of the constraints that exist in community settings. These issues of the external validity of university-tested methods have plagued the dissemination of evidence-based treatments to community settings, and our findings suggest that they affect the dissemination of evidence-based assessment methods as well.

Social workers, being one of the primary providers of services in community mental health settings, have a unique perspective that both acknowledges the need for quality and places high value on applicability. A number of social workers have, with some success, even used this perspective to collaborate with community agencies to assist in the dissemination of evidence-based treatments (e.g., Herie & Martin, 2002). We suggest that evidence-based assessments be developed and disseminated in the same fashion, where quality is emphasized as well as applicability. It is not as if diagnostic assessments do not occur in community settings, rather specific time is protected and alloted for them. The challenge for social workers attempting to improve diagnostic accuracy in these settings, is how to make the most of the time that is already being spent. We suggest that the answer lies in the blend between worlds that have for too long remained separate: research and practice. Future research needs to merge these worlds to combine the measurement expertise of social work researchers with the practical expertise of community clinicians, so that accurate *and* applicable methods can be developed to improve the assessment of individuals in the community and not just the laboratory.

Acknowledgments

References

Basco MR, Bostic JQ, Davies D, Rush AJ, Witte B, Hendrickse W, Barnett V. Methods to improve diagnostic accuracy in a community mental health setting. American Journal of Psychiatry. 2000; 157:1599–1605. [PubMed: 11007713]

Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: Psychometric properties. Journal of Consulting and Clinical Psychology. 1988; 56:893–897. [PubMed: 3204199]

Beck, AT.; Steer, RA.; Brown, GK. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation; 1996.

Becker S, Zaid KA, Faris EA. Screening for somatization and depression in Saudi Arabia: A validation study of the PHQ in primary care. International Journal of Psychiatry in Medicine. 2002; 32:271–283. [PubMed: 12489702]

Coyne JC. Self-reported distress: Analog or ersatz depression? Psychological Bulletin. 1994; 116:29–45. [PubMed: 8078972]

Endicott J, Spitzer RL. The value of the standardized interview for the evaluation of psychopathology. Journal of Personality Assessment. 1972; 36:410–417. [PubMed: 4561481]

First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured Clinical Interview for DSM-IV Axis I Disorders, Research Version, Non-Patient Edition. New York: Biometrics Research Department, New York State Psychiatric Institute; 1996.

Herie M, Martin GW. Knowledge diffusion in social work: A new approach to bridging the gap. Social Work. 2002; 47:85–95. [PubMed: 11829248]

Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Archives of General Psychiatry. 2005; 62:617–627. [PubMed: 15939839]

Kobak KA, Taylor LvH, Dottl SL, Greist JH, Jefferson JW, Burroughs D, Katzelnick DJ, Mandell M. Computerized screening for psychiatric disorders in an outpatient community mental health clinic. Psychiatric Services. 1997; 48:1048–1057. [PubMed: 9255838]

Kraemer, HC. Evaluating medical tests: Objective and quantitative guidelines. Newbury Park, CA: Sage Publications; 1992.

Kranzler HR, Kadden RM, Burleson JA, Babor TF. Validity of psychiatric diagnoses in patients with substance use disorders: Is the interview more important than the interviewer. Comprehensive Psychiatry. 1995; 36:278–288. [PubMed: 7554872]

Leon AC, Portera L, Olfson M, Kathol R, Farber L, Lowell KN, Sheehan DV. Diagnostic errors of primary care screens for depression and panic disorder. International Journal of Psychiatry in Medicine. 1999; 29:1–11. [PubMed: 10376229]

Löwe B, Grafe K, Zipfel S, Spitzer RL, Herrmann-Lingen C, Witte S, Herzog W. Detecting panic disorder in medical and psychosomatic outpatients: Comparative validation of the Hospital Anxiety and Depression Scale, the Patient Health Questionnaire, a screening question, and physicians' diagnosis. Journal of Psychosomatic Research. 2003; 55:515–519. [PubMed: 14642981]

Löwe B, Spitzer RL, Grafe K, Kroenke K, Quenter A, Zipfel S, Buchholz C, Witte S, Herzog W. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. Journal of Affective Disorders. 2004; 78:131–140. [PubMed: 14706723]

Matarazzo JD. The reliability of psychiatric and psychological diagnosis. Clinical Psychology Review. 1983; 3:103–145.

Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. Applied Psychological Measurement. 1977; 1:385–401.

Riskind JH, Beck AT, Berchick RJ, Brown G. Reliability of DSM-III diagnoses for major depression and generalized anxiety disorder using the structured clinical interview for DSM-III. Archives of General Psychiatry. 1987; 44:817–820. [PubMed: 3632255]

Segal DL, Hersen M, van Hasselt VB. Reliability of the Structured Clinical Interview for DSM-III-R: An evaluative review. Comprehensive Psychiatry. 1994; 35:316–327. [PubMed: 7956189]

Shear MK, Greeno C, Kang J, Ludewig D, Frank E, Swartz HA, Hanekamp M. Diagnosis of nonpsychotic patients in community clinics. American Journal of Psychiatry. 2000; 157:581–587. [PubMed: 10739417]

Skre I, Onstad S, Torgersen S, Kringlen E. High interrater reliability for the Structured Clinical Interview for DSM-III-R Axis I (SCID-I). Acta Psychiatrica Scandinavica. 1991; 84:167–173. [PubMed: 1950612]

Spitzer RL, Forman JB, Nee J. DSM-III field trials: I. Initial interrater diagnostic reliability. American Journal of Psychiatry. 1979; 136:815–817. [PubMed: 443467]

Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: The PHQ Primary Care Study. Journal of the American Medical Association. 1999; 282(18):1737–1744. [PubMed: 10568646]

Spitzer RL, Williams JBW, Kroenke K, Linzer M, deGury FV, Hahn SR, Brody D, Johnson JG. Utility of a new procedure for diagnosing mental disorders in primary care: The PRIME-MD 1000 study. Journal of the American Medical Association. 1994; 272(22):1749–1756. [PubMed: 7966923]

Substance Abuse and Mental Health Services Administration. Mental health, United States: 2000. Washington, DC: Author; 2001.

Swartz HA, Shear MK, Wren FJ, Greeno CG, Sales E, Sullivan BK, Ludewig DP. Depression and anxiety among mothers who bring their children to a pediatric mental health clinic. Psychiatric Services. 2005; 56:1077–1083. [PubMed: 16148320]

Zanarini MC, Frankenburg FR. Attainment and maintenance of reliability of Axis I and II disorders over the course of a longitudinal study. Comprehensive Psychiatry. 2001; 42:369–374. [PubMed: 11559863]

Zanarini MC, Skodol AE, Bender D, Dolan R, Sanislow C, Schaefer E, Morey LC, Grilo CM, Shea MT, Mcglashan TH, Gunderson JG. The Collaborative Longitudinal Personality Disorders Study: Reliability of Axis I and II diagnoses. Journal of Personality Disorders. 2000; 14:291–299. [PubMed: 11213787]

Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatrica Scandinavica. 1983; 67:361–370. [PubMed: 6880820]

Table 1

Diagnostic Sensitivity and Specificity of the Patient Health Questionnaire Anxiety Disorders Module. N=50.

| Diagnosis | n^a | Anxiety Disordered (SCID Positive) | | | Healthy (SCID Negative) | | |
		True Positive (PHQ Positive)	False Negative (PHQ Negative)	n	True Negative (PHQ Negative)	False Positive (PHQ Positive)
Any Anxiety	25	14 (56%)	11 (44%)	25	21 (84%)	4 (16%)
Panic	10	6 (60%)	4 (40%)	40	38 (95%)	2 (5%)
Other Anxiety	24	10 (42%)	14 (58%)	26	22 (84%)	4 (16%)

Note. SCID = Structured Clinical Interview for DSM-IV, PHQ = Patient Health Questionnaire.

[a] Prevalence of panic and other anxiety disorder sum to more than 25, because some participants had multiple anxiety disorders.

Table 2

Diagnostic Sensitivity and Specificity of the Patient Health Questionnaire Depressive Disorders Module. N=50.

Diagnosis	n^a	Depressed (SCID Positive)		Healthy (SCID Negative)		
		True Positive (PHQ Positive)	False Negative (PHQ Negative)	n	True Negative (PHQ Negative)	False Positive (PHQ Positive)
Any Depression	17	11 (65%)	6 (35%)	32	22 (69%)	10 (31%)
Major Depression	14	9 (64%)	5 (36%)	36	27 (75%)	9 (25%)
Other Depression	7	1 (14%)	6 (86%)	41	39 (95%)	2 (5%)

Note. SCID = Structured Clinical Interview for DSM-IV, PHQ = Patient Health Questionnaire.

a Prevalence of major and other depression sum to more than 17, because some participants had multiple depressive disorders.