Published in final edited form as:

Psychosom Med. 2016 ; 78(6): 716–727. doi:10.1097/PSY.0000000000000322.

The Patient Health Questionnaire Anxiety and Depression Scale (PHQ-ADS): Initial Validation in Three Clinical Trials

Kurt Kroenke, MD^{a,b,c,*}, Jingwei Wu, PhD^d, Zhangsheng Yu, PhD^d, Matthew J. Bair, MD^{a,b,c}, Jacob Kean, PhD^{a,c,e}, Timothy Stump, MS^d, and Patrick O. Monahan, PhD^d

^aVA HSR&D Center for Health Information and Communication, Roudebush VA Medical Center, Indianapolis, IN

^bDepartment of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

^cRegenstrief Institute, Inc., Indianapolis, IN

^dDepartment of Biostatistics, Indiana University, Indianapolis, IN

^eDepartment of Physical Medicine and Rehabilitation, Indiana University School of Medicine, Indianapolis, IN

Abstract

Objective—We examine the reliability and validity of the Patient Health Questionnaire Anxiety-Depression Scale (PHQ-ADS) – which combines the PHQ-9 and GAD-7 scales – as a composite measure of depression and anxiety.

Methods—Baseline data from 896 patients enrolled in 2 primary-care based trials of chronic pain and 1 oncology-practice based trial of depression and pain were analyzed. The internal reliability, standard error of measurement (SEM), and convergent, construct, and factor structure validity, as well as sensitivity to change of the PHQ-ADS were examined.

Results—The PHQ-ADS demonstrated high internal reliability (Cronbach's alpha of 0.8 to 0.9) in all 3 trials. PHQ-ADS scores can range from 0 to 48 (with higher scores indicating more severe depression/anxiety), and the estimated SEM was approximately 3 to 4 points. The PHQ-ADS showed strong convergent (most correlations 0.7-0.8 range) and construct (most correlations 0.4-0.6 range) validity when examining its association with other mental health, quality of life and disability measures. PHQ-ADS cutpoints of 10, 20, and 30 indicated mild, moderate, and severe levels of depression/anxiety, respectively. Bi-factor analysis showed sufficient unidimensionality of the PHQ-ADS score. PHQ-ADS change scores at 3 months differentiated ($P < .0001$) between individuals classified as worse, stable, or improved by a reference measure, providing preliminary evidence for sensitivity to change.

Conclusions—The PHQ-ADS may be a reliable and valid composite measure of depression and anxiety which, if validated in other populations, could be useful as a single measure for jointly assessing two of the most common psychological conditions in clinical practice and research.

*Corresponding author: Kurt Kroenke, MD, Regenstrief Institute, 1101 West Tenth St, 2nd floor, Indianapolis, IN 46202. Ph 317-630-7447 FAX 317-630-8776. kkroenke@regenstrief.org.

Conflicts of Interest: None of the authors have any conflicts of interest to declare.

Trial Registration—clinicaltrials.gov Identifier: NCT00926588 (SCOPE); NCT00386243 (ESCAPE); NCT00313573 (INCPAD);

Keywords

depression; anxiety; scale; psychometrics

Introduction

Depression and anxiety are the two most common mental health conditions in the general population as well as in clinical practice.¹ Depression and anxiety also result in substantial disability, representing the 2nd and 5th leading causes of years lived with disability in the United States and accounting for enormous losses in work productivity as well as high direct and indirect health care costs.²

There are a number of well-validated measures that assess depression and anxiety as separate domains. However, a measure that provides a single composite score for depression and anxiety also has several potential advantages. First, depression and anxiety frequently co-occur.³⁻⁵ Indeed, the *Diagnostic and Statistical Manual for Mental Disorders, 5th Edition* (DSM 5) acknowledges this comorbidity by including a specifier “with anxious distress” to for depressive disorders accompanied by significant levels of anxiety. Thus, a single score that summarizes the collective effect of depression and anxiety may be useful. Second, some interventions (e.g., cognitive-behavioral therapy; certain classes of antidepressants) are effective for both depression and anxiety. Consequently, selecting a composite score as the primary outcome for interventional studies targeting both depression and anxiety would allow for a smaller sample size than using depression and anxiety as separate co-primary outcomes. As a corollary, a single score that captures both depression and anxiety severity may be attractive to practitioners who are monitoring response to treatment of patients with comorbid depression and anxiety in clinical practice. Third, theoretical and empiric evidence supports an overarching psychological construct that encompasses distinct but related dimensions of depression and anxiety.⁶ Fourth, the moderately strong intercorrelation between depression and anxiety makes a composite score attractive as a covariate in multivariate modeling and other types of adjusted analyses.

The Patient Health Questionnaire 9-item depression scale (PHQ-9) and 7-item Generalized Anxiety Disorder scale (GAD-7) are among the best validated and most commonly used depression and anxiety measures, respectively.⁷ They have been used in hundreds of research studies, incorporated into numerous clinical practice guidelines, and adopted by a variety of medical and mental health care practice settings. Importantly, the PHQ-9 and GAD-7 are public domain measures available in more than 80 translations, many of which can be freely downloaded at www.phqscreeners.com. This paper uses data from 3 clinical trials to examine the reliability and convergent, construct, and factor structure validity as well as sensitivity to change of the Patient Health Questionnaire Anxiety-Depression Scale (PHQ-ADS) – a 16-item scale comprising the PHQ-9 and GAD-7 – as a composite measure of depression and anxiety.

Methods

Patient Sample

Data was drawn from 3 clinical trials enrolling a total of 896 patients (Table 1). Two trials enrolled primary care patients with chronic musculoskeletal pain, and one trial enrolled oncology patients who had depression and/or cancer-related pain. The Stepped Care to Optimize Pain care Effectiveness (SCOPE) trial enrolled 250 patients with chronic musculoskeletal pain from 5 primary care clinics in a single Veterans Affairs (VA) Medical Center, randomizing participants to a telecare collaborative management intervention arm optimizing analgesic therapy (n = 124) or a usual care arm (n = 126). The Evaluation of Stepped Care for Chronic Pain (ESCAPE) trial enrolled 241 Operation Enduring Freedom/Operation Iraqi Freedom veterans, randomizing them to an intervention (n = 120) or usual care (n = 121) group. The intervention involved 12 weeks of optimized analgesic therapy coupled with pain self-management strategies (Step 1) followed by 12 weeks of brief cognitive behavioral therapy (Step 2). The Indiana Cancer Pain and Depression (INCPAD) trial enrolled 405 patients with depression and/or cancer-related pain from 16 community-based oncology practices, randomizing them to a telecare intervention arm optimizing analgesic and antidepressant therapy (n = 202) or a usual care arm (n = 203).¹ Data collection occurred from March 2006 through August 2009 in INCPAD, from December 2007 through April 2012 in ESCAPE, and from June 2010 through May 2013 in SCOPE,

Measures

PHQ-9 and GAD-7—The PHQ-9 consists of 9 items representing the criterion symptoms for DSM 5 major depressive disorder. Respondents are asked how much each symptom has bothered them over the past 2 weeks, with response options of “not at all”, “several days”, “more than half the days”, and “nearly every day”, scored as 0, 1, 2, and 3, respectively. The PHQ-9 can be scored as either a continuous variable from 0 to 27 (with higher scores representing more severe depression) or categorically using a diagnostic algorithm for major depressive or other depressive disorder. The GAD-7 has 7 items with response options identical to the PHQ-9 and therefore can be scored as a continuous variable from 0 to 21 (with higher scores representing more severe anxiety). Although originally developed as a measure to detect generalized anxiety disorder, the operating characteristics of the GAD-7 are nearly as good for the other common anxiety disorders in clinical practice – panic disorder, social anxiety disorder, and posttraumatic stress disorder. The PHQ-9 and GAD-7 have strong internal and test-retest reliability as well as construct and factor-structure validity. Moreover, both measures have proven sensitive to change when monitoring treatment response.²⁻⁴ The PHQ-ADS is the sum of the PHQ-9 and GAD-7 scores and thus can range from 0 to 48, with higher scores indicating higher levels of depression and anxiety symptomatology.

Other Mental Health Measures for Assessing Convergent Validity—The 5-item *Mental Health Inventory (MHI-5)* is one of eight scales that constitute the widely-used 36-item Medical Outcomes Study Short Form health survey (SF-36). Scores on the MHI-5 range from 0 to 100, with lower scores representing worse mental health. The MHI-5 has been found to have reasonable sensitivity and specificity in screening for DSM-IV

depressive and anxiety disorders. The *Mental Component Summary (MCS)* score of the SF-12 was administered, which serves as a measure of impairment related to mental disorders; the MCS is scored from 0 to 100 with higher scores representing better mental functioning and is one of the most widely-used measures of mental health functioning and quality of life. Finally, participants in the SCOPE trial completed the 4-item depression and 4-item anxiety scales from the PROMIS-29 profile; scores for each scale range from 4 to 20 with higher scores representing worse symptoms (www.nihpromis.org). A composite PROMIS anxiety-depression score was also calculated (i.e., the sum of the depression and anxiety scores), which could range from 8 to 40.

Quality of Life and Disability Measures for Assessing Construct Validity—

Two quality of life domains that have shown moderate associations with depression and anxiety are vitality and social functioning which were assessed with the *SF-36 vitality and social functioning* scales; these, like other SF-36 scales, have scores that range from 0 to 100, with lower scores representing worse impairment. *Disability days* were assessed in two trials (SCOPE and INCPAD) with a single item that asked participants to indicate the number of days during the preceding 4 weeks that they were either in bed or had to reduce work or usual activities by 50% or more due to physical health or emotional problems. Another measure of disability used in the INCPAD trial was the *Sheehan Disability Scale (SDS)* which consists of three items asking how much the participant's health condition has interfered with his/her family life, social life, and work over the past month on a scale of 0 (*not at all*) to 10 (*unable to carry on any activities*). The SDS score is a mean of these three items with higher scores reflecting greater disability. In the SCOPE and ESCAPE trials, *work effectiveness* was assessed with a single item asking how effective the respondent was on his or her job during the past 4 weeks on a scale of 0% (not at all effective) to 100% (completely effective).

Statistical Analysis

Because of substantial differences in the patient samples and study interventions, we analyzed data for each trial separately rather than pooling the data. For a number of analyses, results are reported for both the PHQ-ADS as well as its component scales, the PHQ-9 and GAD-7. The mean, standard deviation, and internal reliability (Cronbach's alpha) was calculated for each of the 3 scales. The standard error of measurement (SEM) was calculated as the standard deviation of the baseline score for a measure multiplied by the square root of one minus the Cronbach's alpha. The SEM can be regarded as the standard deviation of an individual score, and either 1 or 2 SEMs have been considered one approach to estimating the minimal clinically important difference (MCID) for a scale. Pearson's correlation coefficients of the PHQ-ADS, PHQ-9 and GAD-7 with other mental health measures and quality of life/disability measures were calculated to assess convergent and construct validity, respectively.

Cutpoints of 10, 20, and 30 on the PHQ-ADS were examined as thresholds of mild, moderate, and severe depression/anxiety symptoms, respectively. This resulted in 4 ordinal PHQ-ADS categories of 0-9, 10-19, 20-29, and 30-48, representing, minimal, mild, moderate, and severe levels of depressive-anxiety symptomatology. The rationale for these

cutpoints was three-fold: 1) Because 5, 10, and 15 represent mild, moderate, and severe cutpoints on the PHQ-9 and GAD-7, it seemed logical to select 10, 20, and 30 on a composite scale that is the simple sum of the two scales; 2) Examination of the frequency distribution of the PHQ-ADS scores in the 3 trials suggested a reasonable distribution of scores using these predefined cutpoints; 3) 10, 20, and 30 are easy-to-remember cutpoints, a pragmatic consideration that may increase clinical uptake. The convergent and construct validity of PHQ-ADS ordinal categories were evaluated by comparing the four groups on mental health and quality of life/disability measures using analysis of variance models.

The structural validity of a single summed PHQ-ADS score was evaluated using confirmatory one-factor, two-factor, and bi-factor models. The one-factor models represent the set of items as being explained by a strictly unidimensional single trait and indicate the measurement validity of a single score when the model fits the data. Bi-factor models represent the set of items as a sufficiently unidimensional trait – one which has some construct-relevant multidimensionality that does not interfere with the interpretation of a single general trait score. Sufficient unidimensionality is indicated when analyses demonstrate that the preponderance of the variance is attributable to the general trait despite the presence of secondary relationships between clusters of items.

Strict unidimensional model fit was evaluated using absolute (i.e., chi square), parsimony-adjusted RMSEA (i.e., root mean square error of approximation; cutoff .06) and WRMR (i.e., weighted root mean square residual; cutoff 1.0), and incremental CFA fit indices (i.e., comparative fit index; cutoff .95). *Sufficient* unidimensionality in the bi-factor model was evidenced by: explained common variance (ECV) .60, omega hierarchical index .70, and a high correlation (e.g. $r > .90$) between the factor loadings of the unidimensional model and the general factor of the bi-factor model. All factor analyses were performed by modeling the items as ordinal categorical with the non-linear logistic link function between items and factors. This non-linear factor analytic model is identical, within a transformation, to an item response theory (IRT) model. We performed factor analysis instead of IRT modeling because our focus was more on dimensionality assessment than item characteristics.

Sensitivity of the PHQ-ADS scores to change was assessed. Specifically, because the MHI-5 and PHQ-ADS were both administered at baseline and at 3 months in two of the trials (SCOPE and INCPAD), and because the MHI-5 is essentially a composite depression-anxiety score (consisting of 3 depression and 2 anxiety items), three MHI-5 change groups (worse, same, improved) were computed for each patient by determining whether the MHI-5 declined or improved by more than 1.0 standard error of measurement (SEM) from baseline to follow-up at 3 months. The SEM for the MHI-5 was 8 in SCOPE and 9 in INCPAD, so we classified those with an MHI-5 decrease or increase of 10 or greater as *worse* or *improved*, respectively, with the remainder of patients classified as *same*. Sensitivity to change of the PHQ-ADS was assessed by computing the standardized response mean (SRM) for each MHI-5 change group, and comparing the SRMs using analysis of variance, with pairwise Tukey-Kramer post hoc tests controlling the overall Type I error rate at 0.05.

Analyses were performed using SAS Version 9.3 (SAS Institute, Cary, North Carolina) and MPlus Version 7.2 (Muthen and Muthen).

Results

Psychometric Characteristics of PHQ-9, GAD-7 and PHQ-ADS in the 3 Trials

As shown in Table 2, the mean PHQ-9 and GAD-7 scores in the 3 trials represent moderate levels of depression and mild levels of anxiety, respectively. The INCPAD trial enrolled patients with depression as well as pain and therefore, not surprisingly, had the highest mean depression scores, whereas the SCOPE trial had the lowest depression and anxiety scores. All 3 scale scores demonstrated good internal reliability, with Cronbach's alphas in the 0.8 to 0.9 range. PHQ-ADS item means (SD) and item-total correlations are summarized in Table S1, Supplemental Digital Content 1; all item-total correlations were good (0.42 to 0.69). Correlations of the 16 PHQ-ADS items with one another are shown in Table S2, Supplemental Digital Content 1.

Using a 1-SEM change to estimate a minimal clinically important difference (MCID), the MCID estimated from these 3 trials would be approximately 2 to 3 points for the PHQ-9 and GAD-7 and 3 to 4 points for the PHQ-ADS. Using a more conservative estimate of a 2-SEM change, the MCID would be approximately 4 to 6 points for the PHQ-9 and GAD-7 and 6 to 8 points for the PHQ-ADS. The distribution of the PHQ-ADS ordinal categories indicated more than a third (38.4%) of patients in the SCOPE trial had minimal depression/anxiety symptoms, approximately a third had mild symptoms (31.2%), and close to a third (30.4%) had moderate to severe symptoms. In the ESCAPE trial, about a quarter (22%-28%) of patients fell into each of the 4 categories, whereas in the INCPAD trial which targeted depressed patients, the majority of patients had some level of depression/anxiety symptoms.

The most commonly used cutpoint on both the PHQ-9 and GAD-7 to screen for depressive and anxiety disorders, respectively, is 10 or greater. The number of patients in the 3 trials that achieved this cutpoint on both the PHQ-9 and GAD-7 was 286 (31.9%); on the PHQ-9 only, 266 (29.7%); on the GAD-7 only, 21 (2.3%); and on neither measure, 323 (36.1%). Thus, if only the PHQ-9 had been used in these trials, 307 (34.3%) of patients with chronic pain who had anxiety only or, more commonly, combined anxiety and depression, would not have been detected. This supports joint use of the PHQ-9 and GAD-7 to increase the detection of comorbid anxiety

Convergent and Construct Validity of the PHQ-ADS, PHQ-9 and GAD-7

As shown in Table 3, the PHQ-ADS had the strongest correlations with the PHQ-9 and GAD-7 (its two component scales), and the PHQ-9 and GAD-7 had moderately strong correlations with one another. The 3 scales also showed moderately strong convergent validity with the 3 composite psychological measures (PROMIS-ADS, MHI-5, and MCS) with the PHQ-ADS having slightly higher correlations than the PHQ-9 and GAD-7. As expected, the highest correlations were with the two scales measuring exclusively depression and anxiety symptoms (PROMIS-ADS and MHI-5). Construct validity was supported by moderate correlations of each of the 3 scales with quality of life and disability measures.

Convergent and Construct Validity of the PHQ-ADS Ordinal Categories

Data in Table 4 demonstrate the convergent and construct validity of the PHQ-ADS ordinal categories. There is a large incremental increase in depression (PHQ-9), anxiety (GAD-7), and psychological composite (PROMIS-ADS, MHI-5, and MCS) scores as one goes from minimal to mild to moderate to severe levels of depression/anxiety as classified by the four PHQ-ADS ordinal categories. A similar incremental “dose-response” effect is seen on all quality of life and disability domains.

Structural Validity of the PHQ-ADS

Table 5 includes the fit statistics for the 1-factor, 2-factor, and bi-factor models. Although the chi-square test was significant in all 3 trials (suggesting some deviation from good fit), this fit index yields high power in larger samples to detect minor deviations. Therefore, consistent with tradition in confirmatory latent variable modeling, we will emphasize the fit indices (CFI, RMSEA, WRMR) which are less dependent on sample sizes. There was generally a small improvement in fit when comparing the 2-factor to 1-factor model, and a greater improvement when comparing the bi-factor to either the 1-factor or 2-factor models. The CFI threshold of .95 was achieved for all 3 models in the SCOPE and ESCAPE trials but only for the bifactor model in the INCPAD trial. The RMSEA threshold of .06 was achieved for the bifactor model in two of the trials but in none of the trials for the 1-factor and 2-factor models. Finally, the WRMR threshold of 1.0 was achieved for the bi-factor model in all 3 trials, the 2-factor model in only 1 trial, and the 1-factor model in none of the trials. As shown in Table S3, Supplemental Digital Content 1, most of the factor loadings were substantially higher than the acceptable threshold of 0.40, and were only slightly higher for the 2-factor compared to the 1-factor model. Moreover, the general factor loadings from the bi-factor model were generally in the range of loadings from the 1-factor model.

In the bi-factor model (Table 5), the general factor strength indices (i.e., ECV, omega hierarchical) and the correlation between factor loadings of the unidimensional model and the general factor of the bi-factor model each exceeded cutoffs (0.60, 0.70, and 0.90, respectively), further suggesting sufficient unidimensionality and supporting the structural validity of a single PHQ-ADS composite score. Finally, the scree plots (Figure S1, Supplemental Digital Content 1) of the eigenvalues indicated that there was one dominant factor, because the eigenvalues dropped greatly from the first to the second factor, after which eigenvalues leveled off with much smaller drops between the second and remaining factors. Taken together, the fit indices and the factor loadings point to the validity of the traditional scoring of the PHQ-9 and GAD-7 as depression and anxiety scale scores as well as the sufficient unidimensionality of scoring the PHQ-ADS as a composite score.

Sensitivity to Change of the PHQ-ADS

According to the MHI-5 change scores at 3 months, there were 56 patients in the SCOPE trial who were classified as worse, 113 as unchanged, and 75 as improved. The mean PHQ-ADS score increased 3.63 points in the worse group, declined 3.12 points in the stable group, and declined 7.96 points in the improved group, resulting in SRMs of -0.45, -0.51, and -0.98, respectively. In the INCPAD trial, there were 73 patients classified as worse, 115 as unchanged, and 147 as improved. The mean PHQ-ADS score decreased in all 3 groups

(--5.10 points in the worse group, --9.72 points in the unchanged group, and --16.40 in the improved group, resulting in SRMs of --0.57, --1.28, and -1.89, respectively. The PHQ-ADS change scores among categories were significantly different ($P < .0001$) by analysis of variance, and pairwise comparisons between the worse, unchanged, and improved categories also differed ($p < .001$) in both trials. Thus, although the direction of PHQ-ADS change for the worse group in the INCPAD trial was unexpected, the PHQ-ADS change scores significantly differentiated between the worse, unchanged, and improved groups in both trials.

Discussion

In this validation study of the PHQ-ADS, several important findings emerge. First, the PHQ-ADS demonstrated good internal reliability as well as strong convergent and construct validity in 3 separate trials. Second, cutpoints of 10, 20, and 30 on the PHQ-ADS indicate mild, moderate, and severe levels of depression/anxiety symptoms, respectively. Third, factor analysis confirmed sufficient unidimensionality of the PHQ-ADS to support its use as a composite depression/anxiety measure. Fourth, there is preliminary evidence for sensitivity to change of the PHQ-ADS in that it significantly differed between groups that were categorized as worse, unchanged, or improved at 3 months post-randomization.

The PHQ-ADS cutpoints of 10, 20, and 30 are easy for clinicians to remember and, interestingly, are double the cutpoints of the individual PHQ-9 and GAD-7 scales for which scores of 5, 10, and 15 represent thresholds for mild, moderate, and severe depressive and anxiety symptoms, respectively. Since the PHQ-9 and GAD-7 ordinal cutpoints have proven useful in patient care as well as in practice guidelines for stratifying treatment decisions, future investigations should examine the utility of ordinal severity categories for the PHQ-ADS. The statistically-determined SEM suggests that a 3 to 4 point change on the PHQ-ADS may represent a clinically important difference. Also, the comparison of PHQ-ADS change scores among worse, stable, and improved groups as defined by the MHI-5 suggest the PHQ-ADS is sensitive to change over time. However, it will also be important to assess responsiveness in treatment trials that jointly target depression and anxiety to further examine what amount of change in PHQ-ADS scores is clinically meaningful.

The high comorbidity of depression and anxiety is one reason a composite measure may be useful. A WHO study involving the administration of a structured psychiatric interview to 5438 primary care patients from 15 international primary sites found that 39% of patients with current depression also had an anxiety disorder, and 44% with a current anxiety disorder also had comorbid depression.⁷ A U.S. study of 2091 patients from 15 primary care clinics found that 30% of patients with depression and/or anxiety (defined as PHQ-8 and GAD-7 scores ≥ 15 , respectively) had both conditions. A Dutch psychiatric cohort study of 1783 patients found that of those with a DSM-IV depressive disorder, 67% had a current and 75% had a lifetime comorbid anxiety disorder, and of persons with a current anxiety disorder, 63% had a current and 81% had a lifetime depressive disorder. Similarly, numerous other studies have confirmed 30-50% or higher co-occurrence rates of depression and anxiety.⁸⁻¹¹

The number of composite depression-anxiety scales is limited. One well-validated composite measure is the 14-item Hospital Depression and Anxiety Scale (HADS) which provides both a single composite score as well as separate depression and anxiety scores.² Notably, a systematic review of studies examining the latent structure of the HADS tend to support both an overarching unidimensional structure as well as two underlying factors, which can vary with both the sample and the analytic strategies used. Another measure is the Mental Health Inventory (the 5-item mental health scale of the SF-36) which provides a composite score³ as well as depression (3 items) and anxiety (2 items) subscores; however, the latter are calculated differently than the composite score and have only occasionally been used in research.^{4,5} and seldom in clinical practice. Both the HADS and MHI are proprietary measures and thus require a user fee to the practice or researcher for their administration. In contrast, the PHQ-9 and GAD-7 are public domain measures. Another set of public domain measures developed with NIH funding are the PROMIS scales, which include depression and anxiety scales of varying lengths (4 to 8 items) as well as computer-adapted testing (CAT) administration that draws upon larger item banks. One study demonstrated good correspondence between PROMIS depression and anxiety scores and PHQ-9 and GAD-7 scores. Also, the PHQ-ADS was strongly associated with scores on the PROMIS Anxiety-Depression composite score (Tables 3 and 4). Thus, future research could compare the PHQ-ADS and PROMIS composite anxiety-depression scores in terms of validity and responsiveness.

Our study has several limitations. First, all 3 trials focused on patients with pain, rather than individuals with depression (except INCPAD) or anxiety. However, previous studies have supported the utility of the PHQ-9⁶ and GAD-7⁷ in individuals with pain, and one would expect similar performance from a composite score of the two measures. Also, there was a substantial number of patients who met clinical cutpoints for depression and combined depression/anxiety in the 3 trials, but only a small proportion with anxiety only. Thus, the PHQ-ADS should be further evaluated in populations without pain as well as those with a more representative distribution of anxiety and depression, including patient samples where a structured diagnostic interview is used rather than cutpoints on a scale. Moreover, it is important to evaluate the PHQ-ADS in patients seen in mental health settings where the types and severity of psychiatric disorders may vary substantially compared to medical populations. For example, although the PHQ-9 has proven useful in psychiatric patients using similar cutpoints as those used in medical settings⁸, the operating characteristics may be somewhat different in psychiatric populations (i.e., similar specificity but lower sensitivity). Second, patient samples in two of the trials were exclusively Veterans and predominantly men; thus, data on the PHQ-ADS in non-Veteran samples including more women is warranted. Third, we did not test responsiveness to treatment of the PHQ-ADS since none of the 3 trials were specifically treating anxiety and only one was targeting depression. Thus, evaluating responsiveness to treatment (e.g., intervention groups versus control group) of the PHQ-ADS in interventional studies targeting depression and anxiety (ideally in the same trial) is needed. Fourth, the results in the INCPAD trial of oncology patients were, though generally comparable to the two primary care trials, weaker on a few of the psychometric analyses. This suggests that further study of the PHQ-ADS in patients with cancer as well as other specialty populations is warranted. Fifth, we did not use a

structured criterion standard diagnostic interview in these 3 trials to determine which patients met criteria for depressive or anxiety disorders, and thus were unable to compare the sensitivity and specificity of the PHQ-ADS with the PHQ-9 and GAD-7. Certainly, a PHQ-ADS screening cutpoint would be higher than that of the PHQ-9 or GAD-7 (which are 10) since its score range is greater; for example, 10 represents a cutpoint for moderate depressive symptoms on the PHQ-9 and moderate anxiety symptoms on the GAD-7, whereas 20 represented a cutpoint for moderate depressive/anxiety symptoms on the PHQ-ADS in our sample. However, the PHQ-ADS is not intended to replace its constituent subscales in screening for depressive and anxiety disorders, since the operating characteristics of the PHQ-9 and GAD-7 are already well-established.⁷ Sixth, our assessment of construct validity relied on relatively brief PROMIS and SF mental health measures; future studies should compare the PHQ-ADS to more detailed depressive and anxiety scales, both in terms of construct validity as well as responsiveness to treatment.

The PHQ-ADS composite score does not override the value of the individual PHQ-9 depression and GAD-7 anxiety scores but instead complements them as a measure of overall psychological symptomatology when the latter is manifested principally by varying levels of depressive and anxiety symptoms. Our findings in terms of reliability and convergent, construct, and structural validity (both fit indices and factorial loadings) support the established value of the PHQ-9 and GAD-7 as measures of depression and anxiety, respectively, while at the same time demonstrating sufficient unidimensionality of the PHQ-ADS as a composite measure. There are conceptual and clinical reasons in support of distinct depression and anxiety scores as well as a single summative score. Despite their comorbidity, depression and anxiety represent different groups of disorders in psychiatric classification; and while responding to several common treatments, depression and anxiety also have some specific treatments that differ. The PHQ-ADS score may be useful in studies for which a single depression/anxiety score is desirable as either an outcome variable or as a covariate to adjust for in multivariable analyses. The PHQ-ADS may also be useful in monitoring the concomitant treatment of depression and anxiety, especially since some treatments work across both conditions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Funding: This work was supported by a Department of Veterans Affairs Health Services Research and Development Merit Review award (IIR 07-119) and National Cancer Institute R01 award (R01 CA115369) to Dr. Kroenke); a Department of Veterans Affairs Rehabilitation Research and Development Merit Review award (IIR F44371) to Dr. Bair; a VA Career Development Award to Dr. Kean (CDA IK2RX000879), and a National Institute of Arthritis and Musculoskeletal Disorders R01 award to Dr. Monahan (R01 AR064081). The sponsor had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

References

1. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, Angermeyer MC, Bernert S, de GG, Morosini P, Polidori G, Kikkawa T, Kawakami N, Ono Y, Takeshima T, Uda H, Karam EG, Fayyad JA, Karam AN, Mneimneh ZN, Medina-Mora ME, Borges G, Lara C, de GR, Ormel J, Gureje O, Shen Y, Huang Y, Zhang M, Alonso J, Haro JM, Vilagut G, Bromet EJ, Gluzman S, Webb C, Kessler RC, Merikangas KR, Anthony JC, Von Korff MR, Wang PS, Brugha TS, Guilar-Gaxiola S, Lee S, Heeringa S, Pennell BE, Zaslavsky AM, Ustun TB, Chatterji S. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA*. 2004; 291(21):2581–90. [PubMed: 15173149]
2. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshelman S, Wittchen H, Kendler KS. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. *Arch Gen Psychiatry*. 1994; 51(1):8–19. [PubMed: 8279933]
3. Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV III, Hahn SR, Brody D, Johnson JG. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study *JAMA*. 1994; 272(22):1749–56. [PubMed: 7966923]
4. Ormel J, Vonkorff M, Ustun TB, Pini S, Korten A, Oldehinkel T. Common mental disorders and disability across cultures. Results from the WHO Collaborative Study on Psychological Problems in General Health Care. *JAMA*. 1994; 272(22):1741–8. [PubMed: 7966922]
5. Spitzer RL, Kroenke K, Williams JBW, the Patient Health Questionnaire Study Group. Validity and utility of a self-report version of PRIME-MD: The PHQ Primary Care Study. *JAMA*. 1999; 282(18):1737–44. [PubMed: 10568646]
6. Strine TW, Mokdad AH, Balluz LS, Gonzalez O, Crider R, Berry JT, Kroenke K. Depression and anxiety in the United States: findings from the 2006 Behavioral Risk Factor Surveillance System. *Psychiatr Serv*. 2008; 59(12):1383–90. [PubMed: 19033164]
7. US Burden of Disease Collaborators. The state of US health, 1990–2010: burden of diseases, injuries, and risk factors. *JAMA*. 2013; 310(6):591–608. [PubMed: 23842577]
8. Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. *JAMA*. 2003; 289(23):3135–44. [PubMed: 12813119]
9. Greenberg PE, Sisitsky T, Kessler RC, Finkelstein SN, Berndt ER, Davidson JR, Ballenger JC, Fyer AJ. The economic burden of anxiety disorders in the 1990s. *J Clin Psychiatry*. 1999; 60(7):427–35. [PubMed: 10453795]
10. Kessler RC, Keller MB, Wittchen HU. The epidemiology of generalized anxiety disorder. *Psychiatr Clin North Am*. 2001; 24(1):19–39. [PubMed: 11225507]
11. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, Rush AJ, Walters EE, Wang PS. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*. 2003; 289(23):3095–105. [PubMed: 12813115]
12. Lowe B, Spitzer RL, Williams JB, Mussell M, Schellberg D, Kroenke K. Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. *Gen Hosp Psychiatry*. 2008; 30(3):191–9. [PubMed: 18433651]
13. Kroenke K, Spitzer RL, Williams JBW, Lowe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics*. 2009; 50:613–21. [PubMed: 19996233]
14. Rodriguez BF, Weisberg RB, Pagano ME, Machan JT, Culpepper L, Keller MB. Frequency and patterns of psychiatric comorbidity in a sample of primary care patients with anxiety disorders. *Compr Psychiatry*. 2004; 45(2):129–37. [PubMed: 14999664]
15. Hanel G, Henningsen P, Herzog W, Sauer N, Schafert R, Szecsenyi J, Lowe B. Depression, anxiety, and somatoform disorders: Vague or distinct categories in primary care? Results from a large cross-sectional study *J Psychosom Res*. 2009; 67:189–97. [PubMed: 19686874]
16. McLaughlin TP, Khandker RK, Kruzikas DT, Tummala R. Overlap of anxiety and depression in a managed care population: Prevalence and association with resource utilization. *J Clin Psychiatry*. 2006; 67(8):1187–93. [PubMed: 16965195]
17. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)*. Washington, DC: American Psychiatric Pub; 2013.

18. Clark LA, Watson D. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *J Abnorm Psychol.* 1991; 100(3):316–36. [PubMed: 1918611]
19. 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(4): 645–54. [PubMed: 7822565]
20. Kroenke K, Spitzer RL, Williams JB, Lowe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *Gen Hosp Psychiatry.* 2010; 32(4):345–59. [PubMed: 20633738]
21. Wittkamp K, van Ravesteijn H, Bass K, van de Hoogen H, Schene A, Bindels P, Lucassen P, van de Lisdonk E, van Weert H. The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. *Gen Hosp Psychiatry.* 2009; 31:451–9. [PubMed: 19703639]
22. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med.* 2007; 22:1596–602. [PubMed: 17874169]
23. Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Lowe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med.* 2007; 146(5):317–25. [PubMed: 17339617]
24. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ.* 2012; 184(3):E191–E196. [PubMed: 22184363]
25. Herr NR, Williams JW Jr, Benjamin S, McDuffie J. Does this patient have generalized anxiety or panic disorder?: The Rational Clinical Examination systematic review. *JAMA.* 2014; 312(1):78–84. [PubMed: 25058220]
26. Kroenke K, Krebs E, Wu J, Bair MJ, Damush T, Chumbler N, York T, Weitlauf S, McCalley S, Evans E, Barnd J, Yu Z. Stepped Care to Optimize Pain Care Effectiveness (SCOPE) Trial: study design and sample characteristics. *Contemp Clin Trials.* 2013; 34:270–81. [PubMed: 23228858]
27. Kroenke K, Krebs EE, Wu J, Yu Z, Chumbler NR, Bair MJ. Telecare collaborative management of chronic pain in primary care: a randomized clinical trial. *JAMA.* 2014; 312(3):240–8. [PubMed: 25027139]
28. Bair MJ, Ang D, Wu J, Outcalt SD, Sargent C, Kempf C, Froman A, Schmid AA, Damush TM, Yu Z, Davis LW, Kroenke K. Evaluation of Stepped Care for Chronic Pain (ESCAPE) in Veterans of the Iraq and Afghanistan Conflicts: A Randomized Clinical Trial. *JAMA Intern Med.* 2015; 175(5):682–689. [PubMed: 25751701]
29. Kroenke K, Theobald D, Norton K, Sanders R, Schlundt S, McCalley S, Harvey P, Iseminger K, Morrison G, Carpenter JS, Stubbs D, Jacks R, Carney-Doebbeling C, Wu J, Tu W. Indiana Cancer Pain and Depression (INCPAD) Trial: design of a telecare management intervention for cancer-related symptoms and baseline characteristics of enrolled participants. *Gen Hosp Psychiatry.* 2009; 31(3):240–53. [PubMed: 19410103]
30. Kroenke K, Theobald D, Wu J, Norton K, Morrison G, Carpenter J, Tu W. Effect of telecare management on pain and depression in patients with cancer: a randomized trial. *JAMA.* 2010; 304(2):163–71. [PubMed: 20628129]
31. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med.* 2001; 16:606–13. [PubMed: 11556941]
32. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* 2006; 166(10):1092–7. [PubMed: 16717171]
33. Lowe B, Unutzer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. *Med Care.* 2004; 42(12):1194–201. [PubMed: 15550799]
34. Lowe B, Kroenke K, Herzog W, Grafe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *Journal of Affective Disorders.* 2004; 81(1):61–6. [PubMed: 15183601]

35. Clark DM, Layard R, Smithies R, Richards DA, Suckling R, Wright B. Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behav Res Ther.* 2009; 47:910–20. [PubMed: 19647230]
36. Dear BF, Titov N, Sunderland M, McMillan D, Anderson T, Lorian C, Robinson E. Psychometric comparison of the generalized anxiety disorder scale-7 and the Penn State Worry Questionnaire for measuring response during treatment of generalised anxiety disorder. *Cogn Behav Ther.* 2011; 40(3):216–27. [PubMed: 21770844]
37. McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs *Med Care.* 1993; 31:247–63.
38. Berwick DM, Murphy JM, Goldman PA, Ware JE Jr, Barsky AJ, Weinstein MC. Performance of a five-item mental health screening test. *Med Care.* 1991; 29(2):169–76. [PubMed: 1994148]
39. Rumpf HJ, Meyer C, Hapke U, John U. Screening for mental health: validity of the MHI-5 using DSM-IV Axis I psychiatric disorders as gold standard. *Psychiatry Res.* 2001; 105(3):243–53. [PubMed: 11814543]
40. Ware JE, Gandek B. The SF-36 Health Survey: development and use in mental health research and the IQOLA Project. *Int J Ment Health.* 1994; 23:49–73.
41. Choi SW, Reise SP, Pilkonis PA, Hays RD, Cella D. Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. *Qual Life Res.* 2010; 19(1):125–36. [PubMed: 19941077]
42. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. *Assessment.* 2011; 18:263–83. [PubMed: 21697139]
43. Kroenke K, Yu Z, Wu J, Kean J, Monahan PO. Operating characteristics of PROMIS four-item depression and anxiety scales in primary care patients with chronic pain. *Pain Med.* 2014; 15(11):1892–901. [PubMed: 25138978]
44. Wang H-L, Kroenke K, Wu J, Tu W, Theobald D, Rawl SM. Cancer-related pain and disability: a longitudinal study. *J Pain Symptom Manage.* 2011; 42:813–21. [PubMed: 21570808]
45. Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. *Int Clin Psychopharmacol.* 1996; 11(Suppl 3):89–95. [PubMed: 8923116]
46. Krebs EE, Bair MJ, Wu J, Damush TM, Tu W, Kroenke K. Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. *Med Care.* 2010; 48:1007–14. [PubMed: 20856144]
47. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol.* 1999; 52(9):861–73. [PubMed: 10529027]
48. Kroenke K, Spitzer RL, Williams JBW, Lowe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *General Hospital Psychiatry.* 2010; 32(4):345–59. [PubMed: 20633738]
49. Babyak MA, Green SB. Confirmatory factor analysis: an introduction for psychosomatic medicine researchers. *Psychosom Med.* 2010; 72:587–597. [PubMed: 20467001]
50. Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *J Pers Assess.* 2010; 92:544–559. [PubMed: 20954056]
51. Takane Y, De Leeuw J. On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika.* 1987; 52:393–408.
52. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation *Control Clin Trials.* 1991; 12(4 Suppl):142S–58S. [PubMed: 1663851]
53. Monahan PO, Boustani MA, Alder C, Galvin JE, Perkins AJ, Healey P, Chehresa A, Shepard P, Bupp C, Frame A, Callahan C. Practical clinical tool to monitor dementia symptoms: the HABC-Monitor. *Clin Interv Aging.* 2012; 7:143–57. [PubMed: 22791987]

54. Sartorius N, Ustun TB, Lecrubier Y, Wittchen HU. Depression comorbid with anxiety: results from the WHO study on psychological disorders in primary health care. *Br J Psychiatry*. 1996; (1)(30): 38–43. [PubMed: 8770426]
55. Goldberg, DP.; Lecrubier, Y. Form and frequency of mental disorders across cultures. In: Ustun, TB.; Sartorius, N., editors. *Mental Illness in General Health Care*. Chichester, United Kingdom: John Wiley & Sons; p. 1995p. 323–34.
56. Lamers F, van OP, Comijs HC, Smit JH, Spinhoven P, van Balkom AJ, Nolen WA, Zitman FG, Beekman AT, Penninx BW. Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). *J Clin Psychiatry*. 2011; 72(3):341–8. [PubMed: 21294994]
57. Murphy JM, Horton NJ, Laird NM, Monson RR, Sobol AM, Leighton AH. Anxiety and depression: a 40-year perspective on relationships regarding prevalence, distribution, and comorbidity. *Acta Psychiatr Scand*. 2004; 109(5):355–75. [PubMed: 15049772]
58. 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(2):69–77. [PubMed: 11832252]
59. Cosco TD, Doyle F, Ward M, McGee H. Latent structure of the Hospital Anxiety And Depression Scale: a 10-year systematic review. *J Psychosom Res*. 2012; 72(3):180–4. [PubMed: 22325696]
60. Vodermaier A, Millman RD. Accuracy of the Hospital Anxiety and Depression Scale as a screening tool in cancer patients: a systematic review and meta-analysis. *Support Care Cancer*. 2011; 19(12):1899–908. [PubMed: 21898134]
61. Yamazaki S, Fukuhara S, Green J. Usefulness of five-item and three-item Mental Health Inventories to screen for depressive symptoms in the general population of Japan. *Health Qual Life Outcomes*. 2005; 3:48. [PubMed: 16083512]
62. Cuijpers P, Smits N, Donker T, ten Have M, de Graaf R. Screening for mood and anxiety disorders with the five-item, the three-item, and the two-item Mental Health Inventory. *Psychiatry Res*. 2009; 168(3):250–5. [PubMed: 19185354]
63. Johns SA, Kroenke K, Krebs EE, Theobald DE, Wu JW, Tu WZ. Longitudinal comparison of three depression measures in adult cancer patients. *J Pain Symptom Management*. 2013; 45(1):71–82.
64. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Amtmann D, Bode R, Buysse D, Choi S, Cook K, DeVellis R, DeWalt D, Fries JF, Gershon R, Hahn EA, Lai JS, Pilkonis P, Revicki D, Rose M, Weinfurt K, Hays R. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*. 2010; 63(11):1179–94. [PubMed: 20685078]
65. Arnow BA, Hunkeler EM, Blasey CM, Lee J, Constantino MJ, Fireman B, Kraemer HC, Dea R, Robinson R, Hayward C. Comorbid depression, chronic pain, and disability in primary care. *Psychosom Med*. 2006; 68(2):262–8. [PubMed: 16554392]
66. Osborne TL, Turner AP, Williams RM, Bowen JD, Hatzakis M, Rodriguez A, Haselkorn JK. Correlates of pain interference in multiple sclerosis. *Rehab Psychology*. 2006; 51(2):166–74.
67. Hauser W, Biewer W, Gesmann M, Kuhn-Becker H, Petzke F, von Wilmoswky H, Langhorst J, Glaesmer H. A comparison of the clinical features of fibromyalgia syndrome in different settings. *Eur J Pain*. 2011; 15(9):936–41. [PubMed: 21652242]
68. Koroschetz J, Rehm SE, Gockel U, Brosz M, Freynhagen R, Tolle TR, Baron R. Fibromyalgia and neuropathic pain - differences and similarities. A comparison of 3057 patients with diabetic painful neuropathy and fibromyalgia. *BMC Neurology*. 2011; 11
69. Forchheimer MB, Richards JS, Chiodo AE, Bryce TN, Dyson-Hudson TA. Cut point determination in the measurement of pain and its relationship to psychosocial and functional measures after traumatic spinal cord injury: a retrospective model spinal cord injury system analysis. *Arch Phys Med Rehab*. 2011; 92(3):419–24.
70. Choi Y, Mayer TG, Williams MJ, Gatchel RJ. What is the best screening test for depression in chronic spinal pain patients? *Spine J*. 2014; 14(7):1175–82. [PubMed: 24225008]
71. Bair MJ, Poleshuck EL, Wu J, Krebs EE, Damush TM, Tu W, Kroenke K. Anxiety but not social stressors predict 12-month depression and pain outcomes. *Clin J Pain*. 2013; 29(2):95–101. [PubMed: 23183264]

72. Duffy FF, Chung H, Trivedi M, Rae DS, Regier DA, Katzelnick DJ. Systematic use of patient-rated depression severity monitoring: is it helpful and feasible in clinical psychiatry? *Psychiatr Serv.* 2008; 59:1148–1154. [PubMed: 18832500]
73. Katzelnick DJ, Duffy FF, Chung H, Regier DA, Rae DS, Trivedi MH. Depression outcomes in psychiatric clinical practice: using a self-rated measure of depression severity. *Psychiatric Services.* 2011; 62:929–935. [PubMed: 21807833]
74. Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta- analysis. *Gen Hosp Psychiatry.* 2015; 37:567–576. [PubMed: 26195347]

Abbreviations

PHQ-9	9-item Patient Health Questionnaire depression scale
GAD-7	7-item Generalized Anxiety Disorder anxiety scale
PHQ-ADS	Patient Health Questionnaire Anxiety-Depression Scale
SCOPE	Stepped Care to Optimized Pain care Effectiveness trial
ESCAPE	Evaluation of Stepped Care for Chronic Pain trial
INCPAD	Indiana Cancer Pain and Depression trial
MHI-5	5-item Mental Health Inventory
SF-36	36-item Short Form Health Survey
SF-12	12-item Short Form Health Survey
MCS	Mental Component Summary
PCS	Physical Component Summary
PROMIS	Patient Reported Outcomes Measurement Information System
SDS	Sheehan Disability Scale
SEM	standard error of measurement
MCID	minimal clinically important difference
CFA	comparative fit index
ECV	explained common variance

Table 1
Characteristics of Patient Samples in the Three Trials

Variable	SCOPE (n = 250)	ESCAPE (n = 241)	INCPAD (n = 405)
Clinical sites	Primary care	Primary care	Oncology
Primary eligibility condition	Chronic musculo-skeletal pain	Chronic musculo-skeletal pain	Pain and/or Depression
Veterans, %	100.0	100.0%	7.7%
Age, mean (range) yr.	55.1 (28-65)	36.7 (21-73)	58.8 (23-86)
Men, %	82.8	88.4	32.1
Race, %			
White	76.8	77.7	79.5
Black	19.2	12.8	18.0
Other	4.0	9.5	2.5
Education, %			
Some college	74.0	75.9	39.0
High school or less	26.0	24.1	61.0
Major depression, %	24.0	32.0	69.9

Table 2
Selected Characteristics of PHQ-9, GAD-7, and PHQ-ADS in Three Trials

Variable	SCOPE (n = 250)	ESCAPE (n = 241)	INCPAD (n = 405)
Scale scores, mean (SD)			
PHQ-9,	9.1 (6.3)	11.2 (5.9)	13.0 (6.7)
GAD-7	5.9 (5.6)	8.8 (5.3)	7.9 (5.8)
PHQ-ADS	14.9 (11.2)	20.0 (10.4)	20.8 (11.0)
Cronbach's alpha			
PHQ-9	0.842	0.846	0.816
GAD-7	0.882	0.853	0.855
PHQ-ADS	0.917	0.908	0.878
Standard error of measurement			
PHQ-9	2.51	2.29	2.91
GAD-7	1.97	2.04	2.94
PHQ-ADS	3.18	3.13	3.81
PHQ-ADS Categories, n %			
Minimal (0-9)	96 (38.4)	53 (22.0)	65 (16.1)
Mild (10-19)	78 (31.2)	66 (27.4)	122 (30.1)
Moderate (20-29)	42 (16.8)	68 (28.2)	122 (30.1)
Severe (30-39)	34 (13.6)	54 (22.4)	96 (23.7)

Table 3
Correlations of PHQ-ADS, PHQ-9, and GAD-7 with Mental Health (Convergent Validity)
and Quality of Life and Disability (Construct Validity) Measures *

Variable	PHQ-ADS	PHQ-9	GAD-7
Convergent Validity			
PHQ-9			
SCOPE	.95	--	--
ESCAPE	.94	--	--
INCPAD	.89	--	--
GAD-7			
SCOPE	.94	.77	--
ESCAPE	.93	.75	--
INCPAD	.86	.54	--
PROMIS-ADS			
SCOPE	.83	.76	.80
SF Mental (MHI-5)			
SCOPE	.83	.78	.78
ESCAPE	.81	.79	.72
INCPAD	.76	.65	.69
SF MCS			
SCOPE	.79	.75	.74
ESCAPE	.82	.81	.73
INCPAD	.67	.60	.57
Construct Validity			
SF Vitality			
SCOPE	.69	.63	.50
ESCAPE	.57	.60	.45
INCPAD	.46	.45	.36
SF Social			
SCOPE	.62	.60	.57
ESCAPE	.66	.65	.58
Disability Days			
SCOPE	.48	.46	.44
INCPAD	.35	.31	.30
Sheehan Disability Scale			
INCPAD	.45	.41	.38
Work Effectiveness			
SCOPE	-.46	-.47	-.39
ESCAPE	-.41	-.34	-.43

* Values shown are Pearson's correlation coefficients

Table 4
Convergent and Construct Validity of PHQ-ADS Ordinal Categories

Measure	PHQ-ADS Category (Score Range)				P-value*	
	Minimal (0-9)	Mild (10-19)	Moderate (20-29)	Severe (30-48)		
Convergent Validity						
Mean (SD)						
PHQ-9						
	SCOPE	3.3 (2.3)	8.8 (2.3)	14.2 (3.2)	19.5 (3.7)	< .001
	ESCAPE	3.8 (1.9)	8.6 (2.3)	13.5 (3.0)	18.7 (3.3)	< .001
INCPAD	1.9 (2.5)	11.1 (3.3)	15.5 (3.5)	19.8 (3.6)	< .001	
GAD-7						
	SCOPE	1.3 (1.6)	4.9 (2.5)	9.8 (2.9)	16.3 (3.2)	< .001
	ESCAPE	2.4 (1.4)	6.5 (2.2)	10.5 (2.6)	15.8 (2.8)	< .001
INCPAD	1.9 (2.1)	4.1 (2.9)	8.6 (3.6)	15.8 (3.2)	< .001	
PROMIS-ADS						
	SCOPE	9.4 (2.3)	12.0 (4.2)	18.8 (5.8)	26.6 (7.1)	< .001
SF Mental (MHI-5)						
	SCOPE	85.5 (7.7)	75.0 (13.2)	52.5 (16.9)	36.5 (18.2)	< .001
	ESCAPE	81.0 (13.5)	67.3 (13.7)	50.5 (14.4)	34.1 (15.1)	< .001
INCPAD	82.2 (10.8)	64.5 (15.3)	49.7 (15.0)	35.1 (17.0)	< .001	
SF MCS						
	SCOPE	56.8 (5.4)	50.7 (8.3)	39.6 (9.7)	29.4 (9.8)	< .001
	ESCAPE	55.4 (7.2)	46.9 (8.8)	37.0 (7.9)	27.8 (7.3)	< .001
INCPAD	54.0 (8.7)	44.8 (8.7)	36.8 (9.9)	30.5 (10.8)	< .001	
Construct Validity						
Mean (SD)						
SF Vitality						
	SCOPE	56.1 (19.4)	38.3 (16.1)	25.3 (19.5)	21.0 (16.9)	< .001

Measure	PHQ-ADS Category (Score Range)				P-value*
	Minimal (0-9)	Mild (10-19)	Moderate (20-29)	Severe (30-48)	
ESCAPE	50.0 (20.0)	40.8 (16.0)	31.2 (15.2)	21.5 (13.4)	< .001
INCPAD	46.7 (18.7)	30.6 (18.4)	23.3 (16.0)	19.1 (14.7)	< .001
SF Social					
SCOPE	82.8 (19.8)	69.2 (21.6)	48.8 (24.8)	37.9 (23.9)	< .001
ESCAPE	75.9 (21.1)	61.0 (21.0)	46.0 (22.1)	29.9 (17.1)	< .001
Disability Days					
SCOPE	4.7 (6.5)	9.1 (8.2)	13.0 (7.4)	16.2 (8.6)	< .001
INCPAD	10.4 (9.9)	15.3 (10.4)	18.9 (9.8)	20.5 (8.3)	< .001
Sheehan Disability Scale					
INCPAD	3.3 (2.6)	4.8 (2.7)	6.1 (2.6)	6.9 (2.4)	< .001
Work Effectiveness					
SCOPE	82.0 (18.8)	73.6 (20.5)	61.8 (22.9)	52.9 (26.3)	< .001
ESCAPE	84.0 (17.8)	79.9 (18.2)	74.3 (20.0)	59.1 (24.5)	< .001

* Analysis of variance was used to compare mean scores across the four categories.

Table 5
Confirmatory One-Factor, Two-Factor, and Bi-factor Model Statistics for the PHQ-ADS *

Fit Index	SCOPE Trial (n = 250)			ESCAPE Trial (n = 241)			INCPAD Trial (n = 405)		
	1-factor	2-factor	Bi-factor	1-factor	2-factor	Bi-factor	1-factor	2-factor	Bi-factor
Number of parameters	64	65	80	64	65	80	64	65	80
Chi-square (df)	318.0 (104)	290.6 (103)	250.39 (88)	278.7 (104)	228.0 (103)	167.1 (88)	817.7 (104)	407.1 (103)	161.6 (88)
RMSEA	.091	.085	.086	.083	.071	.061	.130	.085	.045
CFI	0.956	0.962	0.967	0.954	0.967	0.979	0.862	.941	.986
WRMR	1.179	1.110	0.949	1.114	0.981	0.755	2.040	1.384	0.740
Estimated factor correlations	n/a	0.912	[†]	n/a	0.865	[†]	n/a	0.653	[†]
Explained common variance			0.854			0.792			0.634
Omega hierarchical index			0.906			0.891			0.743
Correlation between 1-factor model loadings and general factor loadings from bi-factor model			0.97			0.73			0.79

* *Strict* unidimensional model fit was evaluated using absolute (i.e., chi square), parsimony-adjusted RMSEA (i.e., root mean square error of approximation; cutoff = .06), incremental CFA fit indices (i.e., comparative fit index; cutoff = .95), and WRMR fit indices (i.e., weighted root mean square residual, cutoff = 1.0). *Sufficient* unidimensionality in the bi-factor model was evidenced by: explained common variance greater than 0.60, omega hierarchical index greater than 0.70, and a high correlation (e.g. $r > .90$) between the factor loadings of the unidimensional model and the general factor of the bi-factor model.

[†] Each pair constrained to zero