

PSYCHOLOGY, PSYCHIATRY & BRAIN NEUROSCIENCE SECTION

Original Research Article

Operating Characteristics of PROMIS Four-Item Depression and Anxiety Scales in Primary Care Patients with Chronic Pain

Kurt Kroenke, MD,^{*†‡} Zhangsheng Yu, PhD,[§]
Jingwei Wu, MS,[§] Jacob Kean, PhD,^{*¶} and
Patrick O. Monahan, PhD[§]

*VA HSR&D Center for Health Information and Communication, Roudebush VA Medical Center; Departments of [†]Medicine and [¶]Physical Medicine and Rehabilitation, Indiana University School of Medicine; [‡]Regenstrief Institute, Inc.; [§]Department of Biostatistics, Indiana University, Indianapolis, Indiana, USA

Reprint requests to: Kurt Kroenke, MD, Regenstrief Institute, Inc., 5th Floor, 1050 Wishard Blvd, Indianapolis, IN 46202, USA. Tel: 317-630-7447; Fax: 317-630-8776; E-mail: kkroenke@regenstrief.org.

Disclosure: Dr. Kroenke has received honoraria for serving on an advisory board for Eli Lilly. None of the authors have any conflicts of interest to disclose.

Trial Registration: clinicaltrials.gov Identifier: NCT00926588

Funding: This work was supported by Department of Veterans Affairs Health Services Research and Development Merit Review award to Dr. Kroenke (IIR 07–119), VA Career Development Award to Dr. Kean (CDA IK2RX000879), and 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.

Abstract

Objective. Depression and anxiety are prevalent in patients with chronic pain and adversely affect pain,

quality of life, and treatment response. The purpose of this psychometric study was to determine the reliability and validity of the four-item Patient Reported Outcomes Measurement Information System (PROMIS) depression and anxiety scales in patients with chronic pain.

Design. Secondary analysis of data from the Stepped Care to Optimize Pain care Effectiveness study, a randomized clinical trial of optimized analgesic therapy.

Setting. Five primary care clinics at the Roudebush VA Medical Center (RVAMC) in Indianapolis, Indiana.

Subjects. Two hundred forty-four primary care patients with chronic musculoskeletal pain.

Methods. All patients completed the four-item depression and anxiety scales from the PROMIS 29-item profile, as well as several other validated psychological measures. The minimally important difference (MID) using the standard error of measurement (SEM) was calculated for each scale, and convergent validity was assessed by interscale correlations at baseline and 3 months. Operating characteristics of the PROMIS measures for detecting patients who had probable major depression or were anxiety-disorder screen-positive were calculated.

Results. The PROMIS scales had good internal reliability, and the MID (as represented by two SEMs) was 2 points for the depression scale and 2.5 points for the anxiety scale. Convergent validity was supported by strong interscale correlations. The optimal screening cutpoint on the 4- to 20-point PROMIS scales appeared to be 8 for both the depression and anxiety scales.

Conclusions. The PROMIS four-item depression and anxiety scales are reasonable options as ultra-brief measures for screening in patients with chronic pain.

Key Words. Pain; Depression; Anxiety; Screening; Psychometrics; PROMIS

Introduction

Depression and anxiety are the two most common mental disorders, cause substantial impairment, and often remain undetected and undertreated [1,2]. Moreover, they frequently co-occur, creating additive morbidity and adverse effects on treatment response of each other as well as other coexisting diseases [3–5]. Chronic pain is a condition where depression and anxiety are particularly prevalent and have profound negative influences on quality of life, disability, and response of pain to therapy [6–9].

Therefore, efforts to measure both depression and anxiety are warranted to optimize treatment in the clinical setting as well as to better understand the respective roles of depression and anxiety as mediators, moderators, and outcomes in clinical research. As depression and anxiety are often only two of many patient-reported outcomes being evaluated, brief measures are desirable, particularly when screening is the initial goal in practice or when multiple secondary outcomes are being assessed in research. Scales as short as two to four items have proven useful as ultra-brief measures for depression and anxiety [10,11]. Individuals with elevated scores in the clinical setting can be further evaluated with more comprehensive measures and/or diagnostic interviews.

The Patient Reported Outcomes Measurement Information System (PROMIS) measures have had extensive development and population validation by the National Institutes of Health, which is encouraging their use across multiple studies to facilitate intra- and inter-disease comparisons [12,13]. One advantage of PROMIS measures is computer adaptive testing (CAT) wherein items are drawn from a large data bank and administered in a tailored fashion to a particular individual based upon his or her initial responses. However, static PROMIS scales with a fixed number of items are also available, perform almost as well as CAT-administered scales [14], and are more feasible to use in many clinical and research settings. There are PROMIS profiles that have four-, six-, and eight-item scales to assess each of seven cross-cutting domains: depression, anxiety, pain, fatigue, sleep, physical functioning, and social role satisfaction. The four-item scales could serve as ultra-brief measures in many settings.

In this article, we use data from the Stepped Care to Optimize Pain care Effectiveness (SCOPE) study to determine the reliability and validity of the four-item PROMIS depression and anxiety measures in a primary care population with persistent pain. Specifically, we report findings regarding internal reliability, minimally important difference (MID), convergent validity, and operating characteristics including optimal cutpoints for screening purposes.

Methods

Patient Sample

Recruitment and a detailed description of the patients enrolled in SCOPE have been previously described [15]. All patients had to have musculoskeletal pain that was at least moderate in intensity (Brief Pain Inventory[BPI] severity score ≥ 5 of either the patient's average or worst pain in the past week) and had persisted at least 3 months despite trying at least one analgesic medication.

Briefly, the 250 study patients had a mean age of 55.1 years (range 28–65); 83% were men; 77% were white, 19% black, and 4% other race. The mean baseline BPI total pain score was 5.2, representing a moderate level of pain. The duration of pain was 3–12 months in 2.0% of patients, 1–5 years in 26.4%, 6–10 years in 19.2%, and more than 10 years in 52.4%. Only 7.6% patients reported a single site of pain, whereas two to four sites were reported by 49.6% and five or more sites by 42.8%. In this article, we used the 244 patients who completed both baseline and 3-month assessments, as both time points were used for some of our analyses.

Depression, Anxiety, and General Mental Health Measures

The four-item depression and four-item anxiety scales are part of the PROMIS-29 profile; scores for each scale range from 4 to 20 with higher scores representing worse symptoms (<http://www.nihpromis.org>) [14,16]. The Patient Health Questionnaire (PHQ)-9 and PHQ-2 depression scales as well as the generalized anxiety disorder (GAD)-7 and GAD-2 anxiety scales were also administered [17]. 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. Both the PHQ-9 and PHQ-2 (scored from 0 to 6 and consisting of the depressed mood and anhedonia items of the PHQ-9) have excellent reliability, construct and criterion validity, and sensitivity to change. Likewise, the GAD-7 (scored 0–21) and GAD-2 (scored 0–6 and consisting of the nervous/anxious and inability to control worry items of the GAD-7) have excellent reliability and validity for assessing generalized anxiety disorder as well as anxiety disorders in general. The five-item Mental Health Inventory (MHI-5) is one of eight scales that constitute the widely used 36-item Short Form health survey. 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 fourth edition of *Diagnostic and Statistical Manual for Mental Disorders* (DSM-IV) depressive and anxiety disorders [18]. The MHI-5 consists of three depression and two anxiety items, for which each item has a raw score ranging from 1 to 5 [19,20]. Thus, scores on the depression subscale (MHI-d) range from 3 to 15, and scores on the anxiety subscale (MHI-a) range from 2 to 10, with higher scores representing worse depression or anxiety. Finally, 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 [21].

Depression and Anxiety Categories

Patients were classified as having *probable major depression* according to the PHQ-9 categorical scoring algorithm, which includes the nine diagnostic criteria for DSM-IV major depressive disorder (MDD). To be classified as having probable major depression, an individual must endorse at least five criterion items as being present “more than half the days” in the past 2 weeks (except the ninth item about thoughts of self-harm or life not worth living counts if endorsed at a lower threshold of “several days”); also, one of the criterion items endorsed must be depressed mood and/or anhedonia. This algorithm has been well validated through multiple studies [17,22,23].

Patients were classified as *anxiety-disorder screen-positive* by their responses on validated screening scales for the five most common anxiety disorders seen in clinical practice (excluding simple phobias). These screeners were the GAD-7 [24] for generalized anxiety disorder, a four-item screener [25,26] for posttraumatic stress disorder (PTSD); the five-item PHQ panic disorder scale for panic disorder [27]; the three-item version of the Social Phobia Inventory [28,29] for social anxiety disorder; and a screening question [30] for obsessive-compulsive disorder (OCD). These assessments were administered at baseline only and were used solely to establish criteria for anxiety; with the exception of the GAD-7, they were not used in the comparison analyses. For each anxiety disorder screening scale, the recommended cutpoint was used. Where more than 1 cutpoint was provided in the literature, the cutpoint with the greater specificity (at least 80% or greater) was selected in order to reduce the number of false positive screens and thereby more conservatively classify patients as screen-positive. The cutpoint was ≥ 10 for the GAD-7 [24,31]; ≥ 3 of four items endorsed as present on the PTSD screener [25,26]; ≥ 6 on the Social Phobia Inventory [28,29]; ≥ 3 items endorsed on the PHQ panic scale [27]; and a response of either “most of the time” or “all of the time” to the OCD screening question (*How often over the past 30 days have you been bothered by having the same thoughts over and over or feeling compelled to do the same thing repeatedly?*) [30].

Statistical Analysis

Internal reliability of each of the measures was estimated by Cronbach’s alpha. The standard error of measurement (SEM) for a measure was calculated as the standard deviation of the baseline score for that measure multiplied by the square root of one minus the Cronbach’s alpha [32]. 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 MID for a scale [33,34].

Convergent validity was tested by calculating correlations of the depression measures with one another and with the two more general measures of mental health (MHI-5 and MCS), and the same was done for anxiety measures.

We calculated the operating characteristics of the PROMIS depression measure for detecting individuals with probable major depression and of the PROMIS anxiety measure for detecting individuals who were anxiety-disorder screen-positive. These operating characteristics were determined across a range of cutpoints for the PROMIS measures and included:

1. *Sensitivity*: proportion of patients with the categorical depression or anxiety condition who had a positive score (i.e., a score at or above the chosen PROMIS cutpoint);
2. *Specificity*: proportion of patients without the categorical condition who had a negative score (i.e., a score below the chosen PROMIS cutpoint);
3. *Positive predictive value*: proportion of patients with a positive PROMIS score who had the categorical condition;
4. *Positive likelihood ratio for a cutpoint*: true positive rate divided by false positive rate for all individuals at or above that cutpoint, calculated as sensitivity/ (1 – specificity);
5. *Positive likelihood ratio for a discrete score*: true positive rate divided by false positive rate for all individuals with that exact PROMIS score. The score for which this operating characteristic exceeds 1 is one potential candidate for the optimal cutpoint;
6. *Youden index*: (sensitivity + specificity) minus 1. For this calculation, sensitivity and specificity are expressed as decimals, so that the Youden index for a cutpoint with a sensitivity of 82% and a specificity of 91% would be $(0.82 + 0.91) - 1 = 0.73$. The cutpoint with the highest Youden index is another potential candidate for the optimal cutpoint.

The area under the curve (AUC) for each measure was determined. AUC values are interpreted as the probability that a measure correctly discriminates between patients with and without a condition (in this study, probable major depression for the depression measures, and anxiety-disorder screen-positive for the anxiety measures). The possible range of values is 0.5 (no ability to discriminate) to 1.0 (perfect ability to discriminate). An AUC ≥ 0.70 is often considered moderate discrimination, and an AUC ≥ 0.90 is considered excellent discrimination [35].

Results

As shown in Table 1, all depression and anxiety scales had good internal reliability (Cronbach’s alpha > 0.80) except the two-item scales, for which a lower alpha is expected by virtue of its very limited number of items; alpha is a function of the number of test items and the average intercorrelation among the items, and two-item scales in particular have lower alphas [36]. The SEM for the PROMIS depression and anxiety scales were 1.08 and

PROMIS Four-Item Depression and Anxiety Scales

Table 1 Internal reliability (Cronbach's alpha), standard error of measurement (SEM), and minimally important difference (MID) for depression and anxiety measures

Scale	# Items	Possible Score Range*	Cronbach's Alpha	SEM†	MID‡
Depression					
PROMIS depression	4	4– 20	0.93	1.08	2.2
PHQ-9	9	0– 27	0.83	2.60	5.2
PHQ-2	2	0– 6	0.69	1.04	2.1
MHI-d	3	3– 15	0.84	1.09	2.2
Anxiety					
PROMIS anxiety	4	4– 20	0.89	1.24	2.5
GAD-7	7	0– 21	0.88	1.96	3.9
GAD-2	2	0– 6	0.78	0.87	1.7
MHI-a	2	2– 10	0.63	1.17	2.3
General Mental					
MHI-5	5	0– 100	0.93	7.89	15.8
Mental Component Summary		0– 100	0.87	3.93	7.9

* Bolded number represents the worst possible score on a scale.

† SEM = (standard deviation of baseline score) × (square root of [1 minus Cronbach's alpha])

‡ MID is two SEMs.

1.24, respectively, which means that, using the 2-SEM criterion, an MID would be 2–2.5 points.

The correlations summarized in Table 2 show that all depression measures were strongly correlated with one

another as well with the general mental health measures (MHI-5 and MCS). The same was true for the anxiety measures. Correlations at baseline and 3 months were generally similar. The slightly lower correlations of the PROMIS depression measure with the PHQ scales may

Table 2 Correlations among depression and anxiety scales at baseline and 3 months*

Depression or General Mental Scale	PROMIS Dep	PHQ-9	PHQ-2	MHI-a	MHI-5	MCS
PROMIS Depression	—	0.75	0.68	0.86	– 0.85	– 0.83
	—	0.75	0.73	0.82	– 0.78	– 0.75
PHQ-9		—	0.83	0.78	– 0.78	– 0.76
		—	0.85	0.69	– 0.74	– 0.74
PHQ-2			—	0.74	– 0.72	– 0.70
			—	0.67	– 0.69	– 0.71
MHI-d				—	– 0.96	– 0.88
				—	– 0.95	– 0.87
MHI-5					—	0.89
					—	0.89
Anxiety or General Mental Scale	PROMIS Anx	GAD-7	GAD-2	MHI-a	MHI-5	MCS
PROM Anxiety	—	0.79	0.75	0.71	– 0.80	– 0.78
	—	0.81	0.76	0.73	– 0.79	– 0.69
GAD-7		—	0.89	0.73	– 0.78	– 0.74
		—	0.89	0.75	– 0.78	– 0.67
GAD-2			—	0.71	– 0.74	– 0.68
			—	0.68	– 0.74	– 0.65
MHI-a				—	– 0.91	– 0.78
				—	– 0.91	– 0.78
MHI-5					—	0.89
					—	0.89

* First (top) correlation is at baseline and second (bottom) correlation is at 3 months.

Table 3 Operating characteristics of PROMIS four-item depression and anxiety measures

PROMIS Cutpoint			Sensitivity	Specificity	Youden's Index*	Likelihood Ratio (Cutpoint) [†]	Likelihood Ratio (Discrete) [†]	Positive Predictive Value
Depression								
Raw score	T-score [‡]	Percentile [¶]						
6	51.8	57	93.2	69.1	0.623	3.02	0.33	49.1
7	53.9	65	89.8	79.5	0.693	4.37	1.39	58.2
8	55.7	72	83.1	84.3	0.674	5.29	1.18	62.8
9	57.3	77	78.0	88.6	0.666	6.84	3.76	68.7
10	58.9	81	67.8	91.4	0.592	7.88	5.64	71.4
Anxiety								
Raw score	T-score [‡]	Percentile [¶]						
6	51.2	55	77.0	64.9	0.419	2.19	0.41	65.4
7	53.7	64	71.7	77.9	0.496	3.24	0.84	73.6
8	55.8	73	64.6	86.3	0.509	4.72	1.74	80.2
9	57.7	78	54.0	92.4	0.464	7.11	4.25	85.9
10	59.5	82	44.2	94.7	0.389	8.34	2.61	87.7

* Youden's index = sensitivity + specificity – 1

[†] Likelihood ratio for a *cutpoint* is for all individuals who have that score or greater, whereas likelihood ratio for a *discrete* value is only for those patients who have that exact score.

[‡] T-score distributions are standardized such that a 50 represents the average (mean) for the US general population, and the standard deviation around that mean is 10 points. A *high score always represents more of the concept being measured*. Thus, for example, a person who has a T-score of 60 is one standard deviation higher than the general population for the concept being measured.

[¶] Percentiles are derived from the *PROMIS Instrument-Level Statistics* manual available at the PROMIS Assessment Center at <http://www.nihpromis.org>. Of note, these are the percentiles for the depression and anxiety item banks derived from the general population and vary somewhat depending upon the age, sex, and disease characteristics of a particular sample.

be due to the fact the PHQ-9 captures the somatic as well as mood criteria for major depression, whereas the PROMIS measure, like the MHI-5 and MCS, excludes somatic symptoms.

Of the 244 patients, 59 (24.1%) had probable major depression and 113 (46.3%) were anxiety-disorder screen-positive. Table 3 shows the operating characteristics for the PROMIS scales at cutpoints ranging from 6 to 10; scores below or above these cutpoints had sensitivities or specificities inadequate for screening purposes. The positive likelihood ratio for a discrete PROMIS score first exceeded 1.0 at a score of 7 for the depression scale and 8 for the anxiety scale. Likewise, the Youden's index was greatest for a depression score of 7 and an anxiety score of 8. However, because both measures were above the 70th percentile (72nd and 73rd percentile for depression and anxiety scales, respectively) at a cutpoint of 8, we favor the latter cutpoint for both scales in order to keep the screen-positive rate less than 30%. The lower sensitivity of the anxiety measure may in part be due to the fact that 21 patients who were anxiety-disorder screen-positive unexpectedly had the lowest PROMIS score of 4.

All depression measures had excellent AUCs as summarized in Table 4. The AUC for the PHQ-9 (as a continuous score) is artificially inflated as the PHQ-9 diagnostic algorithm was used as the criterion measure for determining

Table 4 Area under the curve (AUC) for depression and anxiety measures*

Measure	AUC	SE
Depression		
PROMIS four-item depression profile	0.899	0.024
PHQ-9	0.987	0.006
PHQ-2	0.947	0.013
MHI-d	0.900	0.026
MHI-5	0.899	0.028
Anxiety		
PROMIS four-item anxiety profile	0.793	0.029
GAD-7	0.850	0.025
GAD-2	0.844	0.025
MHI-a	0.791	0.028
MHI-5	0.831	0.026

* For depression scales, AUC was for detecting individuals with major depression as diagnosed by the PHQ-9 diagnostic algorithm. For anxiety scales, AUC was for detecting individuals who screened positive for one or more anxiety disorders. Of the 244 patients, 59 (24.1%) had probable major depression and 113 (46.3%) were anxiety-disorder screen-positive. SE = standard error of AUC.

PROMIS Four-Item Depression and Anxiety Scales

the presence or absence of probable major depression. However, the PROMIS and other depression measures all have AUCs ≥ 0.90 (with rounding), which represents an excellent level of discrimination. The anxiety measures had somewhat lower AUCs (0.79–0.85), reflecting again their lower sensitivity in this particular study.

As a further measure of construct validity, the mean PROMIS anxiety score was 5.4 in those who screened negative for all anxiety disorders ($N = 131$) vs 9.3 in those who screened positive for at least one anxiety disorder ($N = 113$). Also, the PROMIS anxiety score increased as the number of anxiety disorders increased, with a mean score of 6.9 ($N = 54$), 8.9 ($N = 26$), 13.3 ($N = 18$), and 13.7 ($N = 16$) in those with one, two, three, and four to five anxiety disorders, respectively. The mean PROMIS depression score was 11.8 and 5.5 in those with ($N = 59$) and without ($N = 185$) probable major depression

Discussion

Our study provides preliminary psychometric data regarding the four-item PROMIS depression and pain scales in primary care patients with chronic pain. First, both scales have good internal reliability as well as strong convergent validity with other legacy measures. Second, an MID as gauged by two SEMs is between 2 and 2.5 points (making a three-point difference a conservative estimate). Third, an optimal screening cutpoint appears to be 8 on both the depression and anxiety measures. Fourth, the AUCs suggest that both scales discriminate reasonably well between patients with and without depression and anxiety, although for several reasons the AUC for anxiety was somewhat lower.

Regarding the three-point MID noted above, clinical response in an individual patient may, by some definitions, require more than this degree of change. For example, clinically important depression improvement is often defined as at least a 50% decrease in depressive symptom severity (although some patients with more severe depression who achieve less than a 50% reduction in depression severity may still experience benefits) [37].

An important limitation of our study is that, instead of a structured psychiatric interview as the criterion standard, we used the PHQ-9 diagnostic algorithm and the results from anxiety screeners to classify patients as probable major depression and anxiety-disorder screen-positive, respectively. Thus, the operating characteristics in Tables 3 and 4 (which are based upon this categorization) should be considered preliminary. In particular, the anxiety-disorder screen-positive category should be interpreted cautiously as it was based upon multiple anxiety screeners and had a relatively high prevalence (46%) in our sample. Also, an unexpected finding of 21 anxiety-disorder screen-positive patients with a PROMIS anxiety score of 4 (the lowest possible score) led to a lowering of the sensitivity. Whether this reflected our broad approach to categorizing anxiety, a floor effect of the PROMIS measure for detecting anxiety in some pain patients, or a

chance finding in our particular study warrants further research. Despite the caveat to not overinterpret our data regarding the operating characteristics of the four-item PROMIS anxiety scale, it is reassuring that this measure did correlate strongly with other legacy anxiety measures and the mean score increased with the number of screen-positive anxiety disorders, both supporting convergent validity.

A second study limitation is that our sample consisted of chronic pain patients who were Veterans and predominantly white men. Thus, replication in non-Veteran populations with a greater proportion of women and minority participants is warranted in order to determine the generalizability of our findings. However, differential item functioning analyses have shown that PROMIS item endorsement and item discrimination values are not substantially influenced by demographic or disease characteristics after controlling for the overall summed item or trait score of the corresponding PROMIS scale [38,39]. It is also possible that PROMIS cutpoints for depression and anxiety may differ in patients without pain, as it has been shown that mean PROMIS *T*-scores are a little higher in patients with pain conditions (1–4 points depending upon the specific condition); however, the modest increase in scores is not any greater for pain disorders than a number of other chronic medical conditions [40].

Measuring depression in patients with chronic pain is complicated by the question of how to handle somatic symptoms (e.g., fatigue and poor sleep) that are core criteria for MDD but also frequently experienced by patients with pain. Indeed, this issue is not unique to chronic pain but also true of MDD associated with many other chronic medical disorders. Determining how much of the fatigue or insomnia reported by a patient with mood symptoms and concurrent chronic pain or heart failure or cancer is difficult and, in fact, it is likely that both the medical and the psychological conditions are contributory. Indeed, a recent study including a literature synthesis rather convincingly demonstrated that a reciprocal rather than unidirectional relationship exists between pain and depression, making it even more difficult to assign “shared symptoms” to one condition or the other [7]. Thus, many experts now favor an “inclusive” approach toward counting symptoms of MDD rather than assigning symptoms to one condition or the other, an approach that has some empiric support [41,42].

We used the PHQ-9 diagnostic algorithm that has been validated against structured psychiatric interviews and so is a reasonable surrogate for a diagnosis of probable MDD [17]. Moreover, even structured psychiatric interviews rely on the same nine criterion symptoms to diagnose MDD and thus are not immune to potential confounding by somatic symptoms. Some argue that a measure like PROMIS, which excludes somatic symptoms, is a “purer” measure of depression. However, MDD is a “syndromic” diagnosis that includes both affective and somatic symptoms, and a large number of depression treatment trials have used the syndromic diagnosis of MDD (based upon

the 9 criterion symptoms) to determine study eligibility as well as treatment success. Thus, one important issue salient to any depression measure is ascertaining its optimal cutpoint(s) for detecting probable MDD. This requires using either a brief (PHQ-9) or longer (Structured Clinical Interview for DSM disorders or other structured diagnostic interview) criterion measure that assesses the nine MDD criterion symptoms. One potential way of capturing the MDD syndrome with PROMIS measures would be to add other PROMIS scales such as fatigue and sleep; however, the exact strategy for doing this as well as the pragmatic implications have not yet been studied.

Three other streams of research support the use of the PHQ-9 in this study. First, the PHQ-9 has proven to be a reliable and valid depression measure in patients with chronic pain [20,43–57]. Second, the PHQ-9 appears to be unidimensional in patients with rheumatologic disorders rather than characterized by separate affective and somatic factors [58]. Third, although a few measures like the Beck Depression Inventory (BDI) have a long track record in chronic pain research [59], recent studies have shown comparability among the PHQ-9, PROMIS, BDI, and other depression measures in a variety of populations, especially if the appropriate “cross-walking” between scores on different scales is established [60–63].

There are a variety of options for using PROMIS scales. CAT draws upon a larger data bank of items for a particular domain (such as depression or anxiety); responses to earlier items lead to computerized selection of subsequent items, thereby tailoring the scale to a specific patient. Although most individuals need to complete only seven to nine items to achieve a reliable score, the use of a larger data bank of items broadens the range of content and difficulty to accommodate the diversity in how a domain such as depression or anxiety is experienced at the level of an individual. However, CAT requires computerized administration (not practical in some settings) and loses some of its advantages when fixed scales shorter than 10 items have relatively comparable psychometric performance [14]. As an alternative to CAT, the PROMIS project has also developed a number of short-forms and profiles ranging from four to nine items that assess a number of common domains. Ultra-brief scales (one to four items) have been found to perform well in screening for depression, anxiety, and other common mental disorders [10,17,64–66]; our study provides additional evidence for the reliability and validity of the PROMIS four-item depression and anxiety scales as ultra-brief measures. In clinical practice and research settings where depression and anxiety may be only two of multiple domains being evaluated and in which time and respondent burden are concerns, brevity may be a critical factor in determining what is and is not assessed. Besides brevity, ease of scoring and being freely available (i.e., in the public domain) further facilitate adoption and wider use [67–69]. All three pragmatic criteria are satisfied by the PROMIS and PHQ/GAD measures assessed in our study.

Assessment of depression and anxiety is important in the clinical management as well as research of chronic pain. The four-item PROMIS depression and anxiety scales provide practical options as ultra-brief measures. Of note, the American Psychiatric Association has included PROMIS depression and anxiety short scales as one option for cross-cutting symptom measures in its field trials for the fifth edition of its *Diagnostic and Statistical Manual for Mental Disorders*, while still acknowledging more validation in clinical samples is needed [70]. Our study in a sample of primary care patients with chronic pain exemplifies such clinical validation of the PROMIS measures as screeners for depression and anxiety.

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(22):1749–56.
- 2 Anseau M, Dierick M, Buntinx F, et al. High prevalence of mental disorders in primary care. *J Affect Disord* 2004;78(1):49–55.
- 3 Kessler RC, Ormel J, Demler O, Stang PE. Comorbid mental disorders account for the role impairment of commonly occurring chronic physical disorders: Results from the National Comorbidity Survey. *J Occup Environ Med* 2003;45(12):1257–66.
- 4 Lowe B, Spitzer RL, Williams JB, et al. Depression, anxiety and somatization in primary care: Syndrome overlap and functional impairment. *Gen Hosp Psychiatry* 2008;30(3):191–9.
- 5 Rapaport MH, Clary C, Fayyad R, Endicott J. Quality-of-life impairment in depressive and anxiety disorders. *Am J Psychiatry* 2005;162(6):1171–8.
- 6 Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: A literature review. *Arch Intern Med* 2003;163(20):2433–45.
- 7 Kroenke K, Wu J, Bair MJ, et al. Reciprocal relationship between pain and depression: A 12-month longitudinal analysis in primary care. *J Pain* 2011;12: 964–73.
- 8 Bair MJ, Wu J, Damush TM, Sutherland JM, Kroenke K. Association of depression and anxiety alone and in combination with chronic musculoskeletal pain in primary care patients. *Psychosom Med* 2008;70(8):890–7.
- 9 Bair MJ, Poleshuck EL, Wu J, et al. Anxiety but not social stressors predict 12-month depression and pain outcomes. *Clin J Pain* 2013;29:95–101.
- 10 Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care?

PROMIS Four-Item Depression and Anxiety Scales

- A pooled analysis and meta-analysis of 22 studies. *Br J Gen Pract* 2007;57(535):144–51.
- 11 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.
 - 12 Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: Item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res* 2007;16(suppl 1):133–41.
 - 13 Cella D, Riley W, Stone A, et al. 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.
 - 14 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.
 - 15 Kroenke K, Krebs E, Wu J, et al. Stepped Care to Optimize Pain Care Effectiveness (SCOPE) trial: Study design and sample characteristics. *Contemp Clin Trials* 2013;34:270–81.
 - 16 Pilkonis PA, Choi SW, Reise SP, et al. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): Depression, anxiety, and anger. *Assessment* 2011;18:263–83.
 - 17 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.
 - 18 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.
 - 19 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.
 - 20 Johns SA, Kroenke K, Krebs EE, et al. Longitudinal comparison of three depression measures in adult cancer patients. *J Pain Symptom Manage* 2013;45(1):71–82.
 - 21 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.
 - 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.
 - 23 Wittkamp K, van Ravesteijn H, Bass K, et al. 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.
 - 24 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.
 - 25 Prins A, Ouimette P, Kimerling R, et al. The primary care PTSD screen (PC-PTSD): Development and operating characteristics. *Int J Psychiatr Clin Pract* 2004;9:9–14.
 - 26 Bliese PD, Wright KM, Adler AB, et al. Validating the primary care posttraumatic stress disorder screen and the posttraumatic stress disorder checklist with soldiers returning from combat. *J Consult Clin Psychol* 2008;76:272–81.
 - 27 Lowe B, Grafe K, Zipfel S, et al. 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. *J Psychosom Res* 2003;55(6):515–9.
 - 28 Connor KM, Kobak KA, Churchill LE, Katzelnick D, Davidson JR. Mini-SPIN: A brief screening assessment for generalized social anxiety disorder. *Depress Anxiety* 2001;14(2):137–40.
 - 29 Seeley-Wait E, Abbott MJ, Rapee RM. Psychometric properties of the mini-social phobia inventory. *Prim Care Companion J Clin Psychiatry* 2009;11:231–6.
 - 30 Houston JP, Kroenke K, Faries DE, et al. PDI-4A: An augmented provisional screening instrument assessing 5 additional common anxiety-related diagnoses in adult primary care patients. *Postgrad Med* 2011;123:89–95.
 - 31 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.
 - 32 Krebs EE, Bair MJ, Wu J, et al. Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. *Med Care* 2010;48:1007–14.
 - 33 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.

Kroenke et al.

- 34 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.
- 35 Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr* 2007;96:644–7.
- 36 Eisinga R, Grotenhuis M, Pelzer B. The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *Int J Public Health* 2013;58(4):637–42.
- 37 Rush AJ, Kraemer HC, Sackeim HA, et al. Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology* 2006;31:1841–53.
- 38 Teresi JA, Ocepek-Welikson K, Kleinman M, et al. Analysis of differential item functioning in the depression item bank from the Patient Reported Outcome Measurement Information System (PROMIS): An item response theory approach. *Psychol Sci Q* 2009;51(2):148–80.
- 39 Cook KF, Bamer AM, Amtmann D, Molton IR, Jensen MP. Six patient-reported outcome measurement information system short form measures have negligible age- or diagnosis-related differential item functioning in individuals with disabilities. *Arch Phys Med Rehabil* 2012;93(7):1289–91.
- 40 PROMIS Cooperative Group. PROMIS Instrument-Level Statistics Including Gender, Educational Level, Age Bracket, Clinical, and Levels of Self-rated General Health Subgroups. 2011. Available at: www.nihpromis.org (accessed June 2014).
- 41 Koenig HG, George LK, Peterson BL, Pieper CF. Depression in medically ill hospitalized older adults: Prevalence, characteristics, and course of symptoms according to six diagnostic schemes. *Am J Psychiatry* 1997;154:1376–83.
- 42 Simon GE, Von Korff M. Medical co-morbidity and validity of DSM-IV depression criteria. *Psychol Med* 2006;36:27–36.
- 43 Poleshuck EL, Bair MJ, Kroenke K, et al. Musculoskeletal pain and measures of depression. *Gen Hosp Psychiatry* 2010;32:114–5.
- 44 Gibbons LE, Feldman BJ, Crane HM, et al. Migrating from a legacy fixed-format measure to CAT administration: Calibrating the PHQ-9 to the PROMIS depression measures. *Qual Life Res* 2011;20(9):1349–57. (Published erratum appears in *Qual Life Res* 2013;22(2):459–60).
- 45 Turner JA, Dworkin SF. Screening for psychosocial risk factors in patients with chronic orofacial pain—Recent advances. *J Am Dent Assoc* 2004;135(8):1119–25.
- 46 Williams LS, Jones WJ, Shen J, Robinson RL, Kroenke K. Outcomes of newly referred neurology outpatients with depression and pain. *Neurology* 2004;63(4):674–7.
- 47 Rosemann T, Korner T, Wensing M, et al. Rationale, design and conduct of a comprehensive evaluation of a primary care based intervention to improve the quality of life of osteoarthritis patients. The PraxArt-project: A cluster randomized controlled trial. *BMC Public Health* 2005;5:77.
- 48 Arnow BA, Blasey CM, Lee J, et al. Relationships among depression, chronic pain, chronic disabling pain, and medical costs. *Psychiatr Serv* 2009;60(3):344–50.
- 49 Gameroff MJ, Olsson M. Major depressive disorder, somatic pain, and health care costs in an urban primary care practice. *J Clin Psychiatry* 2006;67(8):1232–9.
- 50 Dobscha SK, Corson K, Perrin NA, et al. Collaborative care for chronic pain in primary care: A clustered randomized trial. *JAMA* 2009;301(12):1242–52.
- 51 Goebel S, Steinert A, Vierheilg C, Faller H. Correlation between depressive symptoms and perioperative pain: A prospective cohort study of patients undergoing orthopedic surgeries. *Clin J Pain* 2013;29(5):392–9.
- 52 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 Rehabil* 2011;92(3):419–24.
- 53 Hoffman JM, Bombardier CH, Graves DE, Kalpakjian CZ, Krause JS. A longitudinal study of depression from 1 to 5 years after spinal cord injury. *Arch Phys Med Rehabil* 2011;92(3):411–8.
- 54 Koroschetz J, Rehm SE, Gockel U, et al. Fibromyalgia and neuropathic pain—Differences and similarities. A comparison of 3057 patients with diabetic painful neuropathy and fibromyalgia. *BMC Neurol* 2011;11:55.
- 55 Karp JF, Rollman BL, Reynolds CF, et al. Addressing both depression and pain in late life: The methodology of the ADAPT study. *Pain Med* 2012;13:405–18.
- 56 de Heer EW, Dekker J, van der Sluijs JFV, et al. Effectiveness and cost-effectiveness of transmural collaborative care with consultation letter (TCCCL) and

PROMIS Four-Item Depression and Anxiety Scales

- duloxetine for major depressive disorder (MDD) and (sub)chronic pain in collaboration with primary care: Design of a randomized placebo-controlled multi-Centre trial: TCC:PAINDIP. *BMC Psychiatry* 2013;13:147.
- 57 Kroenke K, Wu J, Bair MJ, et al. Impact of depression on 12-month outcomes in primary care patients with chronic musculoskeletal pain. *J Musculoskelet Pain* 2012;20:8–17.
- 58 Hyphantis T, Kotsis K, Voulgari PV, et al. Diagnostic accuracy, internal consistency, and convergent validity of the Greek version of the patient health questionnaire 9 in diagnosing depression in rheumatologic disorders. *Arthritis Care Res (Hoboken)* 2011;63(9):1313–21.
- 59 Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005;113(1–2):9–19.
- 60 Gibbons LE, Feldman BJ, Crane HM, et al. Migrating from a legacy fixed-format measure to CAT administration: Calibrating the PHQ-9 to the PROMIS depression measures. *Qual Life Res* 2011;20(9):1349–57.
- 61 Wahl I, Lowe B, Bjorner JB, et al. Standardization of depression measurement: A common metric was developed for 11 self-report depression measures. *J Clin Epidemiol* 2014;67:73–86.
- 62 Choi SW, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: Linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. *Psychol Assess* 2014;26:513–27.
- 63 Amtmann D, Kim J, Chung H, et al. Comparing CESD-10, PHQ-9, and PROMIS depression instruments in individuals with multiple sclerosis. *Rehabil Psychol* 2014;59:220–9.
- 64 Means-Christensen AJ, Sherbourne CD, Roy-Byrne PP, Craske MG, Stein MB. Using five questions to screen for five common mental disorders in primary care: Diagnostic accuracy of the Anxiety and Depression Detector. *Gen Hosp Psychiatry* 2006;28(2):108–18.
- 65 Houston JP, Kroenke K, Faries DE, et al. A provisional screening instrument for four common mental disorders in adult primary care patients. *Psychosomatics* 2011;52(1):48–55.
- 66 Hays RD, Reise S, Calderon JL. How much is lost in using single items? *J Gen Intern Med* 2012;27(11):1402–3.
- 67 Newman JC, Feldman R. Copyright and open access at the bedside. *N Engl J Med* 2011;365(26):2447–9.
- 68 Kroenke K. Enhancing the clinical utility of depression screening. *CMAJ* 2012;184(3):281–2.
- 69 Glasgow RE, Riley WT. Pragmatic measures: What they are and why we need them. *Am J Prev Med* 2013;45(2):237–43.
- 70 Narrow WE, Clarke DE, Kuramoto SJ, et al. DSM-5 field trials in the United States and Canada, Part III: Development and reliability testing of a cross-cutting symptom assessment for DSM-5. *Am J Psychiatry* 2013;170(1):71–82.