


RESEARCH ARTICLE

Examining the psychometric properties of brief screening measures of depression and anxiety in chronic pain: The Patient Health Questionnaire 2-item and Generalized Anxiety Disorder 2-item

Madelyne A. Bisby PhD  | Eyal Karin PhD | Amelia J. Scott PhD | Joanne Dudeney PhD | Alana Fisher PhD | Milena Gandy PhD | Taylor Hathway B Psych (Hons) | Andreea I. Heriseanu PhD | Lauren Staples PhD | Nickolai Titov PhD | Blake F. Dear PhD

eCentreClinic, School of Psychological Sciences, Macquarie University, Sydney, Australia

Correspondence

Madelyne A. Bisby, eCentreClinic, School of Psychological Sciences, Macquarie University, Sydney, Australia.
Email: madelyne.bisby@mq.edu.au

Abstract

Objective: Individuals with chronic pain experience anxiety and depressive symptoms at rates higher than the general population. The Patient Health Questionnaire 2-item (PHQ-2) and Generalized Anxiety Disorder 2-item (GAD-2) are brief screening measures of depression and anxiety, respectively. These brief scales are well-suited for use in routine care due to their brevity and ease of administration, yet their psychometric properties have not been established in heterogeneous chronic pain samples when administered over the Internet.

Materials and Methods: Using existing data from randomized controlled trials of an established Internet-delivered pain management program ($n = 1333$), we assessed the reliability, validity, diagnostic accuracy, and responsiveness to treatment change in the PHQ-2 and GAD-2, as well as the long-form counterparts. Exploratory analyses were conducted to obtain cutoff scores using those participants with diagnostic data ($n = 62$).

Results: The PHQ-2 and GAD-2 demonstrated appropriate reliability (eg, Cronbach's $\alpha = 0.79$ – 0.84), validity (eg, higher scores in individuals with a diagnosis; $p < 0.001$), and responsiveness to treatment change (eg, pre- to post-treatment scores, $p < 0.001$). The psychometric properties of the short forms compared well with the longer forms. Cutoff scores on the short forms were consistent with general population samples, while cutoff scores on the long forms were higher than previously observed using general population samples. All four scales favored specificity over sensitivity.

Conclusions: The PHQ-2 and GAD-2 demonstrated acceptable psychometric properties in the current sample, as did the long forms. Based on our findings, the PHQ-2 and GAD-2 can be used as screening tools with chronic pain samples when administered over the Internet.

KEYWORDS

anxiety, chronic pain, depression, measurement

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Pain Practice* published by Wiley Periodicals LLC on behalf of World Institute of Pain.

INTRODUCTION

Individuals with chronic pain experience anxiety and depressive symptoms at rates higher than the general population.¹⁻³ Given mental health problems are associated with poorer functional outcomes and can complicate clinical care, it is important to screen for, and routinely monitor, symptoms of anxiety and depression in individuals with chronic pain.^{4,5} Although the chronic pain literature has tended to emphasize the role of depressive symptoms in chronic pain, it is also important to identify and respond to elevated symptoms of anxiety in this population.^{6,7} Such activities will be facilitated by brief screening tools that are valid and reliable, and can be quickly administered, scored, and interpreted.

The Patient Health Questionnaire 9-item (PHQ-9) and Generalized Anxiety Disorder 7-item (GAD-7) were designed to screen for major depressive disorder and generalized anxiety disorder, respectively.^{8,9} These scales were originally developed in the primary care context and are well-established for use in psychological treatment-seeking samples with excellent psychometric properties, including test-retest reliability, validity, and responsiveness to treatment change.¹⁰⁻¹² Ultra-brief two-item Patient Health Questionnaire 2-item (PHQ-2) and Generalized Anxiety Disorder Scale 2-item (GAD-2) scales have been created as screening measures from the longer, original forms. Although substantially shorter, the PHQ-2 and GAD-2 have demonstrated acceptable reliability and validity in treatment-seeking general population samples.¹³⁻¹⁵ There are numerous potential advantages to these short forms, including that they offer a reliable method of symptom assessment in a much shorter timeframe—a significant benefit in time-constrained clinical practice settings. Containing fewer items also allows the PHQ-2 and GAD-2 to be combined with other brief scales to assess a broader range of outcomes, which is often essential in pain assessment and management within routine care.

As individuals with chronic pain often experience symptoms which overlap with those of depression and anxiety, screening tools need to be validated within chronic pain samples to ensure that they are suitable. Indeed, the interpretation of these measures (eg, the use of cutoff scores) may be different for individuals with chronic pain due to the endorsement of somatic symptoms. For instance, the PHQ-9 and GAD-7 include items relating to difficulties with sleep and fatigue, both of which are common in chronic pain regardless of psychiatric comorbidity.¹⁶⁻¹⁸ In contrast, the PHQ-2 and GAD-2 do not contain these somatic symptom items and assess only the central cognitive and behavioral symptoms of depression and anxiety. In the light of the potential somatic symptom bias in the original forms, the short forms may be particularly well-suited for use in a chronic pain context.

It is important to extend the psychometric evaluation of these scales to chronic pain to ensure appropriate interpretation of scores in research and clinical practice. The available evidence supports the use of the PHQ-9 and GAD-7 in specific populations, including those with migraine^{19,20} or rheumatoid arthritis.²¹ Recently, Kroenke et al.^{22,23} examined the discriminative validity and responsiveness of the PHQ-9 and PHQ-2 in a mixed sample of primary care patients with chronic pain and stroke survivors. The scales demonstrated good to excellent diagnostic performance, as well as responsiveness to treatment (with the exception of the PHQ-2 in one out of three clinical trials). However, a psychometric evaluation of the GAD forms in a heterogeneous chronic pain sample has yet to be undertaken (in contrast to the PHQ forms). In addition, the PHQ and GAD forms have yet to be validated for use with individuals with chronic pain when administered over the Internet, rather than face to face. Lastly, despite the value of the short forms in this population, sample-specific cutoff scores on the PHQ-2 and GAD-2 have yet to be examined.

The current study aimed to evaluate the psychometric properties of the PHQ-2 and GAD-2, as well as the PHQ-9 and GAD-7, including reliability, validity, diagnostic accuracy, and responsiveness to change following treatment. This was achieved in a secondary analysis of previous randomized controlled trials of a pain management program for chronic pain, in which the forms were administered over the Internet. We hypothesized that the short forms would perform well against the original measures, consistent with past psychometric evaluations in the general population. Using a sub-sample of participants who underwent a diagnostic assessment, we propose preliminary cutoff scores for the GAD-2 and PHQ-2, as well as their long-form counterparts. As the available sub-sample was small, we acknowledge the preliminary nature of our examination. In any case, we predicted that the cutoff scores for the short forms would be consistent with past work in the general population, whereas the cutoff scores for the PHQ-9 and GAD-7 may be higher in the chronic pain sample due to the endorsement of overlapping somatic symptoms of chronic pain and psychological distress.

MATERIALS AND METHODS

Sample characteristics

The participant sample ($n = 1333$) was derived from four previously conducted randomized controlled trials examining the effectiveness of a remotely delivered pain management program.²⁴⁻²⁷ As this project used existing data, ethics approval was not required. The following inclusion criteria were used to screen applicants across three of the four trials: (1) pain for longer than 6 months, (2) pain had been assessed by a medical professional, (3)

Australian resident, (4) at least 18 years of age, (5) access to a computer and the Internet, and (6) not currently experiencing a psychotic illness, severe depression, or suicidal ideation. One trial ($n = 62$) included additional inclusion criteria: The duration of chronic pain required was lessened to 3 months, applicants were required to be on a stable dose of medication (longer than one month), and applicants could not be receiving current cognitive behavior therapy. The demographics of the sample are shown in Table 1.

Treatment

Following an initial assessment, participants were randomly allocated to begin treatment either immediately or following a waitlist period. In the current sample, 912 participants were assigned to the treatment groups, and 421 participants were assigned to the control groups. The pain course is an 8-week psychological pain management program based on the principles of cognitive behavior therapy and includes five lessons, practice exercises, additional resources, and case stories. Across

TABLE 1 Demographic and clinical characteristics of chronic pain sample ($n = 1333$)

Demographics	
Age	52.06 (14.11)
Female	82.67%
Diagnoses prior to treatment ($n = 62$)	
Major depressive disorder	59.67% ($n = 37$)
Generalized anxiety disorder	35.48% ($n = 22$)
Baseline symptoms	
PHQ-9	11.62 (5.26)
PHQ-9 ≥ 10	63.84% ($n = 851$)
GAD-7	8.08 (5.03)
GAD-7 ≥ 10	36.08% ($n = 481$)
Chronic pain conditions	
Muscular pain	62.26%
Fibromyalgia	24.45%
Osteoarthritis	20.03%
Headache or migraine	9.97%
Neuropathy	7.87%
Pain characteristics	
Average pain intensity ^a	5.80 (1.53)
Average duration of pain (years)	9.26 (7.54)
Attended specialist pain clinic ($n = 1169$)	45.42%
Medication use ($n = 1271$)	
Pain	77.10%
Mental health	45.16%

Note: Mean (standard deviation).

^aWisconsin Brief Pain Questionnaire Item 3. Scored 0 (no pain at all) to 10 (pain as bad as you can imagine).

the four trials, the course was delivered mostly over the Internet using a purpose-built clinical software platform, although some received the course in a hard-copy workbook format. In addition, participants were randomized to proceed through the course in either a clinician-guided or self-guided manner. The primary trials provided considerable evidence that the pain course leads to clinically meaningful improvements in anxiety, depression, disability, and average pain intensity. The results of these trials, and further details regarding the impact of delivery format and clinical guidance can be found elsewhere.²⁴⁻²⁷

Measures

The measures used in the current psychometric evaluation were taken from a larger questionnaire battery which participants completed at three time-points: initial assessment, pre-treatment, and post-treatment.²⁴⁻²⁷ All measures were administered online except for the Mini International Neuropsychiatric Interview (MINI) interview, which was administered via telephone. There were 2-4 weeks between assessment and pre-treatment, and then 8-12 weeks between pre-treatment to post-treatment.

Patient Health Questionnaire 9-item and Patient Health Questionnaire 2-item

The PHQ-9 includes nine items assessing the severity of depressive symptoms.⁸ Each item is scored on a four-point Likert scale from 0 (not at all) to 3 (nearly every day). Total scores on the PHQ-9 range from 0 to 27, and a score above 10 is indicative of clinical depression in general population samples.¹² The PHQ-2 consists of two cognitive items from the PHQ-9 (*Over the last two weeks, how often have you been bothered by (1) little interest or pleasure in doing things, and (2) feeling down, depressed, or hopeless*). Scores on the PHQ-2 range from 0 to 6, and scores above either 2 or 3 have been argued to indicate clinical depression.^{8,13}

Generalized Anxiety Disorder Scale 7-item and Generalized Anxiety Disorder Scale 2-item

The GAD-7 includes seven items which are rated on a four-point Likert scale from 0 (not at all) to 3 (nearly every day). The GAD-7 is a questionnaire designed to assess the severity of symptoms of generalized anxiety disorder, but is also sensitive to the presence of social anxiety disorder and panic disorder.⁹ Total scores on the GAD-7 range from 0 to 21, and a score of 10 was originally proposed as indicative of clinical anxiety in general population samples,⁹ although recent work shows that a

cutoff score of 8 is also acceptable.¹⁵ The GAD-2 consists of two cognitive items from the GAD-7 (*Over the last two weeks, how often have you been bothered by (1) feeling nervous, anxious, or on edge, and (2) not being able to stop or control worrying*). Total scores on the GAD-2 range from 0 to 6, and scores ≥ 3 are considered indicative of clinical anxiety.^{15,28}

Mini International Neuropsychiatric Interview Version 5

The MINI is a structured interview used to obtain diagnostic information.²⁹ The MINI is a valid and reliable interview that is considered the gold-standard for detecting DSM diagnoses. In one of the included randomized controlled trials,²⁵ the modules for major depressive disorder and generalized anxiety disorder were administered via telephone at assessment.

Kessler Psychological Distress Scale 10-item

The Kessler Psychological Distress Scale 10-item (K-10) is a broad measure of psychological distress which consists of 10-items scored on a five-point Likert scale from 1 (none of the time) to 5 (all of the time). A total score equal to or above 22 are considered indicative of anxiety or depression.^{30,31} In the current sample, the K-10 had good internal consistency ($\alpha = 0.87$).

Statistical analyses

Reliability

Internal consistency was determined using Cronbach's alpha (α) and item-total correlations. Cronbach's alpha is a widely used measure of internal consistency, and values of 0.76–0.90 were considered good, and values of 0.90 or above as excellent. Item-total correlations above 0.30 were considered satisfactory. For the two-item scales, Spearman–Brown coefficients were calculated as an additional measure of reliability.³² Test–retest reliability was determined by Pearson correlations between scores at assessment and pre-treatment, as well as from pre-treatment to post-treatment in the treatment group ($n = 912$) and control group ($n = 421$). Test–retest correlations above 0.70 were considered acceptable.

Validity

The sub-sample of participants that were administered the MINI ($n = 62$) as taken from²⁵ were used for ROC analysis. The area under the curve indicates the degree

of discrimination; values between 0.70 and 0.79 are acceptable, and those of or above 0.80 are excellent.³³ Sensitivity and specificity were reported as indications of cutoff scores in detecting individuals with a diagnosis. The likelihood ratio was reported as a measure of how likely a positive result would be among those who do indeed have a diagnosis relative to those who do not. Criterion validity was determined using one-way ANOVA to compare scores in individuals with or without diagnoses of generalized anxiety disorder and major depressive disorder, respectively. Construct validity was assessed using Pearson correlations between the PHQ and GAD forms and the K10, a conceptually related measure of general distress.

Responsiveness to change

To assess the responsiveness of these scales to treatment effects, generalized estimating equations (GEE) with a gamma distribution and log link response scale were used. These analyses were conducted using only the cases from the treatment group sample without missing data at post-treatment ($n = 824$, 90.35%). Symptom reductions at post-treatment were also calculated as percentage change from pre-treatment. Consistent with past work examining clinically significant change in anxiety, depression, and pain,^{27,34} the proportions of the sample achieving reductions of $\geq 30\%$ and $\geq 50\%$ change following treatment were compared across the short and long forms.

RESULTS

Descriptives

The chronic pain sample ($n = 1333$) was comprised mainly of females (82.67%) with an average age of 52.06 years ($SD = 14.11$; see Table 1). Muscular pain was the most common chronic pain condition, followed by fibromyalgia and osteoarthritis. In the sub-sample of participants for whom diagnostic information was available, 59.67% ($n = 37$) met diagnostic criteria for major depressive disorder and 35.48% ($n = 22$) for generalized anxiety disorder. Most participants in the sample reported moderate symptoms of depression on the PHQ-9 and mild symptoms of anxiety on the GAD-7. Using the Wisconsin Brief Pain Questionnaire,³⁵ individuals rated their average pain as 5.80 out of 10 ($SD = 1.53$). On average, individuals had experienced chronic pain for over 9 years, and 45.42% had attended a specialist pain clinic. Most participants (77.10%) were taking prescription medication for their pain, and almost half of the participants took prescription medication for their mental health (45.16%).

Reliability

Internal consistency was good for both PHQ forms with minimal difference between the short (PHQ-2: $\alpha = 0.79$) and long forms (PHQ-9: $\alpha = 0.84$) using Cronbach's alpha. Similar results were obtained for the two GAD forms, which demonstrated comparable internal consistency (GAD-2: $\alpha = 0.84$, GAD-7: $\alpha = 0.89$). The Spearman–Brown coefficients for the short scales were also acceptable (PHQ-2: 0.79; GAD-2: 0.84). For all four scales, the item-total correlations were acceptable, and the internal consistency of the long forms would not have improved with deletion of later scale items (see Table 2).

Test–retest reliability was examined using scores from assessment and pre-treatment, and then from pre-treatment and post-treatment. In the treatment group, all four scales demonstrated acceptable test–retest reliability from assessment to pre-treatment. However, there was no association between scores at pre-treatment and post-treatment, indicating that the scores were responsive to symptom change as a result of treatment (see Table 3). In contrast, scores were significantly correlated between assessment and pre-treatment, as well as pre-treatment and post-treatment, in the control group. The

test–retest reliability of the brief forms (PHQ-2 and GAD-2) was slightly, but not substantially, lower than the long forms (PHQ-9 and GAD-7).

Validity

Using the available sub-sample of participants with diagnostic interview data ($n = 62$), scores on the PHQ and GAD forms were compared between participants with and without formal diagnoses. Scores on the PHQ-2 and PHQ-9 were significantly different between those individuals with and without a diagnosis of major depressive disorder (PHQ-2: $F_{1,60} = 14.96$, $p < 0.001$; PHQ-9: $F_{1,60} = 28.21$, $p < 0.001$). Similarly, scores on the GAD-2 and GAD-7 differed significantly between individuals with and without a diagnosis of generalized anxiety disorder (GAD-2: $F_{1,60} = 11.80$, $p = 0.001$; GAD-7: $F_{1,60} = 18.12$, $p < 0.001$), indicating good criterion validity. There was no difference in discriminative validity between the long and short forms. The correlations with the K-10, a conceptually related measure, were acceptable for all scales: PHQ-9: 0.73, PHQ-2: 0.60, GAD-7: 0.76, and GAD-2: 0.69.

TABLE 2 Means, standard deviations, item-total correlations, and Cronbach's alpha if item deleted ($n = 1333$)

	Mean	SD	Item-total	Alpha
PHQ-9 ($\alpha = 0.84$)				
Item 1	1.26	0.87	0.67	0.81
Item 2	1.10	0.85	0.65	0.81
Item 3	1.85	0.99	0.51	0.83
Item 4	2.04	0.90	0.57	0.82
Item 5	1.43	1.07	0.57	0.82
Item 6	1.10	0.95	0.61	0.82
Item 7	1.19	0.96	0.61	0.82
Item 8	0.78	0.93	0.47	0.83
Item 9	0.22	0.50	0.34	0.84
PHQ-2 ($\alpha = 0.79$; Spearman–Brown = 0.79)				
Item 1	1.26	0.87	0.65	
Item 2	1.10	0.85	0.65	
GAD-7 ($\alpha = 0.89$)				
Item 1	1.23	0.92	0.74	0.87
Item 2	1.11	0.99	0.81	0.86
Item 3	1.26	0.96	0.80	0.86
Item 4	1.44	0.96	0.73	0.87
Item 5	0.75	0.89	0.56	0.89
Item 6	1.37	0.92	0.57	0.89
Item 7	0.73	0.90	0.63	0.89
GAD-2 ($\alpha = 0.84$; Spearman–Brown = 0.84)				
Item 1	1.23	0.92	0.72	
Item 2	1.11	0.96	0.72	

Diagnostic accuracy

ROC curve analyses were conducted using pre-treatment scores across the whole sample to determine the diagnostic accuracy of the scales. The PHQ-2 and PHQ-9 both demonstrated acceptable discriminative validity (AUCs 0.77–0.83; see Table 4). Similar results were obtained for the GAD-2 and GAD-7 (AUCs 0.72–0.80).

Sensitivity and specificity are presented in Table 4. Consistent with previous research using general population samples, the optimal cutoff score for the GAD-2 was ≥ 3 (sensitivity: 50, specificity: 90, likelihood ratio: 5.00). However, ROC analyses suggested a lower cutoff score for the PHQ-2 at ≥ 2 (sensitivity: 62, specificity: 88, likelihood ratio: 5.18). For the long forms, ROC analyses suggested cutoff scores at higher points than previously reported using general population samples. For the PHQ-9, a score of $\geq 12/13$ indicated optimal sensitivity and specificity (sensitivity: 68, specificity: 80, likelihood ratio: 3.38), while the GAD-7 showed optimal sensitivity and specificity at ≥ 11 (sensitivity: 68, specificity: 90, likelihood ratio: 6.82).

Responsiveness to change

Generalized estimating equations analyses were conducted to assess the responsiveness of the self-report scales to change over time in the treatment group using only those participants with complete post-treatment data ($n = 824$). For the PHQ forms, there was a significant

TABLE 3 Outcome measures at assessment, pre-treatment, and post-treatment

	Assessment		Pre-treatment		Post-treatment		Correlations	
	Mean	SD	Mean	SD	Mean	SD	Assessment—pre	Pre-post
Treatment	<i>n</i> = 912		<i>n</i> = 912		<i>n</i> = 824		<i>n</i> = 912	
PHQ-9	11.52	5.19	11.08	5.39	7.49	5.24	0.72*	0.16
PHQ-2	2.51	1.47	2.42	1.56	1.62	1.39	0.63*	0.27
GAD-7	8.12	5.01	7.96	5.02	5.63	4.73	0.73*	0.01
GAD-2	2.44	1.74	2.39	1.76	1.72	1.64	0.68*	0.00
Control	<i>n</i> = 421		<i>n</i> = 421		<i>n</i> = 403		<i>n</i> = 421	
PHQ-9	11.86	5.40	10.72	5.43	10.75	5.43	0.77*	0.54*
PHQ-2	2.55	1.54	2.21	1.54	2.28	1.60	0.65*	0.47*
GAD-7	7.97	5.06	7.68	5.27	7.45	5.00	0.76*	0.55*
GAD-2	2.37	1.73	2.23	1.79	2.23	1.75	0.70*	0.53*

*Statistical significance ($p < 0.01$).

TABLE 4 Diagnostic properties of anxiety and depression measures ($n = 62$)

	PHQ-2 AUC: 0.77 (0.65–0.87)			GAD-2 AUC: 0.72 (0.59–0.83)		
	Sensitivity	Specificity	Likelihood ratio	Sensitivity	Specificity	Likelihood ratio
>2	62	88	5.18	64	60	1.59
>3 ^a	38	88	3.15	50	90	5.00
>4	22	96	5.41	27	95	5.45

	PHQ-9 AUC: 0.83 (0.72–0.92)			GAD-7 AUC: 0.80 (0.67–0.89)		
	Sensitivity	Specificity	Likelihood ratio	Sensitivity	Specificity	Likelihood ratio
>5	100	24	1.32	91	28	1.25
>6	97	32	1.43	86	38	1.38
>7	97	36	1.52	86	50	1.73
>8	89	56	2.03	82	65	2.34
>9	81	60	2.03	73	70	2.42
>10 ^b	81	60	2.03	73	83	4.16
>11	73	64	2.03	68	90	6.82
>12	68	80	3.38	45	90	4.55
>13	68	80	3.38	27	93	3.64

Note: The Mini International Neuropsychiatric Interview (MINI) was used as the gold-standard reference measure.

Values in bold indicate the optimal cutoff score.

^aA cutoff of 3 is recommended for the PHQ-2 and GAD-2.^{14,15}

^bA cutoff of 10 is recommended for the PHQ-9 and GAD-7.^{8,9}

TABLE 5 Estimates of clinical change (95% CIs) on outcome measures following treatment ($n = 824$)

	PHQ-9	PHQ-2	GAD-7	GAD-2
Group-level	35% (31%–39%)	34% (29%–39%)	33% (28%–38%)	32% (26%–38%)
Proportion $\geq 30\%$	0.59 (0.55–0.62)	0.58 (0.55–0.61)	0.58 (0.54–0.61)	0.59 (0.55–0.62)
Proportion $\geq 50\%$	0.41 (0.37–0.44)	0.48 (0.44–0.51)	0.42 (0.38–0.46)	0.50 (0.46–0.54)

decrease in scores from pre-treatment to post-treatment for both the PHQ-2 (Wald's $\chi^2 = 195.96$, $p < 0.001$) and PHQ-9 (Wald's $\chi^2 = 409.01$, $p < 0.001$). Similar results were obtained for the GAD forms, such that there was a significant decrease in scores from pre-treatment to post-treatment for the GAD-2 (Wald's $\chi^2 = 151.46$, $p < 0.001$) and the GAD-7 (Wald's $\chi^2 = 216.83$, $p < 0.001$).

At the group level, PHQ-2 scores decreased by 34% (95% CI 29%–39%, $d = 0.55$), and PHQ-9 scores decreased by 35% (95% CI 31%–39%, $d = 0.73$; see Table 5). Similar proportions of the sample achieved $\geq 30\%$ symptom change (0.59 vs. 0.58) and $\geq 50\%$ symptom change (0.41 vs. 0.48) on the PHQ forms. Average scores on the GAD-2 decreased by 32% (95% CI 26%–38%, $d = 0.46$), while average scores on the GAD-7 scores decreased by 33% (95% CI 28% to 38%, $d = 0.54$). Likewise, on the GAD forms, comparable proportions of the sample achieved $\geq 30\%$ symptom change (0.58 vs. 0.59) and $\geq 50\%$ symptom change (0.42 vs. 0.50).

DISCUSSION

The PHQ-2 and GAD-2 are brief screening measures of depression and anxiety symptoms that are well-suited for use with individuals experiencing chronic pain in clinical practice. These brief scales were derived from the longer versions, the PHQ-9 and GAD-7. However, no research to date has compared the validity and reliability of the GAD-2 to the GAD-7 in individuals with heterogeneous chronic pain conditions. In a secondary analysis of a large sample of psychological treatment-seeking individuals with chronic pain who participated in clinical trials, both PHQ and GAD forms identified those with a diagnosis of depression or anxiety, respectively, and were sensitive to symptom change across treatment. The psychometric properties of the short and long forms were also comparable. Thus, the findings of this study suggest that the brief versions can be used in individuals with chronic pain to screen for anxiety and depressive symptoms and diagnostic status over the Internet, and that these scales can accurately track changes in symptoms over time. Thus, the brief versions have significant potential for time-pressured settings, such as routine clinical care.

The current study is the first to suggest cutoff scores for screening anxiety and depressive symptoms in a heterogeneous chronic pain sample who completed the PHQ and GAD forms over the Internet. However, our results should be considered exploratory as only a small sub-sample of participants ($n = 62$) had diagnostic data available. Nevertheless, it is worthwhile to compare our preliminary findings to the existing literature. First, the original validation of the GAD-2 ($n = 965$) recommended a cutoff score of ≥ 3 , and subsequent meta-analytic evaluations support this finding.^{15,36} With respect to chronic pain, a cutoff score of ≥ 3 on the GAD-2 performed well in

individuals with rheumatoid arthritis,²¹ and is consistent with that obtained in the current study. On the contrary, the current study suggests that the PHQ-2 demonstrates optimal sensitivity and specificity at a lower cutoff score of ≥ 2 . This suggestion is in contrast to both the original validation study of the PHQ-2 and a study of individuals with rheumatoid arthritis which recommended a cutoff score of ≥ 3 on the PHQ-2.^{21,37} However, subsequent validation studies have suggested the cutoff score be lowered to ≥ 2 to achieve an optimal balance of sensitivity and specificity.^{38–40} Therefore, the clinical interpretations of the PHQ-2 and GAD-2 as screening measures are consistent with that recommended for the general population and is not altered in a chronic pain sample.

In contrast to the cutoff scores obtained on short forms, our analyses indicated that the cutoff scores for the longer forms may be higher in chronic pain. Whereas previous studies have recommended cutoff scores ≥ 10 for both the GAD-7 and PHQ-9, we observed optimal cutoff scores of ≥ 11 for the GAD-7 and $\geq 12/13$ for the PHQ-9. One explanation for this is that the longer forms contain items that assess somatic symptoms often caused by pain and pain medications, whereas the short forms do not. For instance, sleep difficulties are a common symptom of both presentations and emerge as a central component of anxiety, depression, and chronic pain in network models.⁴¹ Sleep difficulties are captured on the GAD-7 as *trouble relaxing* (Item 4), as well as on the PHQ-9 as *trouble falling or staying asleep or sleeping too much* (Item 3). This is consistent with previous work illustrating that the physical symptoms of medical illness can lead to biased mood assessments. For instance, individuals with systemic sclerosis have higher PHQ-9 somatic item scores compared with healthy controls, even when matched on their PHQ-9 cognitive/affective item scores.¹⁸ To maintain discriminant validity, the cutoff scores employed for the PHQ-9 and GAD-7 may need to be higher in chronic pain patients compared to the general population.

It should also be noted that in ROC analyses, all four scales favored specificity over sensitivity. In psychometric evaluations, there is a balance between sensitivity (ie, the ability to correctly identify positive cases) and specificity (ie, the ability to correctly identify negative cases).^{42–45} The finding that specificity was favored over sensitivity in the current study indicates that some true cases may go undetected when using the PHQ or GAD forms as screening measures. This is a potentially important finding and highlights the need for replication, as well as the value of future research using large samples with diagnostic data.

The results of the current study should be acknowledged in light of several limitations. Firstly, and perhaps most importantly, the diagnostic interview data was only available for a sub-group of individuals ($n = 62$) who were subject to other eligibility criteria (eg, experience of chronic pain for a minimum of 3 months rather than 6 months). Importantly, participants who

reported very severe depressive symptoms on the PHQ-9 (ie, total score ≥ 22 , item 9 score ≥ 2) were excluded from the original study, and therefore, our analyses were conducted in a somewhat restricted sample. Future work is needed to replicate our findings using more representative samples. On a related note, the small sample size is likely to have contributed to the minimal distinction between the cutoff scores of ≥ 12 versus ≥ 13 on the PHQ-9, and further replication with larger samples and updated diagnostic interviews (eg, MINI version 7) is needed. Second, we did not combine the PHQ-2 and GAD-2 to examine the PHQ-4, a composite measure of general distress.⁴³ Although the PHQ-4 has not been investigated to the same extent as the individual short forms, psychometric evaluations of ultra-brief scales assessing more general distress, such as the PHQ-4 and the Kessler Psychological Distress Scale 6 item⁴² are warranted.

The results of this study indicate that the psychometric properties of the ultra-brief PHQ-2 and GAD-2 forms are robust when administered over the Internet to individuals with heterogeneous chronic pain conditions, at least within a clinical trial context. The performance of the scales was not compromised in chronic pain; instead, the scales demonstrated acceptable reliability, validity, and responsiveness to symptom change. However, the interpretation of PHQ-9 and GAD-7 scores need to be carefully considered when working with individuals with chronic pain to avoid artificial inflation of clinical anxiety and/or depressive symptoms. The increase in cutoff scores in the current sample was likely due to pain-related endorsement of somatic items, although this remains to be confirmed in further evaluations. As the PHQ-2 and GAD-2 do not include the somatic items of the longer scales, they are less vulnerable to the impact of somatic symptoms and accurately track change over time. These short forms may therefore be more appropriate than the long forms for use in chronic pain. Indeed, the PHQ-2 and GAD-2 appear to offer brief, reliable, and valid measures for symptom assessment and where necessary, monitoring of symptom change, in people with chronic pain.

ACKNOWLEDGEMENT

Open access publishing facilitated by Macquarie University, as part of the Wiley - Macquarie University agreement via the Council of Australian University Librarians. [Correction added on 20 May 2022, after first online publication: CAUL funding statement has been added.]

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Those interested in accessing data and study materials may contact the first author via email.

ORCID

Madelyne A. Bisby  <https://orcid.org/0000-0002-4354-3684>

REFERENCES

1. Elbinoune I, Amine B, Shyen S, Gueddari S, Abouqal R, Hajjaj-Hassouni N. Chronic neck pain and anxiety-depression: prevalence and associated risk factors. *Pan Afr Med J*. 2016;24:1–8.
2. Lerman SF, Rudich Z, Brill S, Shalev H, Shahar G. Longitudinal associations between depression, anxiety, pain, and pain-related disability in chronic pain patients. *Psychosom Med*. 2015;77(3):333–41.
3. Siqueira-Campos VM, Da Luz RA, de Deus JM, Martinez EZ, Conde DM. Anxiety and depression in women with and without chronic pelvic pain: prevalence and associated factors. *J Pain Res*. 2019;12:1223–33.
4. Tseli E, Boersma K, Stålnacke B-M, Enthoven P, Gerdle B, Ång BO, et al. Prognostic factors for physical functioning after multidisciplinary rehabilitation in patients with chronic musculoskeletal pain. *Clin J Pain*. 2019;35(2):148–73.
5. Rayner L, Hotopf M, Petkova H, Matcham F, Simpson A, McCracken LM. Depression in patients with chronic pain attending a specialised pain treatment centre: prevalence and impact on health care costs. *Pain*. 2016;157(7):1472–9.
6. Jordan KD, Okifuji A. Anxiety disorders: differential diagnosis and their relationship to chronic pain. *J Pain Palliat Care Pharmacother*. 2011;25(3):231–45.
7. Oliveira DS, Mendonça LVF, Sampaio RSM, Dias De Castro-Lopes JMP, De Azevedo LFR. The impact of anxiety and depression on the outcomes of chronic low back pain multidisciplinary pain management – a multicenter prospective cohort study in pain clinics with one-year follow-up. *Pain Med*. 2019;20(4):736–46.
8. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16:606–13.
9. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006;166(10):1092–7.
10. Johnson SU, Ulvenes PG, Øktedalen T, Hoffart A. Psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. *Front Psychol*. 2019;10:1–8.
11. Beard C, Björgvinsson T. Beyond generalized anxiety disorder: psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. *J Anxiety Disord*. 2014;28(6):547–52. <https://doi.org/10.1016/j.janxdis.2014.06.002>
12. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *Gen Hosp Psychiatry*. 2010;32(4):345–59. <https://doi.org/10.1016/j.genhosppsych.2010.03.006>
13. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann Fam Med*. 2010;8(4):348–53.
14. Staples LG, Dear BF, Gandy M, Fogliati VJ, Fogliati R, Karin E, et al. Psychometric properties and clinical utility of brief measures of depression, anxiety, and general distress: the PHQ-2, GAD-2, and K-6. *Gen Hosp Psychiatry*. 2019;56:13–8. <https://doi.org/10.1016/j.genhosppsych.2018.11.003>
15. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. *Gen Hosp Psychiatry*. 2016;2:24–31. <https://doi.org/10.1016/j.genhosppsych.2015.11.005>
16. Finan PH, Goodin BR, Smith MT. The association of sleep and pain: an update and a path forward. *J Pain*. 2013;14:1539–52.
17. Eccles JA, Davies KA. The challenges of chronic pain and fatigue. *Clin Med J R Coll Physicians London*. 2021;21(1):19–27.

18. Leavens A, Patten SB, Hudson M, Baron M, Thombs BD. Influence of somatic symptoms on patient health questionnaire-9 depression scores among patients with systemic sclerosis compared to a healthy general population sample. *Arthritis Care Res.* 2012;64(8):1195–201.
19. Seo J-G, Park S-P. Validation of the Generalized Anxiety Disorder-7 (GAD-7) and GAD-2 in patients with migraine. *J Headache Pain.* 2015;16(1):1–7. <https://doi.org/10.1186/s10194-015-0583-8>
20. Seo J-G, Park S-P. Validation of the Patient Health Questionnaire-9 (PHQ-9) and PHQ-2 in patients with migraine. *J Headache Pain.* 2015;16(1):1–7. <https://doi.org/10.1186/s10194-015-0552-2>
21. Hitchon CA, Zhang L, Peschken CA, Lix LM, Graff LA, Fisk JD, et al. Validity and reliability of screening measures for depression and anxiety disorders in rheumatoid arthritis. *Arthritis Care Res.* 2020;72(8):1130–9.
22. Kroenke K, Stump TE, Chen CX, Kean J, Damush TM, Bair MJ, et al. Responsiveness of PROMIS and Patient Health Questionnaire (PHQ) Depression Scales in three clinical trials. *Health Qual Life Outcomes.* 2021;19(1):1–14. <https://doi.org/10.1186/s12955-021-01674-3>
23. Kroenke K, Stump TE, Kean J, Krebs EE, Damush TM, Bair MJ, et al. Diagnostic operating characteristics of PROMIS scales in screening for depression. *J Psychosom Res.* 2021;147:110532. <https://doi.org/10.1016/j.jpsychores.2021.110532>
24. Dear BF, Gandy M, Karin E, Ricciardi T, Fogliati VJ, McDonald S, et al. The pain course: a randomised controlled trial comparing a remote-delivered chronic pain management program when provided in online and workbook formats. *Pain.* 2017;158(7):1289–301.
25. Dear BF, Titov N, Perry KN, Johnston L, Wootton BM, Terides MD, et al. The pain course: a randomised controlled trial of a clinician-guided internet-delivered cognitive behaviour therapy program for managing chronic pain and emotional well-being. *Pain.* 2013;154(6):942–50. <https://doi.org/10.1016/j.pain.2013.03.005>
26. Dear BF, Gandy M, Karin E, Staples LG, Johnston L, Fogliati VJ, et al. The pain course: a randomised controlled trial examining an internet-delivered pain management program when provided with different levels of clinician support. *J Pain.* 2015;156:1920–35.
27. Dear BF, Karin E, Fogliati R, Dudeney J, Nielssen O, Gandy M, et al. The pain course: a randomised controlled trial and economic evaluation of an internet-delivered pain management program. *Pain.* 2021. <https://doi.org/10.1097/j.pain.0000000000002507>. Online ahead of print.
28. Hughes AJ, Dunn KM, Chaffee T, Bhattarai J, Beier M. Diagnostic and clinical utility of the GAD-2 for screening anxiety symptoms in individuals with multiple sclerosis. *Arch Phys Med Rehabil.* 2018;99(10):2045–9.
29. Sheehan DS, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry.* 1998;59:22–33.
30. Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, et al. Screening for serious mental illness in the general population. *Arch Gen Psychiatry.* 2003;60(2):184–9.
31. Australian Bureau of Statistics. Information paper: use of the Kessler Psychological distress scale in ABS health surveys, Australia, 2007–08. Canberra, ACT: Australian Bureau of Statistics; 2012.
32. Eisinga R, Te GM, Pelzer B. The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *Int J Public Health.* 2013;58(4):637–42.
33. Jessen HC, Menard S. Applied logistic regression analysis. *Stat.* 1996;45(4):534.
34. Moore AR, Eccleston C, Derry S, Wiffen P, Bell RF, Straube S, et al. “Evidence” in chronic pain – establishing best practice in the reporting of systematic reviews. *Pain.* 2010;150(3):386–9.
35. Daut RL, Cleeland CS, Flanery RC. Development of the Wisconsin Brief Pain Questionnaire to assess pain in cancer and other diseases. *Pain.* 1983;17(2):197–210.
36. Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Lowe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med.* 2007;151(10):678–85.
37. Kroenke K, Spitzer R, Williams J. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care.* 2003;41(11):1284–92.
38. Pedersen SS, Denollet J, De Jonge P, Simsek C, Serruys PW, Van Domburg RT. Brief depression screening with the PHQ-2 associated with prognosis following percutaneous coronary intervention with paclitaxel-eluting stenting. *J Gen Intern Med.* 2009;24(9):1037–42.
39. Giuliani M, Gorini A, Barbieri S, Veglia F, Tremoli E. Examination of the best cut-off points of PHQ-2 and GAD-2 for detecting depression and anxiety in Italian cardiovascular inpatients. *Psychol Health.* 2021;36(9):1088–101. <https://doi.org/10.1080/08870446.2020.1830093>
40. Manea L, Gilbody S, Hewitt C, North A, Plummer F, Richardson R, et al. Identifying depression with the PHQ-2: a diagnostic meta-analysis. *J Affect Disord.* 2016;2:382–95. <https://doi.org/10.1016/j.jad.2016.06.003>
41. Gómez Penedo JM, Rubel JA, Blättler L, Schmidt SJ, Stewart J, Eglhoff N, et al. The complex interplay of pain, depression, and anxiety symptoms in patients with chronic pain: a network approach. *Clin J Pain.* 2020;36(4):249–59.
42. Kessler R C, Andrews G, Colpe LJ, Hiripi E, Mroczek D K, Normand S-L T, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med.* 2002;32(6):959–76.
43. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics.* 2009;50(6):613–21.
44. Trevethan R. Sensitivity, specificity, and predictive values: foundations, plabilities, and pitfalls in research and practice. *Front Public Heal.* 2017;5:1–7.
45. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. *Contin Educ Anaesthesia Crit Care Pain.* 2008;8(6):221–3.

How to cite this article: Bisby MA, Karin E, Scott AJ, Dudeney J, Fisher A, Gandy M, et al. Examining the psychometric properties of brief screening measures of depression and anxiety in chronic pain: The Patient Health Questionnaire 2-item and Generalized Anxiety Disorder 2-item. *Pain Pract.* 2022;22:478–486. <https://doi.org/10.1111/papr.13107>