

Brief Psychological Screening Questions Can be Useful for Ruling Out Psychological Conditions in Patients with Chronic Pain

Vægter, Henrik Bjarke ; Handberg, Gitte; Kent, Peter

Published in:
The Clinical Journal of Pain

DOI:
10.1097/AJP.0000000000000514

Publication date:
2018

Document version:
Accepted manuscript

Citation for published version (APA):
Vægter, H. B., Handberg, G., & Kent, P. (2018). Brief Psychological Screening Questions Can be Useful for Ruling Out Psychological Conditions in Patients with Chronic Pain. *The Clinical Journal of Pain*, 34(2), 113–121. <https://doi.org/10.1097/AJP.0000000000000514>

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use

This work is brought to you by the University of Southern Denmark.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk

**BRIEF PSYCHOLOGICAL SCREENING QUESTIONS CAN BE USEFUL FOR RULING
OUT PSYCHOLOGICAL CONDITIONS IN PATIENTS WITH CHRONIC PAIN**

Henrik Bjarke Vaegter, PhD^{ab}, Gitte Handberg, MD^a,

Peter Kent, PhD^{c,d}

^a *Pain Research Group, Pain Center South, Odense University Hospital, Odense, Denmark*

^b *Department of Clinical Research, Faculty of Health Sciences, University of Southern Denmark*

^c *School of Physiotherapy and Exercise Science, Curtin University, Bentley, Perth, Western
Australia*

^d *Department of Sports Science and Clinical Biomechanics, Faculty of Health Sciences, University
of Southern Denmark*

Original investigation for: Clinical Journal of Pain

Corresponding author:

Henrik Bjarke Vaegter, PhD

Pain Research Group, Pain Center South, University Hospital Odense

Heden 9, Indgang 201, DK – 5000 Odense C

Tel.: +45 65413869; fax: +45 65415064.

E-mail: henrik.bjarke.vaegter@rsyd.dk

Conflicts of Interest and Source of Funding: There are no actual or potential conflicts of interest for any of the authors.

ABSTRACT

Objectives: Psychological symptoms are highly prevalent in chronic pain patients. Timely and accurate identification may enable individualized treatment and improve outcomes. The aims of this study were to (i) investigate the concurrent validity of brief psychological screening questions assessing anxiety, fear of movement, stress, pain catastrophization, and depression in chronic pain patients, and (ii) to determine screening question cut-points at which the likely probability of having these psychological states was less than 10%.

Methods: Responses to one-item or two-item screening questions within each of these five psychological constructs were compared with those of validated full-length questionnaires in 894 patients with diverse chronic pain conditions.

Results: Compared with scores from full-length questionnaires, brief screening question scores had correlations between 0.54 and 0.66 and AUC between 0.79 and 0.83. At the dichotomized threshold scores that we chose, the post-test probability after a negative test result ranged from 6.5% to 8.6% for all these psychological constructs, except fear of movement. The pre-test probability was so high (70%) for fear of movement that no threshold resulted in a post-test probability (negative test result) that was below 10%.

Discussion: Use of these screening tests and scoring thresholds would have correctly identified that between 38.5% and 60.5% of the sample were unlikely to have these psychological states (true negatives), with a false negative rate between 3.4% and 5.3%. This would allow clinicians to focus on whether there are other patient attributes in those patients requiring more thorough investigation using comprehensive validated questionnaires or structured clinical interviews.

Keywords: Chronic pain, assessment, psychology, questionnaires

1. INTRODUCTION

Chronic pain is one of the most disabling health problems facing the world today (1) and the costs associated with treatment, sick leave and early retirement are comparable to the costs of diabetes and cancer combined. Moreover, with the high prevalence of chronic pain in elderly people (2) and the growing proportion of elderly people across the developed world (3), the burden of chronic pain is expected to increase. Psychological co-morbidities or symptoms are highly prevalent in patients with chronic pain (4, 5), are generally associated with more severe pain (6, 7) and higher levels of disability (8), and have been accepted as having an important role in the experience and consequences of pain. Specific psychological constructs such as anxiety, fear of movement, stress, catastrophization, and depression, (9-11) have been associated with the development of chronic pain (12), use of analgesics (13), health care utilization (14), and response to treatment (15-17) in a variety of pain conditions. There is also a bi-directional relationship between chronic pain and mental health (18). For these reasons, timely and accurate knowledge of the presence or absence of these psychological factors may improve patient's understanding of their own pain condition, better inform clinicians about their individual patients, enable appropriate and individualized treatment, and potentially improve health outcomes.

Several full-length psychological screening questionnaires exist, however due the length, time burden and scoring complexity of validated psychological instruments, these factors are not routinely assessed in clinical practice (19). Formal screening has shown better accuracy than clinicians' informal judgments (20, 21) and, in the absence of some form of screening, most psychological comorbidities may be undiagnosed and untreated (21). Routine screening of psychological factors may be facilitated through development of brief psychological screening questions with acceptable concurrent validity. For example, the concurrent validity of a simple question screening for depression was investigated by Reme and colleagues in a sample of 595 patients with diverse chronic pain conditions, and good sensitivity and specificity was demonstrated

when compared with a full-length validated questionnaire (22). Similarly, satisfactory concurrent validity of some brief screening questions for depression and anxiety was reported in a recent systematic review (23) but the authors suggested that studies of brief screening questions for other psychological constructs were needed. Recently, Kent and colleagues (24) compared brief screening questions for anxiety, depression, catastrophization, social isolation and fear of movement with scores from full-length questionnaires in patients with chronic back pain. Anxiety, depression, and social isolation demonstrated sensitivity of 70.4% to 82.5%, specificity of 75.0% to 94.7%, and catastrophization and fear of movement demonstrated sensitivity of 78.3% to 88.0%, and specificity of 91.0% to 96.2%, indicating that brief psychological screening questions may be useful in the initial assessment of patients with chronic back pain. However, no studies have assessed the concurrent validity for psychological constructs other than depression in large samples of diverse chronic pain conditions and therefore the generalizability of those findings from chronic back pain to other pain populations is unknown.

Often the intended purpose of screening questions is to alert the clinician to there being an elevated probability of a health condition being present. But the clinical presentation of people with chronic pain is often complex and the diagnosis of psychological conditions requires more comprehensive information and clinical judgement. Therefore, some have argued that the purpose of screening is to determine whether more comprehensive assessment is required (25). In recognition of this, it might be that screening questions better assist the efficiency of the clinical encounter if they are used to determine that there is a low probability of having the psychological state, so that the clinician can then move on to focus on whether there are other attributes in those patients that require more thorough investigation.

Therefore, the aims of this study were (i) to quantify the concurrent validity of brief psychological screening questions assessing five psychological constructs (anxiety, fear of movement, stress, pain catastrophization, and depression) across diverse chronic pain conditions,

and (ii) to determine screening question cut-points at which the likely probability of having these psychological states was less than 10%.

2. MATERIALS AND METHODS

2.1 Method summary

This cross-sectional study compared patient responses on one- or two-item screening questions with their responses on longer validated questionnaires for each psychological construct, thereby describing the concurrent validity of the screening questions. The criteria used for assessing that validity were the correlation of the scores, the area under the curve seen on a Receiver Operating Characteristic (ROC) curve analysis, and performance characteristics derived from 2 x 2 contingency tables where scores on both the screening question and validated questionnaire were dichotomized using clinically appropriate cut-points.

2.2 Context

The context for this study was the establishment of a clinical pain registry (PainData, Denmark) for patients referred for assessment and treatment of their chronic pain at a secondary pain clinic in Denmark (Pain Center South, Odense University Hospital). After referral to this multidisciplinary pain clinic but prior to their initial consultation, all patients were invited to answer questions about their clinical characteristics and adaptations to pain via a web-based questionnaire system (PainData). Questionnaires were completed at home prior to the first consultation. None of the questions were mandatory. In addition to basic demographics, these data included full-length validated questionnaires for measures of pain intensity, activity limitation, work participation, health-related quality of life, and several psychological constructs, which collectively involved more than 150 questions. The data in this study were from the 894 patients between 1 May 2015 and 30 March 2016 who gave consent for their completed data to be used for research purposes, and as there were 1,076 initial consultations during this period, this was an 83.1% sampling rate.

The demographic and clinical characteristics of the participants are described in Table 1. Their mean age was 49.8 ± 14.5 years, 66.6% were women, 51.0% had some form of post-secondary school education, 70.3% were married or living with someone, and 31.5% were currently employed. Participants' pain conditions were diverse, with the largest single category (22.5%) being chronic widespread pain. Their median duration of pain was 6 years and most had pain every day (91.3%) with an average pain intensity of 7 (0 to 10 scale).

Written informed consent (electronic signature) was obtained from all patients included in this study, the Danish Data Protection Agency approved the data collection (ref. no. 14/44319), and the conduct of this study complied with the Declaration of Helsinki. As treatment was not affected by participation in the study, under Danish law, this study did not need ethics approval (*Act on Research Ethics Review of Health Research Projects, October 2013, Section 14.2*) (26).

2.3 Screening Questions

For four of the five constructs in this study (anxiety, fear of movement, catastrophization and depression), the one- or two-item screening questions investigated were those items that performed best in a previous study of the concurrent validity of psychological screening questions in people with chronic back pain (24). As we were unable to identify a validated screening question for the remaining construct (stress), we constructed a single-item screening question, as our interest was in brevity and ease of comprehension. All screening questions had response options on a 0 to 10 scale.

For anxiety, we tested a single item: '*Do you feel anxious?*' ('Not at all'=0 to 'Quite anxious'=10). This item had been devised in the back pain study (24).

For fear of movement, we tested two items because in the back pain study the average of scores on both of these two items had performed best: (i) '*Physical activity might damage me*', and (ii) '*I should not do physical activities which (might) make my pain worse*' ('Completely disagree'=0 to 'Completely agree'=10). Both these items were from the Fear Avoidance Beliefs

Questionnaire (physical activities subscale) (27), but the first item we adapted to a diverse pain sample by changing ‘... might damage my back’ to ‘...might damage me’.

For stress we devised and tested the question ‘*Do you feel stressed?*’ (‘Not at all stressed’=0 to ‘Very stressed’=10). The format of this question was simply adapted from that used to screen for anxiety.

For catastrophization, we tested two items because in the back pain study the average of scores on both of these two items had performed best: (i) ‘*When I feel the pain, it is terrible and I feel that it’s never going to get better*’, and (ii) ‘*When I feel pain, I feel that I can’t stand it anymore*’ (‘Never do that’=0 to ‘Always do that’=10). Both these items were from the Coping Strategies Questionnaire (catastrophization subscale) (28).

For depression, we also tested two items: ‘*During the past month, have you often felt sad, depressed or had a sense of hopelessness?*’ and (ii) ‘*During the past month, have you felt bothered by little interest or pleasure in to do something?*’ (‘Never’=0 to ‘All the time’=10). These items are the two Prime 1000 depression screening questions (21), with one for each of the two dimensions of depression: depressed mood and lack of pleasure (anhedonia).

There was no capacity for patients to enter comments about the screening questions or if report assistance was used when completing them. However, patients could leave any screening question blank and the non-response rate ranged from 0.1% to 2.0% (average 0.7%). Because the non-response rate was so low, we simply used the available data and did not impute any data.

2.4 Reference Standard Questionnaires

The reference standard questionnaires and the thresholds that were used to dichotomize their sumscores are itemized in Table 2. Wherever possible, we used population-based thresholds that had been published or had been recommended by the developers of the reference standard questionnaire. In the case of the Perceived Stress Scale, no validated threshold was available, so we used the distribution-based threshold of the mean plus 1 standard deviation (SD). This threshold

identifies the highest scoring 16% of a sample and is often used in the absence of normative data (24, 29, 30).

2.5 Measures of Concurrent Validity

As a measure of the how strongly the scores on the screening questions and sumscores on the reference standard questionnaires were associated, we calculated Pearson Correlation Coefficients. We also calculated the area under the curve (AUC) statistic from Receiver Operating Curve (ROC) analysis as a performance measure (an index of the classification error rate) of the screening questions relative to dichotomized scores on the reference standard questionnaires.

Additional performance characteristics (sensitivity, specificity, post-test probability if the test was positive or negative) (24) were derived from contingency tables for every possible dichotomization threshold for scores on each screening question. We initially sought the dichotomization threshold that best balanced the sensitivity and specificity for each screening question but favored the threshold that resulted in a post-test probability (after a negative test result) of less than 10%. To provide a direct measure of how many patients would have been correctly 'screened out' at that threshold, the number of 'true negatives' (people classified by the reference standard questionnaire as not having the psychological state who were also classified by the screening question as not having the psychological state) and 'false negatives' (people classified by the reference standard questionnaire as not having the psychological state who were also classified by the screening question as actually having the psychological state) were reported, and the proportion of the whole cohort those numbers represented. We highlighted the performance characteristics for the chosen thresholds but also reported the results for all thresholds because other clinical circumstances may require a threshold choice that favors greater sensitivity or specificity. For the chosen thresholds, we also calculated the positive and negative likelihood ratios to provide additional measures of test performance.

We did not report the overall accuracy statistic, due to evidence that it is problematic as a measure of validity when the sample prevalence deviates from 50% (31). Some authors have argued that other statistical characteristics we did describe (sensitivity, specificity and likelihood ratios) are prevalence-independent (32, 33) but other authors have presented evidence that they are not (34, 35). All statistical analyses were performed using STATA version 14.1 (STATA Corp., College Station, Tx 77845, USA) and Excel 2011 version 14.5 (Microsoft Corp, Redmond, Washington, USA).

3. RESULTS

3.1 Concurrent Validity

For anxiety, the correlation between scores on the single-item screening question '*Do you feel anxious?*' and the Generalized Anxiety Disorder Scale sumscore was 0.64. The performance measures from the contingency table are shown in Table 3. For the threshold that best balanced the sensitivity and specificity (5 or more), the sensitivity was 80.8% (CI 74.2% to 86.1%), specificity 72.2% (CI 68.6% to 75.6%), positive likelihood ratio 2.91 (CI 2.51 to 3.36) and negative likelihood ratio 0.26 (CI 0.19 to 0.36). The pre-test probability (prevalence) of anxiety on the reference standard was 20.8% (CI 18.0% to 23.6%) and, at the screening test threshold of 5 or more, the post-test probability following a positive screening test result was 43.3% (CI 37.8% to 48.8%) and following a negative screening test result was 6.5% (CI 4.3% to 8.7%). The AUC was 0.83. (See Figure, Supplemental Digital Content 1, <http://links.lww.com/CJP/A434> for the ROC curve).

For fear of movement, the correlation between scores on the single-item screening question '*Physical activity might damage me*' and the sumscore from the Tampa Scale for Kinesiophobia was 0.62 and the AUC was 0.81 (See Figure, Supplemental Digital Content 2, <http://links.lww.com/CJP/A435> for the ROC curve). For the item '*I should not do physical activities which (might) make my pain worse*', the correlation was 0.55 and the AUC was 0.77. Combining the results of both screening questions, by averaging their scores, resulted in only a

small improvement in performance: correlation 0.66 and AUC 0.83. Therefore, we chose the option of only using the item *'Physical activity might damage me'* because its content seemed more appropriate to a chronic pain population for whom some do experience flares of pain following physical activity, and also because the performance gain by using both screening questions was marginal compared to the additional burden of a second question. For the threshold score on *'Physical activity might damage me'* that best balanced the sensitivity and specificity (5 or more), the sensitivity was 75.7% (CI 71.8% to 79.1%), specificity 74.4% (CI 68.4% to 79.7%), positive likelihood ratio 2.96 (CI 2.36 to 3.72) and negative likelihood ratio 0.33 (CI 0.28 to 0.39) (Table 4). The pre-test probability of fear of movement on the reference standard was 70.0% (CI 66.7% to 73.3%) and, at the screening test threshold of 5 or more, the post-test probability following a positive screening test result was 87.4% (CI 84.3% to 90.4%) and following a negative screening test result was 43.3% (CI 37.7% to 48.9%). Because the pre-test probability was so high, no threshold on this single-item screening question resulted in a post-test probability (negative test) that was below 10%, as even a threshold of 1 or more resulted in a post-test probability of 27.3%.

For stress, the correlation was 0.54 between scores on the single-item screening question *'Do you feel stressed?'* and the Perceived Stress Scale sumscore. For the threshold that best balanced the sensitivity and specificity (7 or more), the sensitivity was 71.2% (CI 62.5% to 78.6%), specificity 70.6% (CI 67.1% to 74.0%), positive likelihood ratio 2.42 (CI 2.06 to 2.86) and negative likelihood ratio 0.41 (CI 0.31 to 0.54) (Table 5). The pre-test probability of stress on the reference standard was 15.0% (CI 12.5% to 17.5%) and, at the screening test threshold of 7 or more, the post-test probability following a positive screening test result was 30.0% (CI 24.6% to 35.4%) and following a negative screening test result was 6.7% (CI 4.5% to 8.9%). The AUC was 0.79. (See Figure, Supplemental Digital Content 3, <http://links.lww.com/CJP/A436> for the ROC curve).

For catastrophization, the correlation between scores on the single-item screening question *'When I feel the pain, it is terrible and I feel that it's never going to get better'* and the Pain

Catastrophization Scale sumscore was 0.66 and the AUC was 0.85 (See Figure, Supplemental Digital Content 4, <http://links.lww.com/CJP/A437> for the ROC curve). For the item *'When I feel pain, I feel that I can't stand it anymore'*, the correlation was 0.74 and the AUC was 0.87. Combining the results of both screening questions, by averaging their scores, resulted in only a negligible improvement in performance: correlation 0.76 and AUC 0.89. As AUC was very close for both these items, and averaging their results resulting in a negligible increase in performance, we opted to use only the first item. The reason for choosing this item, was because *'When I feel the pain, it is terrible and I feel that it's never going to get better'* appeared to have higher face validity for catastrophization than *'When I feel pain, I feel that I can't stand it anymore'* which appeared to have higher face validity for coping. The threshold score on *'When I feel the pain, it is terrible and I feel that it's never going to get better'* that best balanced the sensitivity and specificity (8 or more) did not result in a post-test probability following a negative screening test result that was <10%. Therefore, we opted for a threshold of 7 or more, for which the sensitivity was 90.7% (CI 86.8% to 93.5%), specificity 60.9% (CI 56.5% to 65.1%), positive likelihood 2.32 (CI 2.07 to 2.60) and negative likelihood ratio 0.15 (CI 0.11 to 0.22) (Table 6). The pre-test probability of catastrophization on the reference standard was 36.7% (CI 33.4% to 40.1%) and, at the screening test threshold of 7 or more, the post-test probability following a positive screening test result was 57.4% (CI 52.9% to 61.9%) and following a negative screening test result was 8.1% (5.2% to 11.1%).

For depression, the correlation between the single-item screening question scores on *'During the past month, have you often felt sad, depressed or had a sense of hopelessness?'* and the Patient Health Questionnaire 9-item sumscore was 0.57 and AUC was 0.81 (See Figure and Table, Supplemental Digital Content 5, <http://links.lww.com/CJP/A438> and 6, <http://links.lww.com/CJP/A439> for the ROC curve and performance characteristics). The correlation was also 0.56 for the item *'During the past month, have you felt bothered by little*

interest or pleasure in to do something?’ and the AUC was 0.82 (See Figure and Table, Supplemental Digital Content 7, <http://links.lww.com/CJP/A440> and 8, <http://links.lww.com/CJP/A441> for the ROC curve and performance characteristics). As depression has two dimensions that are each assessed by one of these screening questions, we chose the thresholds on each of them (6 or more for each) that best balanced the sensitivity and specificity on a new dichotomized variable of patients who were positive on both screening questions or not. The AUC for this new composite variable was 0.75 (See Figure, Supplemental Digital Content 9, <http://links.lww.com/CJP/A442> for the ROC curve). The sensitivity was 77.2% (CI 71.1% to 83.4%), specificity 73.1% (CI 69.5% to 76.6%), positive likelihood ratio 2.87 (CI 2.46 to 3.35) and negative likelihood ratio 0.31 (CI 0.24 to 0.41). The pre-test probability of depression on the reference standard was 23.3% (CI 20.3% to 26.3%), the post-test probability following a positive screening test result was 46.5% (CI 40.8% to 51.2%) and post-test probability following a negative screening test result was 8.6% (CI 6.1% to 11.2%) (Table 7).

For completeness, the prevalence in each cell of the 2 x 10 contingency tables for every screening question we tested is displayed as a table in the Supplemental Digital Content (See Tables, Supplemental Digital Content 10, <http://links.lww.com/CJP/A443>). These data will allow the calculation of more detailed performance characteristics at other scoring thresholds than those we chose, which may be useful in other clinical circumstances and also may be useful for meta-analysis with other studies.

4. DISCUSSION

4.1 Summary of findings

The first aim of this study was to investigate the concurrent validity of one-item or two-item screening questions for these five psychological constructs in a large sample of patients with diverse chronic pain conditions. Compared with sumscores from the relevant full-length questionnaire, the 0 to 10 scores on these brief screening questions demonstrated correlations between 0.54 and 0.66

and AUC between 0.79 and 0.83. This is the first study to investigate this in a sample with diverse chronic pain conditions.

The second aim was to determine screening question cut-points at which the likely probability of having these psychological states was less than 10%. At the dichotomized threshold scores that we chose for each screening question, the post-test probability after a negative test result ranged from 6.5% to 8.6% for all these psychological constructs, except fear of movement. Because the prevalence (pre-test probability) of fear of movement was so high (70.0%) in this sample, no threshold on this single-item screening question resulted in a post-test probability (after a negative test) that was below 10%. For the other constructs (anxiety, stress, catastrophization and depression), the chosen thresholds would have resulted in excluding ‘true negatives’ (people classified by the reference standard questionnaire as not having the psychological state who were also classified by the screening question as not having the psychological state) that represented between 38.5% and 60.5% of the whole sample, with a false negative rate between 3.4% and 5.3%. So broadly speaking, use of these screening test thresholds would have efficiently and correctly identified between a third and two thirds of the total patient sample as being unlikely to have these psychological states, and incorrectly identified one thirtieth to one twentieth of the sample as being unlikely when they were actually likely to have these psychological states. Clinicians who value the breadth of clinical options available to rule in these psychological states (comprehensive validated questionnaires, structured clinical interviews, expert diagnostic opinion) may welcome the capacity to use these screening questions to efficiently identify patients unlikely to require these more time consuming and detailed clinical investigations. Also, in our context of a tertiary pain service where we assess many aspect of the patient’s presentation using a comprehensive battery of on-line questionnaires, use of screening questions to rule out conditions would reduce the length of the survey and responder burden.

4.2 Test performance of dichotomized thresholds relative to previous studies

For anxiety, the sensitivity (80.8%), and specificity (72.2%) for the one-item screening question used in this study with a cut-point of 5 and above were similar to that previously reported by Kent et al. for the same screening question with a cut-point of 2 and above in 179 patients with low back pain in a secondary care setting in Denmark (82.5% and 75.0%) and with a cut-point of 5 and above in 191 patients with low back pain in a primary care in Australia (80.0% and 78.3%) (24). The performance of this anxiety question and its response options and scoring technique was better than an alternative screening question investigated in 1,000 patients seeking medical care for a range of physical conditions in USA (94% and 53%) (21).

For fear of movement, the sensitivity (75.7%), and specificity (74.5%) for the one-item screening question found relevant in this study, with a cut-point of 5 and above, were less than those reported by Kent et al. for the sum of two brief questions within the same construct in patients with low back pain in Denmark with a cut-point of 7 and above (84.4% and 91.1%) and in Australia with a cut-point of 8.5 and above (86.7% and 93.4%) (24). The difference in performance may be due to differences in the included pain populations and potential differences in the pain-related beliefs within these pain populations. For example, although many forms of movement are accompanied by pain for chronic pain patients, fear of movement appears to be more prevalent in patients with low back pain (36) than in patients with widespread pain conditions (37).

For stress, a sensitivity (71.2%), and specificity (70.6%) was found for the single-item screening question used in this study, with a cut-point of 7 and above. To our knowledge, the concurrent validity of one-item or two-item brief screening questions for stress has not previously been investigated.

For catastrophization, a sensitivity (90.7%), and specificity (60.9%) with a cut-point of 7 and above was observed for the one-item screening question chosen in this study. This was less balanced than reported previously by Kent et al. for the same construct in patients with low back

pain in Denmark with a cut-point of 7 and above (78.3% and 96.2%) and in Australia with a cut-point of 4 and above (88.0% and 91.0%) (24).

For depression, the sensitivity (77.2%), and specificity (73.1%) for a positive threshold score on both PRIME-MD screening questions (with a cut-point of 6 and above for both) were more balanced than those previously reported by Kent et al. for the same screening question with a cut-point of 7 and above in patients with low back pain in Denmark (70.4% and 89.0%) and with a cut-point of 8 and above in Australia (73.3% and 94.7%) (24). The performance of these depression questions and their response options and scoring technique was also more balanced than those reported for an alternative screening question investigated in 1,000 patients seeking medical care for a range of physical conditions in USA (69% and 82%) (21).

4.3 Clinical Implications

A failure to recognize psychological factors may diminish the therapeutic effectiveness of pain management strategies and result in inefficient use of healthcare services. The use of brief screening questions has a number of potential benefits, including (i) early and easily performed psychological screening of people as a resource-efficient triage to inform clinical evaluation, and (ii) as an indication for referral to specialty pain management for further psychological assessment. However, the use of brief screening questions may also have potential pitfalls, including a loss of accuracy compared with more comprehensive questionnaires or diagnostic interviews. Therefore, we adopted the strategy of using screening questions to determine whether there is a low probability of having a psychological state, so that clinician could focus on whether there are other patient attributes in those patients requiring more thorough investigation. For anxiety, stress, pain catastrophization, and depression, use of these screening test thresholds would have correctly identified that, depending on the psychological factor, between a 22.3% and 57.2% of the total patient sample were unlikely to have these psychological states (true negatives), with a false negative rate between 3.4% and 5.3%.

Recognizing that a balance between efficiency and diagnostic accuracy is highly important if such screening questions are to be useful in a busy clinical setting, we also reported the performance for all dichotomization thresholds because other clinical circumstances may require a threshold choice that favors greater sensitivity or specificity.

4.4 Strengths and limitations

This study has a number of strengths. It is a broad validation study (46), where the performance characteristics these screening questions were examined in a different population (general chronic pain rather than just low back pain) and for a different purpose (screening out the presence of a psychological state rather than screening it in). In addition, recent evidence indicates that 400 to 600 participants are required for relatively stable performance estimates derived from 2x2 contingency tables (47). Therefore, while the earlier study involved 5 samples ranging from 91 to 353 (mean 221), used in an iterative translation and validation approach, the current study included 894 patients and therefore these estimates may be more stable and less sample-specific.

This study also has a number of weaknesses that could influence interpretation of the results. First of all, the use of questionnaires as the reference standard for the psychological constructs represents a potential limitation of the study, and future studies could investigate brief screening questions compared with structured diagnostic interviews as the reference standard. Secondly, the generalizability of these findings to other clinical settings is currently unknown and may be influenced by the variation in severity and prevalence of these psychological constructs between different clinical settings and cultural contexts. Thirdly, despite the relatively large sample size, the distribution of scores on some constructs resulted in relatively small numbers in some contingency table cells. Fourthly, the sampling rate was 83.1%, and so the estimates may potentially contain some selection bias. Patients who did not complete the questionnaires did not give informed consent for any of their data to be used for research and so comparisons with the included sample were not possible. Moreover, although the other screening questions have previously been tested for

their comprehensibility (24), we did not assess the comprehensibility of the novel screening question assessing stress. Nor did we assess if any patients gave erroneous responses to any of the other previously validated questions, either due to a lack of comprehension or due to being distracted by pain. Finally, the cross-sectional study design does not give any insights into test-retest reliability or the predictive validity of these brief screening questionnaires, nor their ability to detect clinically relevant changes, which should be established in relevant pain populations.

4.5 Conclusion

The concurrent validity was described for one-item or two-item screening questions assessing five psychological constructs (anxiety, fear of movement, stress, pain catastrophization, and depression) across diverse chronic pain conditions. For all but fear of movement, we identified threshold scores with acceptable sensitivity and specificity, at which the likely probability of having the psychological state was less than 10%. These results could have immediate practical application in routine screening of psychological constructs enabling timely and individualized treatment. In addition, clinicians and researchers now have comprehensive data to inform whether each of these screening questions are potentially useful to include in routine assessment procedures of their relevant pain populations.

Conflicts of Interest and Source of Funding: There are no actual or potential conflicts of interest for any of the authors.

References:

1. Ma VY, Chan L, Carruthers KJ. Incidence, prevalence, costs, and impact on disability of common conditions requiring rehabilitation in the United States: stroke, spinal cord injury, traumatic brain injury, multiple sclerosis, osteoarthritis, rheumatoid arthritis, limb loss, and back pain. *Arch Phys Med Rehabil.* 2014;95:986-95.
2. Helme RD, Gibson SJ. The epidemiology of pain in elderly people. *Clin Geriatr Med.* 2001;17(3):417-31.
3. Bongaarts J. Human population growth and the demographic transition. *Philos Trans R Soc Lond B Biol Sci.* 2009;364(1532):2985-90.
4. Manchikanti L, Fellows B, Pampati V, Beyer C, Damron K, Barnhill RC. Comparison of psychological status of chronic pain patients and the general population. *Pain Physician.* 2002;5(1):40-8.
5. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *Eur J.* 2006;10(4):287-333.
6. Ho PT, Li CF, Ng YK, Tsui SL, Ng KF. Prevalence of and factors associated with psychiatric morbidity in chronic pain patients. *J Psychosom Res.* 2011;70(6):541-7.
7. van der Windt D, Croft P, Penninx B. Neck and upper limb pain: more pain is associated with psychological distress and consultation rate in primary care. *J Rheumatol.* 2002;29(3):564-9.
8. Hung CI, Liu CY, Fu TS. Depression: An important factor associated with disability among patients with chronic low back pain. *Int J Psychiatry Med.* 2015;49(3):187-98.
9. Ramond A, Bouton C, Richard I, Roquelaure Y, Baufreton C, Legrand E, et al. Psychosocial risk factors for chronic low back pain in primary care--a systematic review. *Fam Pract.* 2011;28(1):12-21.

10. van der Weide WE, Verbeek JH, Salle HJ, van Dijk FJ. Prognostic factors for chronic disability from acute low-back pain in occupational health care. *Scand J Work Environ Health*. 1999;25(1):50-6.
11. Fransen M, Woodward M, Norton R, Coggan C, Dawe M, Sheridan N. Risk factors associated with the transition from acute to chronic occupational back pain. *Spine (Phila Pa 1976)*. 2002;27(1):92-8.
12. Sullivan M, Tanzer M, Stanish W, Fallaha M, Keefe FJ, Simmonds M, et al. Psychological determinants of problematic outcomes following Total Knee Arthroplasty. *Pain*. 2009;143(1-2):123-9.
13. Singh JA, Lewallen DG. Predictors of pain medication use for arthroplasty pain after revision total knee arthroplasty. *Rheumatology (Oxford)*. 2014;53(10):1752-8.
14. Engel CC, von Korff M, Katon WJ. Back pain in primary care: predictors of high health-care costs. *Pain*. 1996;65(2-3):197-204.
15. George SZ, Zeppieri G, Jr., Cere AL, Cere MR, Borut MS, Hodges MJ, et al. A randomized trial of behavioral physical therapy interventions for acute and sub-acute low back pain (NCT00373867). *Pain*. 2008;140(1):145-57.
16. Klaber Moffett JA, Carr J, Howarth E. High fear-avoiders of physical activity benefit from an exercise program for patients with back pain. *Spine (Phila Pa 1976)*. 2004;29(11):1167-72; discussion 73.
17. Sullivan MJ, Adams H, Rhodenizer T, Stanish WD. A psychosocial risk factor--targeted intervention for the prevention of chronic pain and disability following whiplash injury. *Phys Ther*. 2006;86(1):8-18.
18. Hooten WM. Chronic Pain and Mental Health Disorders: Shared Neural Mechanisms, Epidemiology, and Treatment. *Mayo Clin Proc*. 2016;91(7):955-70.

19. Kent PM, Keating JL, Taylor NF. Primary care clinicians use variable methods to assess acute nonspecific low back pain and usually focus on impairments. *Man Ther.* 2009;14(1):88-100.
20. Haggman S, Maher CG, Refshauge KM. Screening for symptoms of depression by physical therapists managing low back pain. *Phys Ther.* 2004;84(12):1157-66.
21. Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV, 3rd, Hahn SR, 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.
22. Reme SE, Eriksen HR. Is one question enough to screen for depression? *Scand J Public Health.* 2010;38(6):618-24.
23. Mirkhil S, Kent PM. The diagnostic accuracy of brief screening questions for psychosocial risk factors of poor outcome from an episode of pain: A systematic review. *Clin J Pain.* 2009;25(4):340-8.
24. Kent P, Mirkhil S, Keating J, Buchbinder R, Manniche C, Albert HB. The concurrent validity of brief screening questions for anxiety, depression, social isolation, catastrophization, and fear of movement in people with low back pain. *Clin J Pain.* 2014;30(6):479-89.
25. Center for Substance Abuse Treatment. Treatment Improvement Protocol (TIP) Series, No. 51. Substance Abuse Treatment. Rockville (MD): Substance Abuse and Mental Health Services Administration (US) 2009.
26. Danish National Committee on Biomedical Research Ethics. Act on research ethics review of health research projects [cited 2015 27 May]. Available from: <http://www.cvk.sum.dk/English/actonbiomedicalresearch.aspx>.
27. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain.* 1993;52(2):157-68.

28. Rosenstiel AK, Keefe FJ. The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. *Pain*. 1983;17(1):33-44.
29. Wald NJ, Simmonds M, Morris JK. Screening for future cardiovascular disease using age alone compared with multiple risk factors and age. *PLoS One*. 2011;6(5):e18742.
30. Mudgal J, Guimaraes Borges GL, Díaz Montiel JC, Flores Y, Salmerón J. Depression among health workers: The role of social characteristics, work stress, and chronic diseases. *Salud Mental*. 2006;29(5):1-8.
31. Alberg AJ, Park JW, Hager BW, Brock MV, Diener-West M. The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. *J Gen Intern Med*. 2004;19(5 Pt 1):460-5.
32. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Phys Ther*. 2001;81(9):1546-64.
33. Sackett DL, Haynes RB, Guyatt GH. *Clinical Epidemiology. A basic science for clinical medicine*. Boston, MA, USA: Little, Brown and Company; 1992.
34. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med*. 1997;16(9):981-91.
35. Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology*. 1997;8(1):12-7.
36. Thomas EN, Pers YM, Mercier G, Cambiere JP, Frasson N, Ster F, et al. The importance of fear, beliefs, catastrophizing and kinesiophobia in chronic low back pain rehabilitation. *Ann Phys Rehabil Med*. 2010;53(1):3-14.
37. de Rooij A, Steultjens MP, Siemonsma PC, Vollebregt JA, Roorda LD, Beuving W, et al. Overlap of cognitive concepts in chronic widespread pain: an exploratory study. *BMC Musculoskeletal Disord*. 2011;12:218.

38. 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.
39. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Lowe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med.* 2007;146(5):317-25.
40. Vlaeyen JW, Kole-Snijders AM, Boeren RG, van Eek H. Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. *Pain.* 1995;62(3):363-72.
41. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav.* 1983;24(4):385-96.
42. Cohen S, Williamson G. Perceived Stress in a Probability Sample of the United States. In: Spacapan S, Oskamp S, editors. *The Social Psychology of Health.* Newbury Park, CA: Sage; 1988.
43. Sullivan MJ, Bishop S, Pivik J. The Pain Catastrophizing Scale: Development and validation. *Psychol Assess.* 1995;7:524-32.
44. Sullivan MJ. The Pain Catastrophizing Scale - User manual. [cited 2016 5 May]. Available from: http://sullivan-painresearch.mcgill.ca/pdf/pcs/PCSManual_English.pdf.
45. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9):606-13.
46. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS: Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. *JAMA* 2000, 284:79-84.
47. Kent P, Boyle E, Keating JL, Albert AB, Hartvigsen J. 400 or more participants needed for stable contingency table estimates of clinical prediction rule performance. *J Clin Epidemiol.* 2017; 82:137-148.

List of Supplemental Digital Content

Supplemental Digital Content 1: Figure SDC1

Supplemental Digital Content 2: Figure SDC2

Supplemental Digital Content 3: Figure SDC3

Supplemental Digital Content 4: Figure SDC4

Supplemental Digital Content 5: Table SCD1

Supplemental Digital Content 6: Figure SDC5

Supplemental Digital Content 7: Table SCD2

Supplemental Digital Content 8: Figure SDC6

Supplemental Digital Content 9: Figure SDC7

Supplemental Digital Content 10: Table SDC3.doc

ACCEPTED

Table 1: Characteristics of cohort (n=894)

Characteristic	Measurement	
Age	Years (mean)	49.8 (SD 14.5) (full range 18 to 97)
Sex	Female	66.6%
Height	Metres (mean)	1.71 (SD 0.15)
Weight	Kilograms	81.8 (SD 27.6)
Pain diagnosis	Chronic widespread pain	22.5%
	Low back pain	9.5%
	Neck pain	4.6%
	Extremity pain	9.3%
	Sciatica	8.9%
	Neck brachialgia	4.4%
	Multiple spinal areas	16.6%
	Abdominal pain	3.2%
	Headache	2.5%
	Spinal and extremity pain (not Chronic Widespread Pain)	18.5%
	Average pain intensity	NRS (0 to 10) (mean)
Resting pain intensity	NRS (0 to 10) (mean)	6.2 (SD 2.2)
Strongest pain intensity	NRS (0 to 10) (mean)	8.2 (SD 1.4)
Pain duration	Years (median)	6 (IQR 2 to 13) (full range 0 to 67)
Pain frequency	1 = Less than 1 day/week	0.1%
	2 = 1 to 2 days per week	0.7%
	3 = 3 to 4 days per week	2.4%
	4 = 5 to 6 days per week	5.5%
	5 = everyday	91.3%
Exposed to trauma	Single item screening question	56.7%
Pain related to trauma	Of those exposed to trauma: (single item screening question)	46.5%
	Overall in sample:	25.2%
	Anxiety	GAD-7 (0 to 21 scale) (mean)
Fear of movement	TSK (17 to 68 scale) (mean)	41.3 (SD 8.0)
Stress	PSS (0 to 40 scale) (mean)	20.6 (SD 4.0)
Catastrophization	PCS (0 to 52 scale) (mean)	25.7 (SD 11.2)
Depression	PHQ-9 (0 to 27 scale) (mean)	8.0 (SD 6.2)
General health	EuroQOL Health thermometer (0 to 100 scale) (mean)	40.3 (SD 22.0)
Making insurance claim	Self-report	24.1%
Education	High school or trade qualification	57.9%
	Bachelor degree	31.3%
	Higher degree	4.5%
	Other	6.3%
Marital status	Married or living with someone	70.3%
Work participation	Full work income	22.2%
	Partial work income	9.3%
	Sick leave	22.0%
	Unemployed	15.1%
	Pension	31.4%

SD = Standard Deviation, IQR = InterQuartile Range, NRS = Numerical Rating Scale, GAD-7 = Generalised Anxiety Disorder Scale, TSK = Tampa Scale for Kinesiophobia, PSS = Perceived Stress Scale, PCS = Pain Catastrophization Scale, PHQ-9 = Patient Health Questionnaire 9-item

Table 2: Reference standard questionnaires and the criteria used for cut-points to dichotomize their sum scores

Construct	Reference standard	Cut-points
Anxiety	Generalized Anxiety Disorder Scale (0 to 21 scale) (38)	Primary care threshold for generalized anxiety disorder ≥ 10 (39) derived in a sample of general medical primary care patients
Fear of movement	Tampa Scale for Kinesiophobia (17 to 68 scale) (40)	Recommended threshold for high pain-related fear of movement ≥ 38 (40) derived in a sample of chronic pain patients
Stress	Perceived Stress Scale (0 to 40 scale) (41)	Mean plus 1 standard deviation in this sample ≥ 25 (42)
Catastrophization	Pain Catastrophization Scale (0 to 52 scale) (43)	Threshold for clinically relevant level of catastrophization (75 th percentile in chronic pain patients) ≥ 30 (44)
Depression	Patient Health Questionnaire (0 to 27 scale) (45)	Threshold for moderately severe to severe depression ≥ 15 (45) derived in a sample of general medical primary care and obstetrics/gynecology patients

*Mean + 1SD equivalent = highest scoring 16% of sample

Table 3: Anxiety: performance characteristics for the screening question 'Do you feel anxious?' and the Generalized Anxiety Disorder Scale

Threshold	Sensitivity	Specificity	Post-test probability (+ve test result)	Post-test probability (-ve test result)	People correctly ruled out*
1 or more	95.2% (90.8% to 98.0%)	33.9% (30.3% to 37.7%)	27.4% (23.8% to 31.0%)	3.6% (1.1% to 6.0%)	216 (26.9%)
2 or more	94.0% (89.3% to 96.7%)	45.1% (41.2% to 48.9%)	31.0% (26.9% to 35.0%)	3.4% (1.3% to 5.4%)	287 (35.7%)
3 or more	91.0% (85.7% to 94.5%)	54.5% (50.6% to 58.3%)	34.4% (30.0% to 38.8%)	4.1% (2.1% to 6.2%)	347 (43.2%)
4 or more	83.8% (77.5% to 88.6%)	64.2% (60.4% to 67.8%)	38.0% (33.1% to 43.0%)	6.2% (3.9% to 8.5%)	409 (50.9%)
5 or more	80.8% (74.2% to 86.1%)	72.2% (68.6% to 75.6%)	43.3% (37.8% to 48.8%)	6.5% (4.3% to 8.7%)	460 (57.2%)
6 or more	65.3% (57.8% to 72.1%)	82.1% (78.9% to 84.9%)	48.9% (42.3% to 55.4%)	10.0% (7.5% to 12.4%)	523 (65.0%)
7 or more	53.9% (46.3% to 61.3%)	88.5% (85.8% to 90.8%)	55.2% (47.6% to 62.8%)	12.0% (9.5% to 14.5%)	564 (70.1%)
8 or more	40.7% (33.6% to 48.3%)	93.9% (91.7% to 95.5%)	63.6% (54.4% to 72.7%)	14.2% (11.6% to 16.8%)	598 (74.4%)
9 or more	22.8% (17.1% to 29.7%)	97.8% (96.3% to 98.7%)	73.1% (61.0% to 85.1%)	17.2% (14.5% to 19.8%)	623 (77.5%)

10	10.2% (64.5% to 15.7%)	99.7% (98.8% to 99.9%)	89.5% (75.7% to 100.0%)	19.1% (16.4% to 21.9%)	635 (79.0%)
----	---------------------------	---------------------------	----------------------------	---------------------------	----------------

Prevalence in sample = 20.8%

*Number of patients without condition (as classified by the reference standard questionnaire) that would have been correctly ruled out at each cut-point. Bolded results are for the dichotomization threshold chosen for this study

Table 4: Fear of movement: performance characteristics for the screening question 'Physical activity might damage me' and the Tampa Scale for Kinesiophobia

Threshold	Sensitivity	Specificity	Post-test probability (+ve test result)	Post-test probability (-ve test result)	People correctly ruled out*
1 or more	94.9% (92.7% to 96.5%)	31.7% (26.0% to 38.0%)	76.4% (72.3% to 79.7%)	27.3% (18.5% to 36.0%)	72 (9.5%)
2 or more	91.7% (89.0% to 93.7%)	44.9% (38.6% to 51.4%)	79.5% (76.3% to 82.7%)	30.1% (22.7% to 37.6%)	102 (13.5%)
3 or more	85.5% (82.2% to 88.2%)	63.0% (56.6% to 69.0%)	84.4% (81.3% to 87.4%)	35.0% (28.7% to 41.3%)	143 (18.9%)
4 or more	79.6% (76.0% to 82.8%)	70.9% (64.7% to 76.5%)	86.5% (83.4% to 89.5%)	40.2% (34.3% to 46.0%)	161 (21.3%)
5 or more	75.7% (71.8% to 79.1%)	74.4% (68.4% to 79.7%)	87.4% (84.3% to 90.4%)	43.3% (37.7% to 48.9%)	169 (22.3%)
6 or more	57.9% (53.7% to 62.1%)	85.5% (80.3% to 89.5%)	90.3% (78.1% to 93.4%)	53.5% (48.7% to 58.3%)	194 (25.6%)
7 or more	53.2% (49.0% to 57.4%)	88.5% (83.8% to 92.1%)	91.6% (88.5% to 94.7%)	55.2% (50.6% to 59.8%)	201 (26.6%)
8 or more	44.9% (40.7% to 49.2%)	91.2% (86.8% to 94.2%)	92.2% (89.0% to 95.5%)	58.5% (54.5% to 62.8%)	207 (27.3%)
9 or more	29.2% (25.5% to 33.3%)	94.7% (91.0% to 97.0%)	92.8% (88.9% to 96.7%)	63.6% (59.7% to 67.4%)	215 (28.4%)
10	17.7% (14.7% to 21.2%)	96.5% (93.2% to 98.2%)	92.2% (86.9% to 97.4%)	66.6% (63.0% to 70.2%)	219 (28.9%)

Prevalence in sample = 70.0%

*Number of patients without condition (as classified by the reference standard questionnaire) that would have been correctly ruled out at each cut-point. Bolded results are for the dichotomization threshold chosen for this study

Table 5: Stress: performance characteristics for the screening question ‘Do you feel stressed?’ and the reference standard of the Perceived Stress Scale

Threshold	Sensitivity	Specificity	Post-test probability (+ve test result)	Post-test probability (-ve test result)	People correctly ruled out*
1 or more	99.2% (95.4% to 99.9%)	18.4% (15.7% to 21.6%)	17.7% (14.8% to 20.6%)	0.8% (0% to 2.4%)	123 (15.7%)
2 or more	97.5% (92.8% to 99.1%)	25.6% (22.5% to 29.1%)	18.8% (15.7% to 21.9%)	1.7% (0% to 3.7%)	171 (21.8%)
3 or more	96.6% (91.6% to 98.7%)	33.4% (30.0% to 37.1%)	20.4% (17.1% to 23.8%)	1.8% (0.1% to 3.5%)	223 (28.4%)
4 or more	92.4% (86.1% to 95.9%)	42.3% (38.6% to 46.1%)	22.1% (18.4% to 25.7%)	3.1% (1.1% to 5.1%)	282 (35.9%)
5 or more	88.1% (81.1% to 92.8%)	49.6% (45.8% to 53.4%)	23.6% (19.7% to 27.6%)	4.1% (2.0% to 6.1%)	331 (42.2%)
6 or more	80.5% (72.5% to 86.7%)	62.5% (58.8% to 66.1%)	27.5% (22.8% to 32.2%)	5.2% (3.1% to 7.3%)	417 (53.1%)
7 or more	71.2% (62.5% to 78.6%)	70.6% (67.1% to 74.0%)	30.0% (24.6% to 35.4%)	6.7% (4.5% to 8.9%)	471 (60.0%)
8 or more	57.6% (48.6% to 66.2%)	80.7% (77.5% to 83.5%)	34.5% (27.9% to 41.2%)	8.5% (6.2% to 10.8%)	538 (68.5%)
9 or more	38.1% (29.9% to 47.1%)	91.3% (88.9% to 93.2%)	43.7% (34.1% to 53.3%)	10.7% (8.4% to 13.0%)	609 (77.6%)
10	22.9% (16.2% to 31.2%)	96.6% (94.9% to 97.7%)	54.0% (40.2% to 67.8%)	12.4% (10.0% to 14.8%)	644 (82.0%)

Prevalence in sample = 15.0%

*Number of patients without condition (as classified by the reference standard questionnaire) that would have been correctly ruled out at each cut-point. Bolded results are for the dichotomization threshold chosen for this study

Table 6: Catastrophization: performance characteristics for the screening question ‘When I feel the pain, it is terrible and I feel that it’s never going to get better’ and the Pain Catastrophization Scale

Threshold	Sensitivity	Specificity	Post-test probability (+ve test result)	Post-test probability (-ve test result)	People correctly ruled out*
1 or more	100.0% (98.7% to 100%)	4.8% (3.2% to 7.0%)	37.9% (34.5% to 41.3%)	0.0% (0.0% to 0.0%)	24 (3.0%)
2 or more	99.7% (98.1% to 99.9%)	9.8% (7.5% to 12.7%)	39.1% (35.6% to 42.6%)	2.0% (0.0% to 5.9%)	49 (6.2%)
3 or more	99.3% (97.5% to 99.8%)	16.6% (13.6% to 20.1%)	40.9% (37.3% to 44.5%)	2.4% (0.0% to 5.6%)	83 (10.5%)
4 or more	98.6% (96.5% to 99.5%)	25.9% (22.3% to 30.0%)	43.6% (39.8% to 47.4%)	3.0% (0.1% to 5.9%)	130 (16.4%)
5 or more	98.6% (96.5% to 99.5%)	34.7% (30.7% to 39.0%)	46.7% (42.8% to 50.7%)	2.2% (0.1% to 4.4%)	174 (22.0%)
6 or more	96.9% (94.2% to 98.4%)	50.7% (46.3% to 55.1%)	53.3% (49.1% to 57.6%)	3.4% (1.2% to 5.6%)	254 (32.1%)
7 or more	90.7% (86.8% to 93.5%)	60.9% (56.5% to 65.1%)	57.4% (52.9% to 61.9%)	8.1% (5.2% to 11.1%)	305 (38.5%)
8 or more	83.8% (79.2% to 87.6%)	72.1% (68.0% to 75.8%)	63.5% (58.7% to 68.4%)	11.5% (8.4% to 14.6%)	361 (45.6%)
9 or more	63.6% (57.9% to 68.9%)	86.6% (83.4% to 89.3%)	73.4% (68.0% to 78.9%)	19.6% (16.3% to 23.0%)	434 (54.8%)
10	100.0% (98.7% to 100%)	4.8% (3.2% to 7.0%)	37.9% (34.5% to 41.3%)	0.0% (0.0% to 0.0%)	24 (3.0%)

Prevalence in sample = 36.7%

*Number of patients without condition (as classified by the reference standard questionnaire) that would have been correctly ruled out at each cut-point. Bolded results are for the dichotomization threshold chosen for this study

Table 7: Depression: performance characteristics for being above the chosen threshold score on both screening questions and the reference standard of the Patient Health Questionnaire 9-item

Threshold	Sensitivity	Specificity	Post-test probability (+ve test result)	Post-test probability (-ve test result)	People that would have been correctly ruled out*
Scored 6 or more on both	77.2% (71.1% to 83.4%)	73.1% (69.5% to 76.6%)	46.5% (40.5% to 52.1%)	8.6% (6.1% to 11.2%)	434 (56.1%)
Prevalence in sample = 23.3%					

*Number of patients without condition (as classified by the reference standard questionnaire) that would have been correctly ruled out at each cut-point. Bolded results are for the dichotomization threshold chosen for this study

ACCEPTED