SDU
University of Southern Denmark

Preliminary validity and test-retest reliability of two depression questionnaires compared with a diagnostic interview in 99 patients with chronic pain seeking specialist pain treatment

Kha, Thuy Vy; Stenager, Elsebeth; Hoang, Huong; Bruun-Plesner, Karin; Fuglsang, Kira Søndberg; Søgaard la Cour, Birgitte; Handberg, Gitte; Vaegter, Henrik Bjarke

Published in:
Scandinavian Journal of Pain

DOI:
10.1515/sjpain-2020-0042

Publication date:
2020

Document version:
Final published version

Citation for pulished version (APA):
Kha, T. V., Stenager, E., Hoang, H., Bruun-Plesner, K., Fuglsang, K. S., Søgaard la Cour, B., Handberg, G., & Vaegter, H. B. (2020). Preliminary validity and test-retest reliability of two depression questionnaires compared with a diagnostic interview in 99 patients with chronic pain seeking specialist pain treatment. *Scandinavian Journal of Pain*, *20*(4), 717-726. https://doi.org/10.1515/sjpain-2020-0042

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

Download date: 13. Feb. 2025

Original article

Thuy Vy Kha, Elsebeth Stenager, Huong Hoang, Karin Bruun-Plesner, Kira Søndberg Fuglsang, Birgitte Søgaard la Cour, Gitte Handberg and Henrik Bjarke Vaegter*

Preliminary validity and test–retest reliability of two depression questionnaires compared with a diagnostic interview in 99 patients with chronic pain seeking specialist pain treatment

https://doi.org/10.1515/sjpain-2020-0042
Received March 25, 2020; accepted June 8, 2020; published online xxx

Abstract

Objectives: Depression symptomatology is highly prevalent in patients with chronic pain, but accurate identification of major depression may be challenged due to time constraints and diagnostic interviews are therefore not routinely performed in clinical practice. Assessment of depression may be facilitated through the use of full-length depression screening questionnaires with acceptable construct validity and test-retest reliability. However, as previously indicated screening questionnaires may overestimate depression in patients with chronic pain, possibly due to overlapping symptoms. However, the failure to screen for depression may raise a concern for missing relevant cases with depression. The objectives of this study were to (1) quantify the validity of the 9-items Patient-Health Questionnaire (PHQ9) and the Major Depression Inventory (MDI) compared with a diagnostic interview in patients with chronic pain seeking specialist pain treatment, and (2) assess the relative test-retest reliability of PHQ9 and MDI over two weeks.

Methods: Responses to the PHQ9 and MDI were compared with a Present-State-Examination (PSE) interview in 99 patients with chronic pain referred to interdisciplinary pain treatment. PHQ9 and MDI were completed twice over two weeks. Construct validity were assessed with the area under the curve (AUC) analysis, and performance characteristics derived from 2×2 contingency tables in which scores on the screening questionnaires were dichotomized and compared with the classification of clinical depression based on the diagnostic interview. Relative test-retest reliability was assessed with intraclass correlation coefficients (ICC).

Results: Based on the PSE interview, the prevalence of depression was 22.2%, and according to the PHQ9 and MDI questionnaires the prevalence was 26.3 and 34.3%, respectively. Compared with the diagnostic PSE, the PHQ9 and MDI questionnaires had areas under the curve of 0.83 and 0.88, respectively. Both questionnaires had high negative predictive values (PHQ9: cut-off of 11; MDI: cut-off of 26), but low positive predictive values for all possible scores. ICC values were excellent.

Conclusions: The PHQ9 and MDI questionnaires reliably identified chronic pain patients unlikely to have clinical depression, but showed limited validity identifying patients with clinical depression. These preliminary results may have clinical implications in depression screening in

***Corresponding author: Henrik Bjarke Vaegter, Ph.D**, Department of Clinical Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark; and Pain Research Group, Pain Center, University Hospital Odense, Heden 7-9, Indgang 200, DK – 5000 Odense, Denmark, Phone: +45 65413869, Fax: +45 65415064, E-mail: hbv@rsyd.dk. https://orcid.org/0000-0002-7707-9947 (H.B. Vaegter)

Thuy Vy Kha: Department of Clinical Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark, E-mail: vy_kt@hotmail.com

Elsebeth Stenager: Research Unit in Mental Health, Åbenrå, Department of Regional Health Services, University of Southern Denmark, Odense, Denmark; Department of Psychiatry, University of Southern Denmark, Odense, Denmark, E-mail: Elsebeth.Stenager@rsyd.dk

Huong Hoang: Department of Psychiatry, University of Southern Denmark, Odense, Denmark, E-mail: Huong.Hoang@rsyd.dk

Karin Bruun-Plesner and Gitte Handberg: Department of Clinical Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark; Pain Research Group, Pain Center, Odense University Hospital, Odense, Denmark, E-mail: Karin.Due.Bruun@rsyd.dk (K. Bruun-Plesner), gitte.handberg@rsyd.dk (G. Handberg)

Kira Søndberg Fuglsang and Birgitte Søgaard la Cour: Pain Research Group, Pain Center, Odense University Hospital, Odense, Denmark, E-mail: kira.fuglsang@rsyd.dk (K.S. Fuglsang), birgittesmail@hotmail.com (B. Søgaard la Cour)

patients with chronic pain seeking specialist pain treatment. Clinicians in a specialty care pain clinic can use these screening questionnaires to identify patients without depression, but caution should be used when positive cases are identified by PHQ9 or MDI due to the risk of false positives.

Keywords: chronic pain; depression; diagnostic interview; MDI; PHQ9; questionnaires.

Introduction

Chronic pain and depression are among the top 10 leading causes of disability worldwide [1]. Patients with chronic pain are more likely to be depressed than people without chronic pain [2], and chronic pain with comorbid depression is generally associated with higher levels of disability [3], lower quality of life, increased health care utilization, amplification of pain [4], and higher treatment dropouts [5]. Therefore, accurate knowledge of the presence or absence of clinical depression may better inform clinicians about their individual patients, and enable appropriate and individualized treatment, and potentially improve treatment outcomes.

Rayner and colleagues observed a high prevalence of depression around 60% in 1,204 patients attending a specialist pain setting in the United Kingdom using the 9-item Patient Health Questionnaire (PHQ9) [6], whereas Vaegter and colleagues observed a much lower prevalence of moderate to severe depression around 23% in 894 patients attending one specialist pain setting in Denmark using the PHQ9 [7], and around 23% in 6,637 patients attending six different specialist pain settings in Denmark [8] suggesting that the prevalence of clinical depression in patients with chronic pain may be quite different across countries. In a similar Danish patient sample Søndergård and colleagues observed an even lower prevalence of clinical depression around 6% in 7,197 patients attending a specialist pain setting using a golden standard diagnostic interviews performed by psychiatrists [2].

Accurate identification of clinical depression through a diagnostic interview in patients with chronic pain is, however, a challenge due to the time burden and specific training needed by clinicians, and diagnostic interviews are therefore not routinely performed in clinical practice. Therefore, assessment of depression may be facilitated through the use of full-length depression screening questionnaires with acceptable construct validity and test-retest reliability. However, as previously indicated [2], it is possible that screening questionnaires overestimate depression in patients with chronic pain, possibly due to overlapping symptoms, such as fatigue, sleep disturbance, concentration difficulties and decreased appetite [9, 10]. However, the failure to screen for depression may raise a concern for missing relevant cases with depression.

False positive diagnosis may result in unnecessary medical treatment, and therefore identification of patients with low probability of having depression might also have great relevance in the clinic. The importance of correct diagnosis is therefore crucial as it may reduce the frequency of wrong diagnosis, treatment and costs. Still, the use of depression screening questionnaires in patients with chronic pain do have a number of potential benefits including easily performed screening of people as a resource-efficient triage to inform clinical evaluation, and as a possible indication for further assessment of possible depression.

Currently, the performances of two commonly used depression questionnaires; the PHQ9 and the Major Depression Inventory (MDI) have been established in patients at psychiatry and community adults [11], and in subjects with chronic back pain [12], however these depression questionnaires have not been compared to a diagnostic interview in patients with mixed chronic pain conditions referred to interdisciplinary pain rehabilitation, and optimal cut-points for more accurate assessment of depression in this population is unknown. In addition, depression questionnaires are recommended [13], and often used as outcome measure in patients with chronic pain. Therefore, during the establishment of a national clinical pain registry (PainData, Denmark) for patients with chronic pain referred for assessment and treatment at public and private interdisciplinary pain clinic in Denmark the performances of the freely available PHQ9 and the MDI were assessed in patients with mixed chronic pain conditions referred to one of the public interdisciplinary pain centres. The objectives of this study were to (1) quantify the construct validity of the PHQ9 and the MDI assessing depression compared with a diagnostic interview in patients with chronic pain seeking specialist pain treatment, and (2) assess the relative test-retest reliability of PHQ9 and MDI over two weeks prior to patients commencing treatment.

Materials and methods

Method summary

This study compared patient responses on the PHQ9 and MDI depression screening questionnaires with a diagnostic interview. The

criteria used for assessing construct validity were the area under the curve (AUC) seen on a receiver operating characteristic (ROC) curve analysis, and performance characteristics derived from 2 × 2 contingency tables in which scores on the screening questionnaires were dichotomized and compared with the classification of depression based on the diagnostic interview. The criteria used for assessing relative test-retest reliability was intraclass correlation coefficients (ICC).

Procedures

After referral to an interdisciplinary pain center (Pain Center, Odense University Hospital, Denmark) but prior to the initial consultation, all referred patients were invited to answer questions about their clinical characteristics and adaptations to pain via a web-based clinical pain registry (PainData) implemented in most public and private pain clinics in Denmark. Questionnaires were completed at home prior to the first consultation. In addition to the PHQ9 and MDI questionnaires, collected data included full-length questionnaires for measures of pain intensity, activity limitation, health-related quality of life, pain drawings, and several other psychological constructs. Moreover, patients were asked for consent to whether the Pain Center was allowed to contact them in case of ongoing projects that were relevant for people with chronic pain. Each week of the study period (from October 2017 to September 2018), three patients who had completed the questionnaires and gave consent to contact from the Pain Center were randomly selected from the total group of approximately 22 newly referred patients completing questionnaires each week. These patients were contacted by telephone and invited to participate in a diagnostic interview at the Pain Center approximately two weeks after completion of the questionnaires. However, patients were excluded if they did not speak Danish well enough to complete the diagnostic interview. On the day of the diagnostic interview, patients were asked to complete the PHQ9 and MDI questionnaires again prior to the interview. 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.

Diagnostic interview

Present state examination: Present state examination (PSE) is a semi-structured diagnostic interview developed by Wing and colleagues [14] for use by trained clinicians to assess and classify psychopathology and behaviour associated with major psychiatric disorders experienced by the patient during the last month. The ninth edition is translated into more than 20 languages, and it provides reliable descriptions of symptoms which forms the foundation of a diagnosis [14].

The PSE was performed by one of three pain specialist all experienced doctors or psychologist (KSF, KBP, BSC) from the Pain Center. The three specialists have all participated in a week long PSE training course, together with medical doctors, psychologist and nurses from psychiatry. Furthermore they have been thoroughly trained and supervised by a specialist in psychiatry (HH) who has extensive training in the use of diagnostic interviews and in performing the PSE. The rationale for choosing pain specialist for the assessment after relevant training was for practical reasons. After completing the interview, a diagnosis was given according to the ICD-10 criteria for Depression (F 32.0–F 33.0 assessing mild, moderate or severe depression). Interviewers were blinded to the patient responses on the depression questionnaires. According to ICD-10, patients with moderate or severe depression on PSE were diagnosed with depression.

Depression questionnaires

Patient Health Questionnaire 9-items: The 9-item Patient Health Questionnaire (PHQ9) is a brief, self-reported instrument developed by Kroenke and colleagues [15] for screening, diagnosing and monitoring depressive symptoms using criteria from the DSM-IV. PHQ9 consists of nine-items evaluating the presence of the nine symptoms of depression over the last two weeks: (I) depressed mood, (II) anhedonia, (III) trouble sleeping, (IV) feeling tired, (V) change in appetite or weight, (VI) guilt or worthlessness, (VII) trouble concentrating, (VIII) feeling slowed down or restless, and (IX) suicidal thoughts. Each item is assessed on a four-point Likert scale, ranging from 0="not at all" to 3="nearly every day", resulting in a summed score of 0–27 points. PHQ9 scores of 5–9, 10–14, 15–19 and 20–27 represents mild, moderate, moderately severe, and severe depression respectively. For diagnostic purposes, the algorithm method requires at least five symptoms rated as "more than half the days", except from the suicidal ideation item, which counts as one of the five symptoms if rated as "several days" or above. Furthermore, item-one (loss of interest, pleasure or depressed mood) has to be rated at least as "more than half the days" [16].

PHQ9 has been widely used to assess the prevalence of depressive symptoms in chronic pain populations [17]. It has been demonstrated to be a valid instrument measuring depressive symptoms in primary care patients, with excellent internal consistency reliability (α=0.89) [15]. Additionally, PHQ9 has shown good validity (r=0.84) as compared to health-professional administered versions of the scale, as well as good construct validity as indicated by a strong association between increasing PHQ9 severity scores and worsening function on the SF-20 [15]. The PHQ9 was also proposed as a measure of depression severity in the International Association for the Study of Pain (IASP) proposal for a structural assessment approach [13]. Previous systematic reviews and meta-analysis have concluded that the PHQ9 cut-point (>10) has good diagnostic accuracy identifying depression in an unselected sample in primary care with a sensitivity and a specificity of 84 and 88%, respectively [18], as well as in patients with chronic physical illnesses [19].

Major depression inventory: The major depression inventory (MDI) is a brief, self-reported instrument developed by Bech and colleagues [11] in collaboration with the Danish WHO Collaborating centre. It was developed to measure ICD-10 and DSM-IV criteria for screening and diagnosing moderate to severe depression [11]. The Danish version of the MDI consists of 12-items evaluating nine symptoms of depression over the last two weeks. Each item is rated on a six-point Likert scale, ranging from 0="at no time" to 5="all the time". The total score on this scale is calculated as the sum of all items, except from the items 8a and 8b (agitation/retardation) and the items 10a and 10b (increased/reduced appetite) where only the highest score of each is used

Table 1: Raw scores (mean ± SD and 95% confidence intervals) for demographics, clinical pain profile, psychological distress, and depression scores in patients with chronic pain and comorbid depression, and chronic pain and no comorbid depression. Diagnosis of depression is according to ICD-10 based on moderate or severe depression on the Present State Examination interview. 'BMI': Body Mass Index. 'NRS': Numerical Rating Scale. 'GAD7': Generalized Anxiety Disorder 7-items. 'PCS': Pain Catastrophizing Scale. 'TSK': Tampa Scale of Kinesiophobia. 'PDI': Pain Disability Index. 'PHQ9': Patient Health Questionnaire 9-item. 'MDI': Major Depression Inventory. p-Values are based on by Chi-square tests for categorical variables and Student's *t*-test for continuous variables.

Domain	Variable	Total (*n*=99)	Chronic pain + depression (*n*=22)	Chronic pain - depression (*n*=77)	p-Value
Demographics	Gender (Women/Men)	74/25	16/6	58/19	0.81
	Age (years)	47.3 ± 13.3	44.5 ± 11.3	48.1 ± 13.7	0.27
	BMI (kg/m^2)	28.3 ± 5.9	28.4 ± 4.6	28.3 ± 6.3	0.94
Clinical pain	Pain duration (years)	11.4 ± 10.3	13.0 ± 9.5	10.9 ± 10.5	0.40
	Peak pain intensity (NRS: 0–10)	7.7 ± 1.8	8.2 ± 1.4	7.6 ± 1.9	0.11
	Average pain intensity (NRS: 0–10)	6.6 ± 1.9	7.3 ± 1.5	6.3 ± 2.0	0.02
Psychological	Anxiety (GAD7: 0–21)	7.1 ± 5.6	12.1 ± 5.6	5.6 ± 4.8	<0.001
	Pain Catastrophizing (PCS: 0–52)	25.4 ± 11.6	33.0 ± 11.5	23.2 ± 10.8	<0.001
	Fear of movement (TSK: 17–68)	39.6 ± 8.6	43.2 ± 7.1	38.6 ± 8.7	0.016
	Disability (PDI: 0–50)	36.3 ± 8.6	39.6 ± 6.5	35.3 ± 8.9	0.015
	Quality of Life (EQ5D: 0–100)	46.7 ± 22.2	33.7 ± 20.0	50.3 ± 21.5	0.002
Depression	PHQ: 0–27	12.7 ± 5.3	17.4 ± 4.7	11.3 ± 4.6	<0.001
	MDI: 0–50	23.4 ± 10.5	34.1 ± 6.5	20.4 ± 9.4	<0.001

resulting in a summed score between 0 and 50. According to the MDI guidelines, the cut-points for the total score of MDI are ≤19, 20–24, 25–29, ≥30 representing no, mild, moderate, and severe depression, respectively. For diagnostic purposes, the algorithm method requires that items are dichotomized indicating the presence (=1) or absence (=0) of each symptom. For a diagnosis of moderate to severe depression, at least two of the first three core symptom items must at least be rated as "most of the time" and at least four of the seven associated symptom items (items 4–10) must at least be rated as "more than half the time". The MDI has shown acceptable sensitivity and a specificity of 90 and 82%, respectively in a sample of patients referred to two psychiatric departments [11], and sensitivity and a specificity of 66 and 63% in outpatients with depressive symptoms [20].

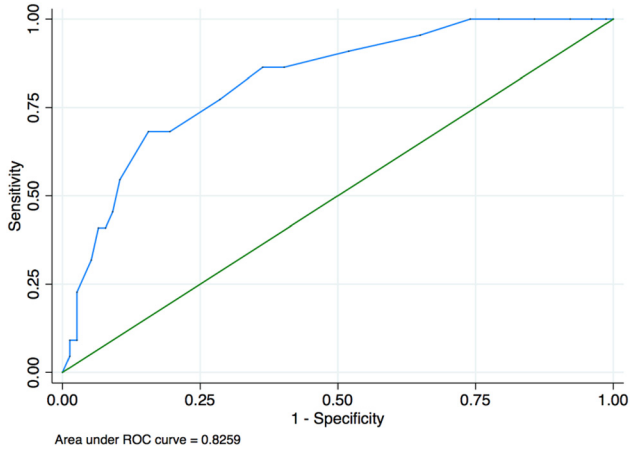

Figure 1: Receiving operator characteristic curve for the 9-items patient-health-questionnaire (PHQ9) and the present-state-examination (PSE).

Statistical analysis

All statistical analysis was performed using STATA IC version 15.0 (STATA Corp., College Station, TX) and Excel 2010 version 14.0 (Microsoft Corp, Redmond, WA). As this was a preliminary study as part of the implementation of a national clinical registry we aimed to include 100 patients with chronic pain seeking specialist pain treatment. There was no missing data on PHQ9, MDI or the PSE.

First, descriptive analysis based on means and standard deviation (SD) was used to describe the population and Student's *t*-tests were used to investigate potential differences in demographics, clinical pain profiles, and items responses on the PHQ9 and MDI between patients who based on the PSE were diagnosed with depression (i.e. moderate or severe depression on PSE) and patients without depression (i.e. mild depression or no depression on PSE). Mean score differences were expressed in terms of effect sizes according to Cohen [21]. Secondly, performance characteristics (sensitivity, specificity, post-test probability if the test was positive or negative as well as positive predictive and negative predictive values) of the questionnaires based on total sum-scores and the diagnostic algorithms compared with the PSE were derived from 2×2 contingency tables for every possible dichotomization threshold for scores on the PHQ9 and MDI questionnaires. The sensitivity measure is the proportion of patients with clinical depression based on the PSE who are correctly identified by the questionnaire as having depression at a specific threshold. The specificity measure is the proportion of patients who do not have clinical depression based on the PSE who are correctly identified by the questionnaire as not having depression at a specific threshold. The positive and negative post-test probability measures are the probabilities of the presence of depression after a score on the questionnaires above or below a specific threshold. The positive predictive value is the probability that if a patient has a positive test (i.e. a score on the questionnaire above a specific threshold) that this

Table 2: Performance Characteristics for the 9-items Patient-Health Questionnaire (PHQ9) sum-score and the Present-State-Examination (PSE) in 99 patients with mixed chronic pain seeking specialist pain treatment.

Cut-point PHQ9	Sensitivity	Specificity	Post-test Probability (+Test resulton PHQ9)	Post-test probability (– Test resulton PHQ9)	Positive predictive value	Negative predictive value
≥1	100 (100–100)	0 (0 to 0)	22.0 (14.0–30.4)	-	0.22 (0.14–0.30)	-
≥2	100 (100–100)	0 (0 to 0)	22.0 (14.0–30.4)	-	0.22 (0.14–0.30)	-
≥3	100 (100–100)	1.0 (–0.9 to 3.5)	22.0 (14.2–30.7)	0.00 (0.0–0.0)	0.22 (0.14–0.31)	1.0 (1.0–1.0)
≥4	100 (100–100)	1.0 (–0.9 to 3.5)	22.0 (14.2–30.7)	0.00 (0.0–0.0)	0.22 (0.14–0.31)	1.0 (1.0–1.0)
≥5	100 (100–100)	4.0 (1 to 7.7)	23.0 (14.6–31.2)	0.00 (0.0–0.0)	0.23 (0.15–0.31)	1.0 (1.0–1.0)
≥6	100 (100–100)	8.0 (2.5 to 13.1)	24.0 (15.3–32.0)	0.00 (0.0–0.0)	0.24 (0.15–0.32)	1.0 (1.0–1.0)
≥7	100 (100–100)	14.0 (7.4 to 21.2)	25.0 (16.5–33.5)	0.00 (0.0–0.0)	0.25 (0.17–0.34)	1.0 (1.0–1.0)
≥8	100 (100–100)	21.0 (12.8 to 28.8)	26.5 (17.8–35.2)	0.00 (0.0–0.0)	0.27 (0.18–0.35)	1.0 (1.0–1.0)
≥9	100 (100–100)	26.0 (17.3 to 34.6)	28.0 (19.0–36.7)	0.00 (0.0–0.0)	0.28 (0.19–0.37)	1.0 (1.0–1.0)
≥10	95.0 (91.4–99.6)	35.0 (25.7 to 44.5)	30.0 (20.6–38.6)	4.0 (0.0–7.2)	0.30 (0.21–0.39)	0.96 (0.93–1.00)
≥11	91.0 (85.2–96.6)	48.0 (38.2 to 57.9)	33.0 (24.0–42.6)	5.0 (8.0–9.5)	0.33 (0.24–0.43)	0.95 (0.91–0.99)
≥12	86.0 (79.6–93.1)	60.0 (50.1 to 69.4)	38.0 (28.4–47.6)	6.0 (1.4–10.8)	0.38 (0.28–0.48)	0.94 (0.89–0.99)
≥13	86.0 (79.6–93.1)	64.0 (54.2 to 73.1)	40.0 (30.8–50.1)	6.0 (1.2–10.4)	0.40 (0.31–0.50)	0.94 (0.90–0.99)
≥14	77.0 (69.0–85.5)	0.71 (62.5 to 80.3)	44.0 (33.8–53.4)	8.0 (2.9–13.8)	0.44 (0.34–0.53)	0.92 (0.86–0.97)
≥15	68.0 (59.0–77.4)	81.0 (72.7 to 88.3)	50.0 (40.2–59.8)	10.0 (4.2–16.1)	0.50 (0.40–0.60)	0.90 (0.84–0.96)
≥16	68.0 (59.0–77.4)	84.0 (77.3 to 91.3)	56.0 (45.8–65.3)	10.0 (3.9–15.6)	0.56 (0.46–0.65)	0.90 (0.84–0.96)
≥17	55.0 (44.7–64.4)	90.0 (83.6 to 95.6)	60.0 (50.3–69.7)	13.0 (6.1–19.2)	0.60 (0.50–0.70)	0.87 (0.81–0.94)
≥18	45.0 (35.6–55.3)	91.0 (85.2 to 96.6)	59.0 (49.1–68.5)	15.0 (7.7–21.6)	0.59 (0.49–0.69)	0.85 (0.78–0.92)
≥19	41.0 (31.2–50–6)	92.0 (86.9 to 97.5)	60.0 (50.3–69.7)	15.0 (8.4–22.6)	0.60 (0.50–0.70)	0.85 (0.77–0.92)
≥20	41.0 (31.2–50–6)	94.0 (88.7 to 98.4)	64.0 (54.8–73.7)	15.0 (8.2–22.4)	0.64 (0.55–0.74)	0.85 (0.78–0.92)
≥21	32.0 (22.6–41.0)	95.0 (90.4 to 99.2)	64.0 (54.2–73.13)	17.0 (9.6–24.5)	0.64 (0.54–0.73)	0.83 (0.76–0.90)
≥22	23.0 (14.5–31.0)	97.0 (94.3 to 100)	71.0 (62.5–80.3)	18.0 (10.8–26.1)	0.71 (0.63–0.80)	0.82 (0.74–0.89)
≥23	9.0 (3.4–14.8)	97.0 (94.3 to 100)	50.0 (40.2–59.8)	21.0 (13.0–29.1)	0.50 (0.40–0.60)	0.79 (0.71–0.87)
≥24	9.0 (3.4–14.8)	99.0 (96.5 to 100)	67.0 (57.4–76.0)	21.0 (12.8–28.8)	0.67 (0.57–0.76	0.79 (0.71–0.87)
≥25	5.0 (0.4–8.6)	99.0 (96.5 to 100)	50.0 (40.2–59.8)	22.0 (13.5–29.8)	0.50 (0.40–0.60)	0.78 (0.70–0.87)
≥26	5.0 (0.4–8.6)	99.0 (96.5 to 100)	50.0 (40.2–59.8)	22.0 (13.5–29.8)	0.50 (0.40–0.60)	0.78 (0.70–0.87)
27	5.0 (0.4–8.6)	99.0 (96.5 to 100)	50.0 (40.2–59.8)	22.0 (13.5–29.8)	0.50 (0.40–0.60)	0.78 (0.70–0.87)

patients is also classified with a clinical depression based in the PSE. The negative predictive value is the probability that if a patient has a negative test (i.e. a score on the questionnaire below a specific threshold) that this patients is classified with no clinical depression based in the PSE. Thirdly, the AUC statistics from ROC analysis as a performance measure (an index of the classification error rate) of the questionnaires relative to diagnosis on the PSE were calculated. An AUC of 0.5 indicated that a test has no better accuracy chance whereas a test with perfect accuracy has an AUC of 1.0. Agreement in classification between the diagnostic algorithms of the questionnaires and the PSE was assessed using Cohen's kappa coefficient. Kappa result was interpreted with values ≤0 indicating no agreement, 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement.

Finally, the two weeks relative test-retest reliability of the PHQ9 and MDI sum-scores were investigated with Intraclass correlation coefficients based on a single rating, consistency, 2-way mixed effect model ($ICC_{3,1}$). The relative reliability is a measure of how well the questionnaires differentiate the scores between different subjects across days) with higher ICC scores indicating better ability to differentiate between subjects (e.g. patients with and without depression). ICC values range from 0 to 1, with higher scores indicating high consistency between the two time points. ICC above 0.75 was taken as excellent reliability, 0.40–0.75 was fair to good reliability, and less than 0.40 defined poor reliability [22].

Results

Participants

Out of 1278 patients referred to the Pain Centre in the study period, 112 were randomly selected and contacted by phone. During the telephone contact, patients were briefly told about the study and invited to take part in an interview in the Pain Centre. Ninety-nine (88.4%) patients (Table 1) were willing to participate in the interview. 13 patients declined due to transportation issue, or lack of physical and mental resources.

According to the PSE interview, 22 (22.2%) patients were classified with depression. Patients with depression reported significantly higher average pain intensity, anxiety,

Table 3: 2 × 2 contingency table for the 9-items Patient-Health Questionnaire (PHQ9) diagnostic algorithm and the Present State Examination (PSE) in 99 patients with mixed chronic pain seeking specialist pain treatment.

Cohen's kappa coefficient: κ = 0.29 [95% CI: 0.08–0.50])		PSE	
		Depression	No depression
PHQ9	Depression	11	15
	No depression	11	62

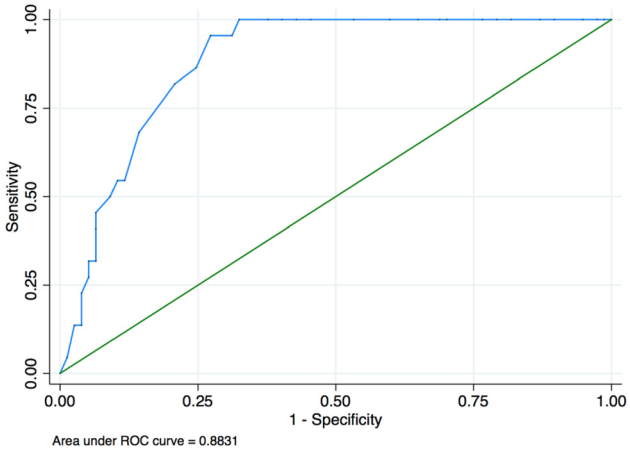

Figure 2: Receiving Operator Characteristic Curve for the Major Depression Inventory (MDI) questionnaire and the Present State Examination (PSE).

pain catastrophizing, fear of movement and disability as well as lower quality of life compared with patients without depression (Table 1). No significant differences in pain duration, peak pain intensity, age, or gender distribution were found between patients with depression and patients without depression based on the PSE.

Mean scores of the PHQ9 and MDI items

The PHQ9 mean scores (Supplementary Table 1) of the depressed patients ranged from 0.6 (suicidal ideation) to 2.9 (lack of energy). All items showed higher mean scores in depressed patients compared with non-depressed patients. The greatest differences between groups, expressed in terms of effect sizes were items 2 (depressed mood), 6 (self-blame), and 8 (agitation/retardation).

The MDI mean scores (Supplementary Table 2) of the depressed patients ranged from 1.5 (increase in appetite) to 4.5 (lack of energy). All items showed higher mean scores in depressed patients compared with non-depressed patients. The greatest differences between groups, expressed in terms of effect sizes were items 1 (depressed mood), 4 (low self-esteem), and 2 (lack of interests).

PHQ9 diagnostic validity: summed-score and diagnostic algorithm

Compared with PSE, the AUC was 0.83 ± 0.04 (Figure 1; 95% CI: 0.73–0.92) for the PHQ9 sum score suggesting an excellent performance in discriminating patients with and without clinical diagnosis of depression. The best balance between sensitivity and specificity was a PHQ9 sum-score threshold ≥14 (Table 2).

According to the PHQ9 diagnostic algorithm, 26 (26.3%) patients were classified with depression on the day of the PSE interview (Table 3). For the diagnostic algorithm, a sensitivity and specificity of 50% (95% CI: 40.2–59.8) and 81% (95% CI: 72.7–88.3) were found, and the post-test probability for a positive and negative test were 42% (95% CI: 32.6–52.0%) and 15% (95% CI: 8.0–22.1%), respectively. Agreement in the classification of depression and no-depression between PHQ9 and PSE was 73.7% (κ = 0.29 [95% CI: 0.08–0.50]).

MDI diagnostic validity: summed-score and diagnostic algorithm

Compared with PSE, the AUC was 0.88 ± 0.03 (Figure 2; 95% CI: 0.82–0.95) for the MDI sum score suggesting an excellent performance in discriminating patients with and without clinical diagnosis of depression. The best balance between sensitivity and specificity was a MDI sum-score threshold ≥26 (Table 4).

According to the MDI diagnostic algorithm, 34 (34.3%) patients were classified with depression on the day of the PSE interview (Table 5). For the diagnostic algorithm, a sensitivity and specificity of 77% (95% CI: 69.0–85.5) and 78% (95% CI: 69.8–86.1) were found, and the post-test probability for a positive and negative test were 50% (95% CI: 40.2–59.8%) and 8% (95% CI: 2.4–12.9%), respectively. Agreement in the classification of depression and no-depression between MDI and PSE was 77.8% (κ = 0.46 [95% CI: 0.27–0.64]).

Test–retest reliability of PHQ9 and MDI

Test-retest reliability of the PHQ9 and the MDI across two repeated assessments showed excellent ICCs of 0.83 and 0.90 (P<0.001), respectively (Table 6).

Table 4: Performance Characteristics for the Major Depression Inventory (MDI) sum-score and the Present-State-Examination (PSE) in 99 patients with mixed chronic pain seeking specialist pain treatment.

Cut-point MDI	Sensitivity	Specificity	Post-test Probability (+Test result on MDI)	Post-test probability (– Test result on MDI)	Positive predictive value	Negative predictive value
≥1	100 (100–100)	0 (0 to 0)	22.0 (14.0–30.4)	-	0.22 (0.14–0.30)	-
≥2	100 (100–100)	0 (0 to 0)	22.0 (14.0–30.4)	-	0.22 (0.14–0.30)	-
≥3	100 (100–100)	1.0 (–0.9 to 3.5)	22.0 (14.2–30.7)	0.00 (0.0 to 0.0)	0.22 (0.14–0.31)	1.0 (1.0–1.0)
≥4	100 (100–100)	3.0 (–0.5 to 5.7)	23.0 (14.4–30.9)	0.00 (0.0 to 0.0)	0.23 (0.14–0.31)	1.0 (1.0–1.0)
≥5	100 (100–100)	3.0 (–0.5 to 5.7)	23.0 (14.4–30.9)	0.00 (0.0 to 0.0)	0.23 (0.14–0.31)	1.0 (1.0–1.0)
≥6	100 (100–100)	3.0 (–0.5 to 5.7)	23.0 (14.4–30.9)	0.00 (0.0 to 0.0)	0.23 (0.14–0.31)	1.0 (1.0–1.0)
≥7	100 (100–100)	5.0 (0.8 to 9.6)	23.0 (14.8–31.5)	0.00 (0.0 to 0.0)	0.23 (0.15–0.32)	1.0 (1.0–1.0)
≥8	100 (100–100)	10.0 (4.4 to 16.4)	24.0 (15.7–32.6)	0.00 (0.0 to 0.0)	0.24 (0.16–0.33)	1.0 (1.0–1.0)
≥9	100 (100–100)	13.0 (6.4 to 19.6)	25.0 (16.2–33.2)	0.00 (0.0 to 0.0)	0.25 (0.16–0.33)	1.0 (1.0–1.0)
≥10	100 (100–100)	13.0 (6.4 to 19.6)	25.0 (16.2–33.2)	0.00 (0.0 to 0.0)	0.25 (0.16–0.33)	1.0 (1.0–1.0)
≥11	100 (100–100)	18.0 (10.6 to 25.8)	26.0 (17.3–34.5)	0.00 (0.0 to 0.0)	0.26 (0.17–0.35)	1.0 (1.0–1.0)
≥12	100 (100–100)	21.0 (12.8 to 28.8)	27.0 (17.8–35.2)	0.00 (0.0 to 0.0)	0.27 (0.18–0.35)	1.0 (1.0–1.0)
≥13	100 (100–100)	23.0 (15.0 to 31.7)	27.0 (18.4–35.9)	0.00 (0.0 to 0.0)	0.27 (0.18–0.36)	1.0 (1.0–1.0)
≥14	100 (100–100)	30.0 (20.9 to 38.9)	29.0 (20.0–37.9)	0.00 (0.0 to 0.0)	0.29 (0.20–0.38)	1.0 (1.0–1.0)
≥15	100 (100–100)	31.0 (22.0 to 40.3)	29.0 (20.4–38.3)	0.00 (0.0 to 0.0)	0.29 (0.20–0.38)	1.0 (1.0–1.0)
≥16	100 (100–100)	35.0 (25.7 to 44.5)	31.0 (21.5–39.6)	0.00 (0.0 to 0.0)	0.31 (0.22–0.40)	1.0 (1.0–1.0)
≥17	100 (100–100)	40.0 (30.6 to 49.9)	32.0 (23.1–41.6)	0.00 (0.0 to 0.0)	0.32 (0.23–0.42)	1.0 (1.0–1.0)
≥18	100 (100–100)	47.0 (36.9 to 56.6)	35.0 (25.5–44.3)	0.00 (0.0 to 0.0)	0.35 (0.26–0.44)	1.0 (1.0–1.0)
≥19	100 (100–100)	55.0 (44.7 to 64.4)	39.0 (29.0–48.2)	0.00 (0.0 to 0.0)	0.39 (0.29–0.48)	1.0 (1.0–1.0)
≥20	100 (100–100)	57.0 (47.4 to 66.9)	40.0 (30.3–49.7)	0.00 (0.0 to 0.0)	0.40 (0.30–0.50)	1.0 (1.0–1.0)
≥21	100 (100–100)	60.0 (50.1 to 69.4)	42.0 (31.8–51.2)	0.00 (0.0 to 0.0)	0.42 (0.32–0.51)	1.0 (1.0–1.0)
≥22	100 (100–100)	62.0 (52.8 to 71.9)	43.0 (33.4–52.9)	0.00 (0.0 to 0.0)	0.43 (0.33–0.53)	1.0 (1.0–1.0)
≥23	96.0 (91.6–99.7)	68.0 (58.3 to 76.8)	47.0 (37.0–56.6)	0.00 (0.0 to 0.0)	0.47 (0.37–0.57)	1.0 (1.0–1.0)
≥24	95.0 (91.4–99.6)	69.0 (59.7 to 78.0)	47.0 (36.8–56.5)	2.0 (–0.8 to 4.5)	0.47 (0.37–0.57)	0.98 (0.96–1.00)
≥25	95.0 (91.4–99.6)	73.0 (64.0 to 81.5)	50.0 (40.2–59.8)	2.0 (–0.8 to 4.3)	0.50 (0.40–0.60)	0.98 (0.96–1.00)
≥26	86.0 (79.6–93.1)	75.0 (66.8 to 83.8)	50.0 (40.2–59.8)	5.0 (0.7 to 9.2)	0.50 (0.40–0.60)	0.95 (0.91–0.99)
≥27	82.0 (74.2–89.4)	79.0 (71.2 to 87.2)	53.0 (43.1–62.8)	6.0 (1.4 to 10.9)	0.53 (0.43–0.63)	0.94 (0.89–0.99)
≥28	82.0 (74.2–89.4)	79.0 (71.2 to 87.2)	53.0 (43.1–62.8)	6.0 (1.4 to 10.9)	0.53 (0.43–0.63)	0.94 (0.89–0.99)
≥29	68.0 (59.0–77.4)	86.0 (78.8 to 92.6)	58.0 (48.0–67.4)	10.0 (3.8 to 15.4)	0.58 (0.48–0.67)	0.90 (0.85–0.96)
≥30	55.0 (44.7–64.4)	88.0 (82.0 to 94.6)	57.0 (47.4–66.9)	13.0 (6.2 to 19.4)	0.57 (0.47–0.67)	0.87 (0.81–0.94)
≥31	55.0 (44.7–64.4)	90.0 (83.6 to 95.6)	60.0 (50.3–69.7)	13.0 (6.1 to 19.2)	0.60 (0.50–0.70)	0.87 (0.81–0.94)
≥32	55.0 (44.7–64.4)	90.0 (83.6 to 95.6)	60.0 (50.3–69.7)	13.0 (6.1 to 19.2)	0.60 (0.50–0.70)	0.87 (0.81–0.94)
≥33	50.0 (40–2–59.8)	91.0 (85.2 to 96.6)	61.0 (51.5–70.7)	14.0 (6.8 to 20.3)	0.61 (0.52–0.71)	0.86 (0.80–0.93)
≥34	45.0 (35.6–55.3)	94.0 (88.7 to 98.4)	67.0 (57.4–76.0)	14.0 (7.4 to 21.2)	0.67 (0.57–0.76)	0.86 (0.79–0.93)
≥35	41.0 (31.2–50.6)	94.0 (88.7 to 98.4)	64.0 (54.8–73.7)	15.0 (8.2 to 22.4)	0.64 (0.55–0.74)	0.85 (0.78–0.92)
≥36	41.0 (31.2–50.6)	94.0 (88.7 to 98.4)	64.0 (54.8–73.7)	15.0 (8.2 to 22.4)	0.64 (0.55–0.74)	0.85 (0.78–0.92)
≥37	32.0 (22.6–41.0)	94.0 (88.7 to 98.4)	58.0 (48.6–68.0)	17.0 (9.8 to 24.7)	0.58 (0.49–0.68)	0.83 (0.75–0.90)
≥38	32.0 (22.6–41.0)	95.0 (90.4 to 99.2)	64.0 (54.2–73.1)	17.0 (9.6 to 24.5)	0.64 (0.54–0.73)	0.83 (0.76–0.90)
≥39	27.0 (18.5–36.0)	95.0 (90.4 to 99.2)	60.0 (50.3–69.7)	18.0 (10.4 to 25.5)	0.60 (0.50–0.70)	0.82 (0.75–0.90)
≥40	23.0 (14.5–31.0)	96.0 (92.3 to 99.9)	63.0 (53.0–72.0)	19.0 (11.0 to 26.4)	0.63 (0.53–0.72)	0.81 (0.74–0.89)
≥41	14.0 (6.9–20.4)	96.0 (92.3 to 99.9)	50.0 (40.2–59.8)	20.0 (12.5 to 28.4)	0.50 (0.40–0.60)	0.80 (0.72–0.88)
≥42	14.0 (6.9–20.4)	96.0 (92.3 to 99.9)	50.0 (40.2–59.8)	20.0 (12.5 to 28.4)	0.50 (0.40–0.60)	0.80 (0.72–0.88)
≥43	14.0 (6.9–20.4)	97.0 (94.3 to 100)	60.0 (50.3–69.7)	20.0 (12.3 to 28.1)	0.60 (0.50–0.70)	0.80 (0.72–0.88)
≥44	5.0 (0.4–8.6)	99.0 (96.5 to 100)	50.0 (40.2–59.8)	22.0 (13.5 to 29.8)	0.50 (0.40–0.60)	0.78 (0.70–0.87)
≥45	0.0 (0.0–0.0)	100.0 (100 to 100)	-	22.0 (14.0 to 30.4)	-	0.78 (0.70–0.86)

Discussion

Summary

This is the first study to examine the validity and reliability performance of PHQ9 and MDI in patients with mixed chronic pain conditions referred to an interdisciplinary pain centre in Denmark.

According to the PSE interview, the prevalence of moderate or severe depression was 22.2%, and according to the PHQ9 and MDI questionnaires the prevalence was between 26.3 and 34.3% which is a somewhat lower

Table 5: 2 × 2 contingency table for the Major Depression Inventory (MDI) diagnostic algorithm and the Present State Examination (PSE).

Cohen's kappa coefficient: κ = 0.46 [95% CI: 0.27–0.64])		PSE	
		Depression	No depression
MDI	Depression	17	17
	No depression	5	60

prevalence compared with previous studies in similar settings using screening questionnaires [6]. However, the prevalence of depression in this study according to the PHQ9 questionnaire is very similar to a previous study in a larger sample in the same setting (between 23 and 27%) [7] suggesting that the prevalence of depression in patients with chronic pain attending specialist pain treatment may be quite different across countries.

The first aim of this study was to quantify the construct validity of the PHQ9 and the MDI assessing depression compared with a diagnostic interview in patients with chronic pain. Compared with a diagnostic interview the questionnaire sum-scores demonstrated AUC between 0.83 and 0.88 and the agreement in the classification of depression and no-depression between questionnaires and PSE was 73.7 and 77.8% with fair to moderate agreement. The second aim was to investigate the test-retest reliability of PHQ9 and MDI over two weeks. Both questionnaires showed excellent between-subject reliability with ICCs of 0.83 and 0.90. The results suggest that both the PHQ-9 and MDI sum-scores may be useful to reliably rule out current clinical depression in patients with chronic pain seeking specialist pain treatment.

For PHQ9, the performance characteristics were similar to previous findings in 542 patients with chronic spinal pain admitted to an interdisciplinary functional restoration program that showed an AUC of 0.77, a sensitivity of 66.2%, and a specificity of 74.4% for the sum-score with a cut-point of 13, and a diagnostic accuracy of 69.4% [12]. To our knowledge this is the first study investigating the relative reliability for the PHQ9 in patients with chronic pain. For MDI, the performance characteristics were is also similar to studies assessing depression in psychiatric patients. To our knowledge the validity as well as the relative reliability for the MDI in patients with mixed chronic pain conditions seeking specialist pain treatment has not previously been investigated. The relative test-retest reliability showed excellent ICC values. Recognizing that different clinical circumstances may require a cut-off choice that favors greater sensitivity or specificity compared to the best balanced sensitivity and specificity values in this study we reported the performance for all dichotomization thresholds as has previously been recommended for psychometric studies on depression questionnaires [23]. The PHQ-9 have previously been found to have acceptable diagnostic properties for detecting moderate depressive disorder with cut-off scores between 8 and 11, which are somewhat lower than the value of 14 observed in this study. However, it has been suggested that the accuracy of screening for depression is improved by using higher cut-off scores [24]. For MDI, the cut-off of 26 which had the best balance between sensitivity and specificity is in agreement with previous studies reporting 26 as cut-off for moderate depression [25].

Clinical implications

In assessment of patients with chronic pain, a failure to identify depression may result in reduced effectiveness of subsequent treatments. As previously recommended, the combination of a high sensitivity and negative predictive value is of greater importance than a high specificity and positive predictive values [26] when choosing a screening tool for depression so that cases are not missed. However, the use of questionnaires may also increase the risk of a false positive depression diagnosis as this patient group tend to have a higher score on these items than the normal population due to shared pain-depression symptomology, and the certainties of identifying patients who are likely to have depression are limited. This can possibly explain the high rates of depression in patients with chronic pain found in other studies using screening instruments [6]. The use of depression questionnaire scores or diagnostic

Table 6: Relative (between-subjects) two weeks test-retest reliability for the 9-items Patient-Health Questionnaire (PHQ9) and the Major Depression Inventory (MDI) sum-scores in patients with severe chronic pain. 'PHQ9': Patient Health Questionnaire 9-items. 'MDI': Major Depression Inventory. 'CI': confidence interval.

Variable	Test 1 Mean ± SD (95% CI)	Test 2 Mean ± SD (95% CI)	Absolute difference Mean ± SD (95% CI)	p-Value	Pearson r	$ICC_{3,1}$ (95% CI)
PHQ9 (*n*=99)	12.1 ± 5.6 (11.0–13.3)	12.7 ± 5.3 (11.6–13.7)	0.5 ± 4.2 (-0.3–1.4)	0.21	0.70 P < 0.001	0.83 (0.74–0.88)
MDI (*n*=99)	21.7 ± 10.5 (19.6–23.8)	23.4 ± 10.5 (21.3–25.5)	1.7 ± 6.1 (0.5–2.9)	0.007	0.83 P < 0.001	0.90 (0.85–0.94)

algorithms in patients with mixed chronic pain has a number of potential benefits including early and easily performed depression screening of people as a resource-efficient triage to inform clinical evaluation and management, and as an indication for referral to specialty mental evaluation for further assessment. Health care practitioners who assess and manage patients with chronic pain may welcome the usefulness of these questionnaire cut-offs to efficiently identify patients unlikely to require more detailed and time-consuming clinical investigations as the number of patients who need a more comprehensive mental health evaluation is significantly reduced.

Strengths and limitations

This study has a number of strengths. First, the PSE interview which was used as criterion standard is a validated instrument for diagnosing depression, and therefore considered as a good reference standard. Secondly, the interviewers were blinded to the questionnaire results reducing bias. Thirdly, the interviewers were trained pain specialists with extensive training and supervision in the use of PSE.

The study also has a number of limitations that need consideration. Firstly, very few or none of the patients had questionnaire scores in the extremes of the scales which may affect the range of prediction. Secondly, the sample size, although fairly large for validation studies may influence the ability to find a good balance between sensitivity and specificity as well as the reliability analysis as more than 100 patients are often recommended [27]. Thirdly, patients presented to the pain clinic with a mix of different pain conditions. As the prevalence of depression may vary across patients with different pain conditions, e.g. fibromyalgia, low back pain, migraine, arthritis this could affect the validity of the results. However, the sample is similar to other patients seeking specialist pain care in this setting that often report a mix of different pain conditions [28].

Conclusion

Validity as well as relative reliability was examined for the PHQ9 and MDI depression questionnaires. Compared with a diagnostic PSE interview, PHQ9 and MDI questionnaires reliably identified chronic pain patients unlikely to have depression, but showed limited validity identifying chronic pain patients with depression. These preliminary results may have clinical implications in depression screening in patients with chronic pain seeking specialist pain treatment. Clinicians in a specialty care pain clinic can use these screening questionnaires to identify patients without depression, but caution should be used when positive cases are identified by PHQ-9 or MDI due to the risk of false positives.

Research funding: This study was supported by the Pregraduate Research Foundation of Odense University Hospital, Denmark.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: Authors state no conflict of interest.

Informed consent: Written informed consent (electronic signature) was obtained from all patients included in this study.

Ethical approval: 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. No ethics approval was needed (Act on Research Ethics Review of Health Research Projects, October 2013, Section 14.2).

References

1. GBD. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015;386:743–800.
2. Sondergard S, Vaegter H, Erlangsen A, et al. Ten-year prevalence of mental disorders in patients presenting with chronic pain in secondary care: a register linkage cohort study. Eur J Pain 2018;22:346–54.
3. Hung C I, Liu C Y, Fu T S. Depression: an important factor associated with disability among patients with chronic low back pain. Int J Psychiatr Med 2015;49:187–98.
4. Bair M J, Robinson R L, Katon W., et al. Depression and pain comorbidity: a literature review. Arch Intern Med 2003;163:2433–45.
5. Poole H, White S, Blake C., et al. Depression in chronic pain patients: prevalence and measurement. Pain Pract 2009;9:173–80.
6. Rayner L, Hotopf M, Petkova H., et al. Depression in patients with chronic pain attending a specialised pain treatment centre: prevalence and impact on health care costs. Pain 2016;157:1472–9.
7. Vaegter H B, Handberg G, Kent P. Brief psychological screening questions can be useful for ruling out psychological conditions in patients with chronic pain. Clin J Pain 2018;34:113–21.
8. Vaegter H B, Group P W. PainData data rapport 2015–2018. [cited 16 December 2019]; 2019. Available from: http://ouh.dk/wm518981.

9. Gunzler D D, Perzynski A, Morris N., et al. Disentangling Multiple Sclerosis and depression: an adjusted depression screening score for patient-centered care. J Behav Med 2015;38:237–50.
10. Amris K, Omerovic E, Danneskiold-Samsoe B., et al. The validity of self-rating depression scales in patients with chronic widespread pain: a Rasch analysis of the Major Depression Inventory. Scand J Rheumatol 2016;45:236–46.
11. Bech P, Rasmussen N A, Olsen L R., et al. The sensitivity and specificity of the Major Depression Inventory, using the Present State Examination as the index of diagnostic validity. J Affect Disord 2001;66:159–64.
12. Choi Y, Mayer T G, Williams M J., et al. What is the best screening test for depression in chronic spinal pain patients?. Spine J 2014; 14:1175–82.
13. Tauben D Chronic pain management: measurement-based step care solutions. IAPS Pain Clinical Updates 2011;1–8. https://s3.amazonaws.com/rdcms-iasp/files/production/public/Content/ContentFolders/Publications2/PainClinicalUpdates/Archives/PCU_20-8_web.pdf.
14. Wing J K, Babor T, Brugha T., et al. SCAN. Schedules for clinical assessment in neuropsychiatry. Arch Gen Psychiatr 1990;47: 589–93.
15. Kroenke K, Spitzer R L, Williams J B. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16: 606–13.
16. Kroenke K, Spitzer R L, Williams J B., et al. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatr 2010;32:345–59.
17. Bair M J, Wu J, Damush T M., et al. Association of depression and anxiety alone and in combination with chronic musculoskeletal pain in primary care patients. Psychosom Med 2008;70:890–7.
18. Wittkampf K A, Naeije L, Schene A H., et al. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatr 2007;29:388–95.
19. Meader N, Mitchell A J, Chew-Graham C., et al. Case identification of depression in patients with chronic physical health problems: a diagnostic accuracy meta-analysis of 113 studies. Br J Gen Pract 2011;61:e808–20.
20. Cuijpers P, Dekker J, Noteboom A., et al. Sensitivity and specificity of the major depression inventory in outpatients. BMC Psychiatr 2007;7:39.
21. Cohen J. Statistical power analysis for the behavioural sciences. Vol. 2; 1988, Hillsdale, NJ: Erlbaum.
22. Fleiss J L. The design and analysis of clinical experiments. New York: Wiley; 1986.
23. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ (Can Med Assoc J) 2012;184: E191–6.
24. Kendrick T, Dowrick C, McBride A., et al. Management of depression in UK general practice in relation to scores on depression severity questionnaires: analysis of medical record data. BMJ 2009;338:b750.
25. Bech P, Timmerby N, Martiny K., et al. Psychometric evaluation of the major depression inventory (MDI) as depression severity scale using the LEAD (longitudinal expert assessment of all data) as index of validity. BMC Psychiatr 2015;15:190.
26. Ren Y, Yang H, Browning C., et al. Performance of screening tools in detecting major depressive disorder among patients with coronary heart disease: a systematic review. Med Sci Mon Int Med J Exp Clin Res 2015;21:646–53.
27. Terwee C B, Mokkink L B, Knol D L., et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651–7.
28. Vaegter H B, Graven-Nielsen T. Pain modulatory phenotypes differentiate subgroups with different clinical and experimental pain sensitivity. Pain 2016;157:1480–8.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/sjpain-2020-0042).