

Validity of Current Assessment Tools Aiming to Measure the Affective Component of Pain: A Systematic Review

Anders Heiberg Agerbeck¹⁻³
 Frederik Handberg Juul
 Martiny^{1,2}
 Christian Patrick Jauernik^{1,2,*}
 Karin Due Bruun^{3,4,*}
 Or Joseph Rahbek^{1,2,*}
 Kristine H Bissenbakker^{1,2}
 John Brodersen^{1,2}

¹The Section of General Practice and Research Unit for General Practice in Copenhagen, Copenhagen, Denmark; ²The Research Unit for General Practice in Region Zealand, Copenhagen, Denmark; ³Pain Research Group, Pain Centre, Odense University Hospital, Odense, Denmark; ⁴Department of Clinical Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark

*These authors contributed equally to this work

Abstract: The objective of this study was to identify patient-reported outcome measures (PROMs), which aim to measure the affective component of pain and to assess their content validity, unidimensionality, measurement invariance, and Internal consistency in patients with chronic pain. The study was reported according to the PRISMA guidelines. A protocol of the review was submitted to PROSPERO before data extraction. Eligible studies were any type of study that investigated at least one of the domains: PROM development, content validity, dimensionality, internal consistency, or measurement invariance of any type of scale that claimed to measure the affective component of pain among patients with chronic pain. The databases Medline, Embase, PsycINFO, and the Cochrane Library were searched for eligible studies. The database search was supplemented by looking for relevant articles in the reference list of included studies, ie backtracking. All included studies were assessed independently by two authors according to the “COSMIN methodology on Systematic Reviews of Patient-Reported Outcome Measures”. Descriptive data synthesis of the identified PROMs was conducted. The search yielded 11,242 titles of which 283 were assessed at the full-text level. Full-text screening led to the inclusion of 11 studies and an additional 28 studies were identified via backtracking, leading to the inclusion of 39 studies in total in the review. Included studies described the development and validity of 10 unique PROMs, all of which we assessed to have potentially inadequate content validity and doubtful psychometric properties. No studies reported whether the PROMs possessed invariant measurement properties. The existing PROMs measuring affective components of chronic pain potentially lack content validity and have inadequate psychometric measurement properties. There is a need for new PROMs measuring the affective component of chronic pain that possess high content validity and adequate psychometric measurement properties.

Keywords: PROMs, content validity, psychometrics, chronic pain, COSMIN

Introduction

The prevalence of chronic pain is high globally, with estimates ranging from 10% to 20% of the adult population, and the number of people with self-reported chronic pain is rising.^{1,2} Chronic pain has many consequences, both for society eg increased healthcare expenditures, and for the individual eg psychosocial consequences, activity impairment and reduced work productivity.^{3,4}

Conventional therapy for pain relief (opioids and/or surgery) primarily targets the sensory component of pain. However, treatment of chronic pain with opioids is

Correspondence: Anders Heiberg Agerbeck Strandvej 61, Svendborg, 5700, Denmark
 Tel +4520894562
 Email anders.heiberg.agerbeck@rsyd.dk

associated with the risk of several adverse effects, eg addiction, constipation, fatigue, and dizziness, often without providing sufficient pain relief.^{5,6}

By utilizing various psychological therapies such as Cognitive Behavioral Therapy (CBT), Acceptance and Commitment Therapy (ACT), or Mindfulness, multidisciplinary approaches offer an alternative to chronic pain management aiming to relieve the emotional distress caused by chronic pain, ie, the affective component of pain, thus helping patients both to accept their pain and to develop more adaptive coping strategies.⁷

As an example of this, following the completion of an ACT or mindfulness treatment course, meaningful changes in the patient's pain-related distress might have occurred even though the level of pain and functioning remained unchanged.

Current evidence for multidisciplinary approaches with psychological interventions is promising; however, most of the evidence is of low quality and many trials report heterogeneous results.⁸ Despite existing literature showing beneficial effects of psychological interventions, a recent review with meta-analysis has shown only small to medium size effects of treatment with ACT or mindfulness and only on some outcomes, namely anxiety and pain interference with activities of daily living.⁹ Another recent systematic review and meta-analysis showed similar results, addressing the need for additional studies with robust methodology as research in chronic pain management was limited by the lack of a biological variable to measure effect sizes. The discrepancy of the trial results may in part be due to the lack of an appropriate tool that can comprehensively and accurately measure the effect of psychological interventions.¹⁰

This underpins the importance of accurate measurement tools for research in chronic pain management.

The IMPAACT recommendations define self-reported measures as the “gold standard” to measure outcomes in clinical trials of pain interventions. A core set of outcomes consisting of six outcome domains has been outlined including 1) pain; 2) physical functioning; 3) emotional functioning; 4) participants ratings of global improvement and satisfaction with treatment; 5) adverse events, and 6) participant disposition.¹¹ Chronic pain trials often measure emotional functioning using generic patient-reported outcome measures (PROMs), developed for evaluating depression and anxiety such as Hospital Anxiety and Depression Scale (HAD).¹² Even though depression and anxiety scales to some degree might

reflect the emotional distress experienced by patients with chronic pain, these PROMs have not been developed specifically for patients with chronic pain or to measure the construct of emotional distress caused by chronic pain.

It is generally acknowledged that chronic pain is a multidimensional construct, which cannot be defined solely by pain intensity.¹³ As described in the biopsychosocial model of chronic pain, psychological factors such as personality, attitude, and past experience influence the perception of a noxious stimulus.¹⁴ In line with this, Harkins' Affective-motivational model of pain proposes the affective component of pain as the end product of the sensation of pain combined with arousal and cognitive appraisal of the pain.¹⁵

Several pain-specific PROMs have been developed to measure the affective component of pain. Such a PROM developed for patients with chronic pain could possibly measure outcome effects of psychological interventions to reduce the affective/emotional components of pain, as the quality of the affective distress caused by pain may differ from the distress measured by generic measurement tools.

The first PROM to acknowledge pain as a distinct phenomenon that transcends sensory input was the McGill Pain Questionnaire developed in 1975.¹⁶ In this PROM the severity of pain was defined by sensory, affective, and evaluative (or cognitive) factors. Since then a plethora of both generic and disease-specific PROMs have emerged, identifying different factors deemed relevant for the measurement of pain and the impact of pain on functioning, work, mental health, and more. It is to the authors' knowledge, however, unknown how many PROMs have been designed specifically to measure the affective component of pain for patients with chronic pain, or whether they are valid by modern standards.

Although the IMPAACT recommendations have assessed and defined outcome domains of interest in pain trials, and outcome measurement tools deemed appropriate for the measurement of the affective component of pain, none of these have to the authors' knowledge been assessed by the rigorous standards of COSMIN. Furthermore, it is unclear if other scales have been developed in the meantime, following modern standards of PROM development and psychometric validation.

Regarding the validity of PROMs, the COSMIN initiative has developed a long list of criteria to assess whether an existing PROM is a valid evaluative tool.

COSMIN is an abbreviation of “Consensus-based standards for the selection of health measurement instruments.” The COSMIN initiative aims to advance the science and application of health outcome measurement. In this regard, the COSMIN group has developed an extensive checklist to use as a guideline for conducting systematic reviews of PROMs. The checklist was published in 2018 and has found widespread appeal due to its applicability and rigorous standards.¹⁷ The recommendations of a PROM should be based on an evaluation of its Content validity, Dimensionality (named Structural Validity by COSMIN), Internal consistency, and analysis of invariant measurement. Of these criteria, content validity is the essential validity criterion: if a PROM lacks content validity, no statistical psychometric model can change this fact: “garbage-in, garbage-out”. If adequate content validity is not ensured, there is a risk of excluding factors important to patients, or imprecise measurement due to misinterpretation of items.¹⁸

In view of the above, this study aims to 1) identify PROMs that measure the affective dimension of pain in patients with chronic pain, and 2) to assess the adequacy of the identified PROMs regarding content validity, dimensionality, internal consistency, and measurement invariance.

Methods

This systematic review was conducted according to the COSMIN manual for systematic reviews of PROMs and reported according to the PRISMA checklist for systematic reviews.^{17,19}

The “COSMIN methodology of systematic reviews of Patient-Reported Outcome Measures” was used to assess the validity criteria of the included PROMs.^{17,20}

A protocol of the systematic review was uploaded to PROSPERO prior to data extraction. The protocol is accessible at https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=131908

Search Strategy and Information Sources

All details of the search strategy are published in the study protocol. The search strategy was developed for PubMed in cooperation with an information specialist and consisted of three columns, combined with the Boolean operator “AND”, representing the construct, population, and type of instrument. In each of these three columns, synonyms were combined with the Boolean operator “OR” to increase the sensitivity of the search strategy: In the “construct”,

“population”, and “type of instrument” columns the terms “affective component of pain”, “chronic pain” and “PROM” were expanded respectively. Lastly, a search filter for PROM development and validity studies was added to the search as a fourth column combined with the Boolean operator “AND” to increase the specificity of the search strategy.²¹

Subsequently, the search strategy was adapted for the databases Embase, PsycINFO, and the Cochrane Library. Additional databases exist, however we limited our search to these databases because of their comprehensive content, making it likely that we would not identify further studies by searching other databases. This judgment was based on the experience of the authors and recommendations from the information specialist. We applied no restrictions concerning date, language, or study design. The final search in the above-mentioned databases was conducted on the 26th of September 2018. All studies identified were compiled in Endnote where duplicates were removed. Furthermore, the first author scrutinized reference lists of included studies to identify studies potentially missed by the search strategy, ie backtracking. Also, the first author did citation tracking: After identifying eligible PROM development studies, these were searched in Google Scholar, combined with keywords for measurement properties including “validity”, “unidimensionality”, “factor analysis”, “invariant”, “differential item function”, “internal consistency” and “Cronbach” to find additional eligible studies. Because PROMs are often cited in multiple publications, indexed in the databases we searched, we did not conduct a grey literature search. The full search strategy is available in [Appendix 1 – Search strategy](#).

Eligibility Criteria

We included all studies investigating at least one of the domains: PROM development, Content Validity, Dimensionality, Internal consistency, or measurement invariance (also defined as Differential Item Function (DIF)) of any scale measuring the affective component of pain. Studies that only used the PROM as an outcome measure were excluded. Only full-text studies were included. PROMs concerning children aged less than 18 years old were excluded. A full list of inclusion and exclusion criteria is available in [Appendix 2 – Eligibility criteria](#).

Study Selection

Identified studies were screened independently by two authors at title, abstract, and full-text levels according to the pre-defined eligibility criteria. Any discrepancies were

resolved by discussion. In cases where consensus could not be reached, a third author was consulted. All studies excluded through full-text screening were noted, with reasons for exclusion.

Data Collection Process and Items

Data extraction was pre-specified in the protocol. For each included study, data extraction included: a) sample population, eg mean age, gender distribution, disease, disease duration, and b) psychometric data analyses relevant to the aim of the review. The main author of the review conducted the data extraction.

Risk of Bias and Summary Measures

Using the COSMIN Risk of Bias checklist, two authors independently assessed the PROM development, Content validity, Structural validity (Dimensionality), Internal consistency, and measurement invariance. Several items assess each domain, and every item on the list is rated on a scale of “very good”, “adequate”, “doubtful”, or “inadequate”. Any disagreements were settled through discussion, and in case agreement could not be achieved, a third author was consulted to reach consensus.

The COSMIN checklist assesses the level of patient involvement in the PROM development phase to evaluate content validity by the degree of content relevance, content coverage (comprehensiveness), and understandability (comprehensibility) of the PROM in the target population.

Regarding dimensionality of factors, they can be assumed by the developers to be either multi- or unidimensional. The factor dimensionality can be examined by factor analysis. The COSMIN checklist considers whether the factor structure of the PROM has been validated by exploratory or confirmatory factor analyses (EFA or CFA) or Item Response Theory (IRT) models, and the adequacy of the sample size. Regarding measurement invariance, the checklist considers whether this has been established with either DIF analyses or CFA.

Regarding internal consistency, the COSMIN checklist considers whether Cronbach’s alpha or omega has been calculated.

Even though the checklist also assesses the reliability of the PROMs, the aforementioned psychometric qualities are considered key, and further analysis of reliability is therefore not the scope of this study. We have chosen to include an assessment of internal consistency as some studies use this as a way to confirm dimensionality.

The COSMIN Risk of bias checklist employs a “worst score counts” method to assess the internal validity (risk of bias) for each domain, as all the dimensions of the domain are necessary for the adequacy of measurement, eg PROM development is rated inadequate if the items regarding relevance are inadequate, irrespective of the assessment of comprehensiveness and understandability.^{17,20}

Results

Study Selection and Characteristics

The search yielded 11,242 articles after duplicates had been removed. These were screened at title/abstract level leading to full-text scrutiny of 283 studies concerning 101 different PROMs. Eleven of these studies were included in the review and an additional 28 studies were identified via references and citation tracking. In total, 39 studies were included, covering development and validity studies of 10 different PROMs measuring the affective component of pain.

A full account of the study selection process and flow-chart are available in [Figure 1](#).

Details regarding study characteristics are available in [Appendix 3](#).

The following is a description of the main results from the analysis of included PROMs. Analyses on invariant measurement were not conducted in any of the 10 identified PROMs. The main findings concerning the content validity and measurement invariance analyses are summarized in [Table 1](#).

Risk of Bias Within Studies

Glasgow Pain Questionnaire (GPQ)

PROM Development and Content Validity

The GPQ was developed to evaluate self-rated pain in community studies. It was designed to measure not only the experience of pain but also the impact pain has on daily life. It consists of the subscales “pain frequency”, “intensity”, “emotional reactions”, “ability to cope”, and “restrictions of daily activities”.

The items were generated by 230 unstructured interviews with members of the public in ten locations. The interviews yielded over 5,000 descriptors, which were divided into broad categories. A total of 59 items was derived from the item pool and tested on 60 healthy volunteers, who were asked to report “yes” or “no” to having pain representing each item in the last month. Hereafter the authors excluded items with low responder

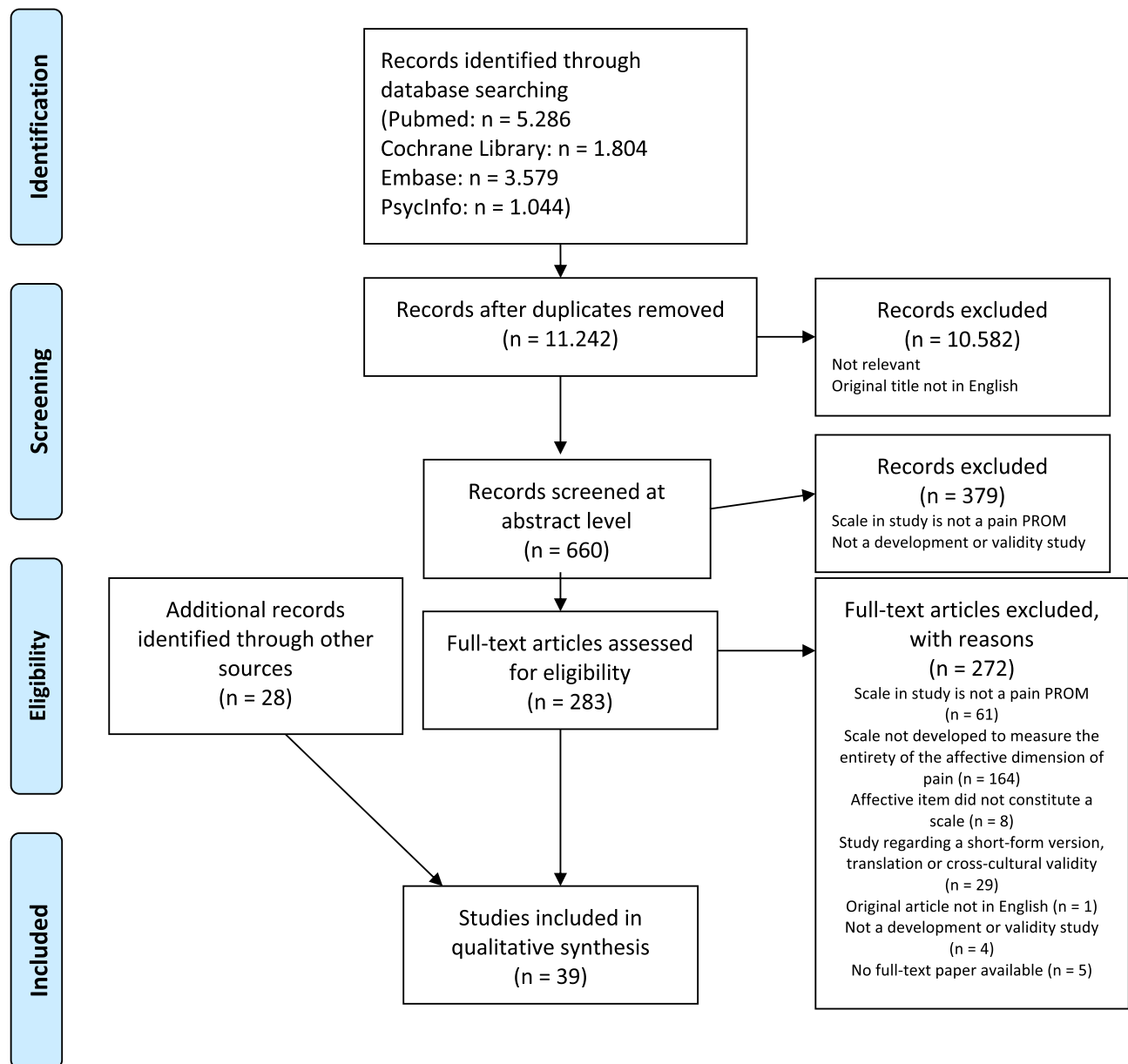


Figure 1 PRISMA flow diagram.

Note: Copied from Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(7): e1000097.

rates and items that the respondents found ambiguous or difficult to understand by the respondents. Finally, the PROM was tested by administering it to three different patient samples and comparing their scores.²²

Conclusion

It is doubtful whether the method of item selection adequately selected relevant items for the target population and it is unclear whether the participants of the pilot test were

asked about the content coverage of the PROM. It was developed for community studies, and even though this is a heterogeneous target population, no information on the age or gender of the study populations was given. Therefore, it is unclear whether the study populations, from concept elicitation to the final test of the PROM, can adequately fit any target population. No cognitive interviews regarding the final set of items were conducted. In summary, the content validity is rated as inadequate. Furthermore, no psychometric

Table I Content Validity and Measurement Invariance

PROM	Abbreviation	Construct	Factors	Content Relevance	Content Coverage	Understandability	Measurement Invariance Analyses
Glasgow Pain Questionnaire	GPQ	Self-rated pain	Frequency, intensity, emotional reaction, ability to cope, restriction of daily life	Interviews	Not reported	Pilot test	Not reported
Global Pain Scale	GPS	Pain	Pain, feelings, clinical outcomes, activities	Not reported	Not reported	Not reported	Not reported
Multidimensional Affect and Pain Survey	MAPS	Pain	Severe sensory pain, negative emotions, moderate sensory pain, good health, manageable illness, numb/cold	Not reported	Not reported	Not reported	Not reported
McGill Pain Questionnaire	MPQ	Pain quality	Sensory, affective, evaluative	Not reported	Not reported	Not reported	Not reported
Multiperspective Multidimensional Pain Assessment Protocol	MMPAP	All domains of chronic pain	Pain description, functional abilities, ADL, environmental components, mood alteration, emotional descriptors, family history, significant other actions, effect on cognitive processes, functional repetitions, pain interference with sleep	Not reported	Not reported	Not reported	Not reported
Multidimensional Pain Inventory	MPI	Subjective experience of pain	Interference, support, severity, self-control, negative mood, activities, significant other responses	Not reported	Not reported	Not reported	Not reported
NIH PROMIS – Pain Quality	PROMIS – PQ	Pain Quality	Pulling/tugging, tingling/ numbness, sharp/stabbing, dull/aching, pounding/ pulsing, affective	Not clear	Focus groups	Cognitive interviews	Not reported
Pain Discomfort Scale	PDS	The affective dimension of pain	Pain intensity, pain affect	Survey	Not reported	Not reported	Not reported
Profile of Chronic Pain: Screen	PCP: S	Chronic pain	Pain severity, interference, emotional burden	Not reported	Not reported	Not reported	Not reported
Vulvar Pain Assessment Questionnaire	VPAQ	Chronic vulvar pain	Pain severity, emotional response, cognitive response, life interference, sexual function interference, self-stimulation/penetration interference	Not reported	Not reported	Not reported	Not reported

analyses were conducted on data collected with the GPQ including testing of unidimensionality.

Global Pain Scale (GPS)

PROM Development and Content Validity

The GPS is based on a biopsychosocial model of pain and was developed as a screening tool and outcome measure to assess physical pain, affective effects of pain, specific clinical outcomes, and interference with daily activity. The scale consists of the subscales “your pain”, “your feelings”, “clinical outcomes”, and “your activity”.

The items were developed by the authors, organized to a scale, and tested on 262 undergraduate university students with self-reported chronic pain.

Psychometric Properties

The scale was tested for construct and criterion validity by correlation analyses with the short-form MPQ, MPI, and Perceived Stress Scale. A CFA was conducted to ensure unidimensionality and internal consistency was calculated with Cronbach’s Alpha with a value of 0.84 for the “your feelings” subscale.²³

Conclusion

The developers of the GPS clearly described the construct, target population, and context of use; however, as they generated the items without any description of patient involvement, the scale is rated to have inadequate content validity.

The a priori factor structures were confirmed by factor analysis, but only university students with self-reported pain were included in the psychometric testing. It is unknown, therefore, if GPS will possess the same measurement properties in other populations.

Multidimensional Affect and Pain Survey (MAPS)

PROM Development and Content Validity

MAPS was developed as a pain measurement tool consisting of 101 descriptors of pain and emotions. The descriptors were selected from checklists of psychological symptomatology and the initial set of 270 words described by Melzack and Torgerson in 1971.²⁴

Healthy volunteers sorted these items using a pile-sort method into groups that contained items thematically related to one another.²⁵

Psychometric Properties

A cluster analysis was performed organizing the sorted items in a dendrogram, after which redundant items were removed. Subsequently, a group of 104 healthy volunteers

from three different ethnic groups, divided into six sex-ethnocultural groups, sorted the remaining items using the same pile-sort method. Another cluster analysis determined the final dendrogram structure of the MAPS after the removal of yet another 34 redundant items and 54 items that had different meanings in the different groups.

The final questionnaire was subjected to principal components analysis. The analysis yielded 6 factors, with a distinct affective factor named “negative emotions”.

Conclusion

The PROM development and content validity were rated inadequate, as the authors did not clearly describe the construct to be measured, the context for use of the questionnaire, or the target population. Furthermore, there was no description of patients being asked about content relevance, content coverage, or understandability. Healthy volunteers sorted the items and a factor analysis was conducted on outpatients with cancer, raising doubt as to whether the PROM was tested on the population for which it was intended.

EFA was performed to prove unidimensionality but on an inadequate sample size.

McGill Pain Questionnaire (MPQ)

PROM Development and Content Validity

The MPQ was developed by Melzack in 1977 and consists of 78 pain adjectives categorized into 20 subclasses dividing the pain experience across three factors: sensory, affective, and evaluative. The MPQ is based on “The Gate Control Theory” by Melzack and Wall.²⁶

Melzack and Torgerson developed the items for the MPQ in a preliminary study. In this study adjectives describing pain were found through literature review and subsequently arranged in their subclasses by the authors. The subclasses were presented to 20 university students who were to assess if they deemed that an adjective fitted the assigned subclass. Subsequently, the adjectives were sorted by intensity rating in each subclass by a sample of psychology students, physicians, and patients.^{16,24}

Psychometric Properties

The development study aimed to categorize pain adjectives in distinct factors consisting of related adjectives. However, no formal tests for dimensionality were conducted in the development studies. Nine studies have investigated the factor structure of the MPQ with EFA in different pain populations leading to various factor

solutions. Only one of these studies derived an affective component of pain factor.²⁷

Five studies have attempted to confirm the original factor structure of the MPQ with CFA. Two of these studies confirmed the original structure.^{28,29} The findings of one of these two studies were disregarded due to inadequate P-value of the factor structure.²⁹

One study conducted an EFA, leading to a 4-factor solution, including an affective factor, which was confirmed by subsequent CFA in a sample of patients with lower back pain.³⁰

Conclusion

The content validity was rated inadequate as no clear description of the target population was provided. Twenty patients were included in the intensity rating of the adjectives; however, no report was given as to whether they were asked if the adjectives were relevant, understandable, or comprehensively covered their experiences of pain. Furthermore, no cognitive interviews of the final set of items were conducted. Although several studies examined the factor structure, only a few of them have adequately confirmed the unidimensionality of an affective component of pain factor with different results.

Multiperspective Multidimensional Pain Assessment Protocol (MMPAP)

PROM Development and Content Validity

The MMPAP was developed in an effort to standardize the assessment of pain in patients with chronic pain. It aimed to provide a comprehensive protocol for physical examination, physician evaluation, and patient-reported outcome assessment.

The items were generated by literature and expert panel review. The final set of items was pilot tested on a sample consisting of 67 patients reporting pain for at least 6 months. No further details on item development were reported, nor was there any mention of the number of items.

Psychometric Properties

A principal-factor analysis with varimax rotation was conducted on the entire sample plus analyses of Cronbach's alpha with a value of 0.85 on the subscale "effect on emotional status".³¹

Conclusion

The patient-reported outcome part of the MMPAP is based

on an inadequate development process, as no report of patient involvement in item development was mentioned. Even though the MMPAP was subjected to pilot testing, no mention was made of patient input or revisions as a consequence of the pilot test.

The factor structure was created by expert panel recommendations and modified following an EFA. However, the factor loadings were not reported. As the number of items was not mentioned, the test of unidimensionality is rated doubtful due to uncertainty about the sufficiency of the sample size.

Multidimensional Pain Inventory (MPI)

PROM Development and Content Validity

The MPI was based on a cognitive-behavioral approach to chronic pain and was developed as an assessment tool for the evaluation of treatment approaches. The scale has three parts, which were developed a priori, where the first part measures 6 general concepts: pain severity and suffering, pain-related life interference, dissatisfaction with present level of functioning, appraisal of support from significant others, perceived life control, and affective distress. The second part regards patients' perceptions of the responses of their significant others, and the third part measures activities.

No patient input was reported in the development of the individual items or scales and no pilot test of the final set of items was recorded.³²

Psychometric Properties

The a priori defined factors were confirmed with CFA in a sample of 120, mainly male, veterans of the United States Armed Services.

Additional studies throughout the following thirty years have assessed the original factor structure with diverging results. Several studies have performed EFAs deriving a different set of factor solutions than the original study. None of them support the unidimensionality of the affective distress subscale.^{33–35} Two studies pooled the MPI with other self-report instruments in order to derive and confirm underlying factors of chronic pain. These studies could not confirm the unidimensionality of the affective distress subscale.^{36,37}

None of the CFAs conducted have confirmed the original factor structure proposed by the developers.^{38,39} Internal consistency was calculated with Cronbach's alpha with a score of 0.79 on the affective distress

subscale. A subsequent study by Wittink et al calculated similar results.⁴⁰

Conclusion

The content validity of the MPI is rated as inadequate due to lack of patient input in the item development phase. Furthermore, no pilot test or cognitive interviews of the final set of items were conducted. The a priori defined factor structure was confirmed by factor analysis; however, it was rated doubtful due to the small sample size. This doubt is enhanced by subsequent studies, where the original factor structure could not be confirmed; additionally, most of these studies are rated inadequate or doubtful due to small sample sizes.

National Institute of Health (NIH) Patient-Reported Outcome Measurement Information System (PROMIS) Pain Quality Item Bank

PROM Development and Content Validity

The PROMIS item bank was developed to provide an efficient and informative item set for the evaluation of a variety of domains of latent traits, including pain.

The items were identified through literature search, yielding 644 pain items, that were subsequently sorted in a process termed “binning and winnowing” wherein items are categorized and subsequently excluded if they are either redundant or deemed inconsistent with their assigned category. Subsequently, PROMIS Network PROM experts revised the remaining items aiming to ensure comprehensibility of items, uniformity of response options, and recall time. Focus groups were utilized to identify gaps in item coverage and to confirm the definitions of PROMIS categories. Patients with a wide variety of illnesses were recruited from different settings. Finally, the developers performed cognitive interviews to ascertain the understandability of the items, recollection of relevant information from the participants, and response options.⁴¹

Psychometric Properties

A subsequent study conducted EFA on pain quality items in the PROMIS item bank. The study derived 6 factors including an affective component of pain factor. This factor structure was confirmed by CFA. The analyses were performed on a mixed sample of general population participants and patients with chronic pain.⁴²

Conclusion

No clear description of the construct to be measured or the target population was described in the item development

study. The participants in the cognitive interviews were recruited from a different setting than the participants in the focus groups. Furthermore, as the developers did not define a target population for the item set, it is doubtful whether the included samples in the focus groups or cognitive interviews are representative.

Two to four focus groups were conducted for each domain. It is unclear if data saturation was reached or to what extent new information provided by the participants was incorporated in the final set of items.

The study does not state the number of pain items that were tested in the cognitive interviews, making it doubtful whether the sample size of 44 participants was adequate. Given the methodological issues outlined above, the content validity is rated inadequate.

The factor solution was derived with EFA and confirmed by CFA. The evidence of unidimensionality was rated inadequate due to insufficient sample size.

Pain Discomfort Scale (PDS)

PROM Development and Content Validity

The PDS was developed to provide a brief measurement tool to distinguish pain-affect from other dimensions of pain. Sixteen items were generated based on patient statements from a chronic pain program, and they were tested for relevance in a survey with 59 patients with chronic pain. The patients were asked to rate relevance on a five-point Likert scale. Ten items were retained in the final PROM.

Psychometric Properties

A principal axis factor analysis was conducted on the PDS along with the affective subscale of MPQ, Beck Depression Inventory, and four measures of pain intensity (VAS, NRS, BS-11, PPI) revealing the two factors “pain intensity” and “pain affect”. The PDS was tested for internal consistency with a Cronbach’s alpha of 0.77.⁴³

Conclusion

Even though the construct that the PDS intends to measure is clearly described, no definition of the target population or context for use is provided. The item generation process was not clearly described since no information on the original 16 items was provided. The items were tested quantitatively for content relevance; however, no final test of the retained items was conducted and there was no reported test for content coverage or understandability. The PROM development, therefore, is rated inadequate. In

addition, the included sample size was inadequate to prove the unidimensionality of the two factors revealed by the EFA.

Profile of Chronic Pain: Screen (PCP:S)

PROM Development and Content Validity

The PCP:S was based on a multidimensional conceptual model of chronic pain and developed as a brief screening tool to assess the impact of chronic pain in the non-clinic-referred general population. The authors considered that the dimensions “severity”, “interference” and “emotional burden” were the key aspects of pain to be measured.

The items were generated by a review of the literature describing the aforementioned constructs.

Psychometric Properties

The scale was tested psychometrically with CFA and Cronbach’s alpha in a national sample of 2449 adults.⁴⁴

A subsequent study investigated the factor structure and internal consistency of the PCP:S in a sample of 244 patients from primary care living with chronic pain.⁴⁵

The CFA confirmed the original factor structure in the sample of primary care pain patients, and Cronbach’s alpha was calculated for all subscales with a value of 0.91 for the “emotional burden” subscale.

Conclusion

The PCP:S was developed to assess chronic pain in a well-described population and context; however, the item development phase was not reported. As a result thereof it is unclear whether patients were involved in the item development phase and whether the method used was adequate to ensure content relevance, content coverage, and understandability. Content validity, therefore, is rated inadequate.

The CFA confirmed the unidimensionality of the “emotional burden” subscale.

The Vulvar Pain Assessment Questionnaire (VPAQ)

PROM Development and Content Validity

The VPAQ was developed to provide clinicians with a brief, yet comprehensive, questionnaire for assessing vulvar pain. The authors stated that some of the subscales might apply to other pain categories. The questionnaire consists of 55 items divided into 6 subscales, including one that considers the affective component of pain.

The items were identified from literature review, a review of similar measures, and lay websites. The selected set of items was reviewed for relevance, content

coverage, and understandability by an expert panel consisting of four health professionals and one layperson with personal experience of vulvodynia. The questionnaire was pilot tested in a small group of students and community members known to the first author. There was no report of whether anyone in the pilot test had vulvodynia.⁴⁶

Psychometric Properties

Six factors were identified after examining the item distributions for normality, as well as maximum likelihood factor analyses with an oblique rotation and listwise deletion.

A subsequent study of the VPAQ conducted exploratory structural equation modeling. This analysis confirmed the fit of the original factor structure of all the subscales, except the “coping” subscale. Both studies calculated Cronbach’s Alpha for each subscale with values ranging from 0.77 to 0.89. A specified list of the values for each subscale was not provided.⁴⁷

Conclusion

Even though a clear description of the construct and target population was provided, the PROM development was rated inadequate as, a) only one patient was included in the item development phase; b) whether the patient input had any consequence for the composition of the final set of items is not described, and c) the pilot test was not conducted in a sample representing the target population.

The unidimensionality of the factors was established in both studies with EFA based on inadequate sample sizes.

Discussion

Summary of Evidence

This review identified studies of 10 PROMs that aimed to measure the affective component of pain. Scrutiny of the methodology in the studies showed that none of the identified PROMs had adequate content validity according to the COSMIN manual. Seven development studies did not report assessments of any of the three components of content validity (relevance, comprehensiveness, understandability); and none of the 10 assessed all three components, rendering the content validity unknown, and thus inadequate as defined by COSMIN. The number of studies that are cited for each included PROM in the results section varied because some PROMs had been validated or investigated in other ways in and more settings than other PROMs.

The dimensionality of the PROMs was most often validated by EFAs. Most of these validity studies concerned the unidimensionality of the MPQ and MPI and were unable to confirm their original factor structure. None of the items in the 10 identified PROMs' affective scales have been tested for invariant measurement properties.

Cronbach's alpha was estimated for six of the ten PROMs; however, if unidimensionality is not established before Cronbach's alpha is calculated, it is hard to interpret the value of Cronbach's alpha since unidimensionality is the prerequisite for a meaningful interpretation of Cronbach's alpha.

The COSMIN manual was published in 2018 and the 10 scrutinized PROMs were all developed prior to this manual that emphasizes the importance of content validity. Even so, content validity is considered the key measurement property of a PROM to ensure that we know what we are measuring.⁴⁸ Therefore, the results from previous studies using these 10 PROMs might be questionable. In addition, most of the investigated PROMs also had inadequate or doubtful psychometric measurement properties. This can lead to over or underestimation when utilizing the PROM, due to measurement invariance: the risk that the same construct is not measured equally among people with eg different diagnoses, cultures, languages, or genders.⁴⁹

Contemporary literature recommends the use of PROMs as an outcome measurement in clinical trials studying patients with chronic pain, highlighting the MPQ and MPI as well-documented and widely used outcome measures.^{11,50} Our search revealed no systematic reviews of pain PROMs investigating the affective component of pain. However, other reviews have assessed the content validity of pain PROMs concerning other domains of pain, finding inadequate content validity.^{51,52} Hence, this review adds to the list of systematic reviews demonstrating inadequate content validity of PROMS in pain measurement.

The main result of the present systematic review was the lack of patient involvement in the item development phase along with insufficient and imprecise definitions of target populations. This might lead to a poor distinction between chronic pain in different patient groups when applying the questionnaires in a clinical setting, eg it is plausible that patients with lower back pain have a distinctively different experience of pain than patients with chronic pancreatitis.⁵³ This problem could be aggravated by the novel classification of chronic pain in ICD-

11, where emphasis is placed on the distinction between different subgroups of patients with chronic pain. The present review demonstrates that we do not have a validated tool to make such a distinction.

This calls for PROMs that can encompass well-defined latent constructs with high-quality evidence: high content validity and adequate psychometric measurement properties targeting specific (sub)-populations of patients with pain.

Strengths and Limitations

The main strength of this review is the composition of the review team including experts on content, ie pain, and methodology, eg systematic review and PROM methodology. This ensured adherence to the criteria of systematic reviews defined by PRISMA and to the rigorous assessment protocol of COSMIN.

The COSMIN checklist is a manageable tool that provides a detailed and systematic assessment of the PROM development process: the importance of high content validity is emphasized, as all PROMs must be based on a thorough and detailed development process incorporating members of the target population through qualitative methods. If the development process is rated as inadequate, the psychometric properties regardless of how they are conducted, are implicitly inadequate. However, regarding the assessment of dimensionality, the COSMIN checklist has some shortcomings. Exploratory and confirmatory factor analyses receive almost identical ratings: "adequate" and "very good", respectively; and CFA receives the same rating as Item Response Theory (IRT) and Rasch models with a good fit. However, EFAs are not confirmatory tests and are thus unable to confirm an alleged factor structure. Exploratory factor analyses are able to generate hypotheses of factors, extracted from the data and not from patient interviews, but they cannot confirm the unidimensionality of items in one or more qualitatively hypothesized domains developed from interviews with patients. Furthermore, factors generated by data without direct patient input are arguably inferior to factors identified through qualitative methods. Considering CFA the COSMIN checklist does not assess whether a p-value, with its corresponding confidence interval, has been reported or if the p-value is adequate, which can raise doubts about the adequacy of the CFA. Item Response Theory and Rasch analysis are rated equally in the COSMIN checklist. However, we consider Rasch analysis as more strict than other IRT models, as Rasch models are

the only models which ensure invariant measurement (called specific objectivity by G. Rasch) and sufficiency.^{54,55} The Non-Rasch IRT models and the CFA can all provide evidence of unidimensionality, but additional analyses on invariant measurement (differential item function or DIF) are needed if these psychometric analyses are used.^{49,56}

Of the included articles in this review, 11 (27,5%) of these were identified through the initial search strategy, whereas citation tracking identified the remaining 29 (72,5%) articles. This could be considered a limitation of the review. However, the search strategy was developed in cooperation with an information specialist at a university library, followed by the procedure outlined by PRISMA and COSMIN. Of note, most of the included PROMs were developed 20–40 years ago, which might explain why their development studies did not appear in the original search due to different MeSH term registration.

This review only included PROMs measuring the affective dimension of pain. This might be considered a limitation, as many generic PROMs measuring emotional burden exist and these are used along with pain questionnaires in clinical settings and research.¹¹ The definition of the affective or emotional component of pain is unclear. IMPACT recommendations distinguish between the emotional function of pain patients and the affective quality of pain.¹¹ Arguably, the emotional distress derived from pain can be difficult to distinguish from the pain itself.⁵⁷ However, measurement tools aiming to measure only pain intensity may be amiss in distinguishing key emotional factors of importance to the patients. Likewise, generic instruments measuring emotional burden may miss factors relevant to the population of patients with chronic pain. In practice, this could open the possibility that there may be unexplored factors of relevance to the population, or various subpopulations of patients with chronic pain, that clinicians should be aware of.

Conclusion

Compared to modern standards of psychometric development of PROMs via the COSMIN manual's criteria, all of the 10 identified PROMs measuring the affective dimension of pain are inadequate regarding content validity and psychometric properties.

Implications for Research

There is a need to develop new PROMs that can adequately measure the effect of psychological interventions

when treating chronic pain. Existing PROMs measuring the affective component of pain cannot be recommended for this use. Most importantly, besides high content validity and unidimensionality of each proposed scale, such a PROM will need to be assessed for DIF across the different classifications of chronic pain in the ICD-11.

Acknowledgments

The Lundbeck Foundation funded salary for AA.

Andrea O'Donnell conducted language revision.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Goldberg DS, McGee SJ. Pain as a global public health priority. *BMC Public Health*. 2011;11:770. doi:10.1186/1471-2458-11-770
2. Jackson TP, Stabile VS, McQueen KAK. The global burden of chronic pain. *ASA Newsl*. 2014;78(6):24–27.
3. Kronborg C, Handberg G, Axelsen F. Health care costs, work productivity and activity impairment in non-malignant chronic pain patients. *Eur J Health Econ*. 2009;10(1):5–13. doi:10.1007/s10198-008-0096-3
4. Andersen LN, Kohberg M, Juul-Kristensen B, Herborg LG, Sogaard K, Roessler KK. Psychosocial aspects of everyday life with chronic musculoskeletal pain: a systematic review. *Scand J Pain*. 2014;5(2):131–148. doi:10.1016/j.sjpain.2014.01.001
5. Garland EL. Treating chronic pain: the need for non-opioid options. *Expert Rev Clin Pharmacol*. 2014;7(5):545–550. doi:10.1586/17512433.2014.928587
6. Els C, Jackson TD, Kunyk D, et al. Adverse events associated with medium- and long-term use of opioids for chronic non-cancer pain: an overview of cochrane reviews. *Cochrane Database Syst Rev*. 2017;10(10):Cd012509.
7. Williams A, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. *Cochrane Database Syst Rev*. 2012;11. doi:10.1002/14651858.CD007407.pub3.
8. Scascighini L, Toma V, Dober-Spielmann S, Sprott H. Multidisciplinary treatment for chronic pain: a systematic review of interventions and outcomes. *Rheumatology*. 2008;47(5):670–678. doi:10.1093/rheumatology/ken021
9. Veehof MM, Trompetter HR, Bohlmeijer ET, Schreurs KM. Acceptance- and mindfulness-based interventions for the treatment of chronic pain: a meta-analytic review. *Cogn Behav Ther*. 2016;45(1):5–31. doi:10.1080/16506073.2015.1098724
10. Hughes LS, Clark J, Colclough JA, Dale E, McMillan D. Acceptance and commitment therapy (ACT) for chronic pain: a systematic review and meta-analysis. *Clin J Pain*. 2017;33(6):552–568. doi:10.1097/AJP.0000000000000425
11. Dworkin HR, Turk CD, Farrar TJ, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2005;113(12):9–19. doi:10.1016/j.pain.2004.09.012
12. Hilton L, Hempel S, Ewing BA, et al. Mindfulness meditation for chronic pain: systematic review and meta-analysis. *Ann Behav Med*. 2017;51(2):199–213. doi:10.1007/s12160-016-9844-2
13. Simons LE, Elman I, Borsook D. Psychological processing in chronic pain: a neural systems approach. *Neurosci Biobehav Rev*. 2014;39:61–78.

14. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychol Bull.* 2007;133(4):581–624. doi:10.1037/0033-2909.133.4.581
15. Price DD, Harkins SW. The affective-motivational dimension of pain A two-stage model. *APS J.* 1992;1(4):229–239. doi:10.1016/1058-9139(92)90054-G
16. Melzack R. The McGill pain questionnaire: major properties and scoring methods. *PAIN.* 1975;1(3):277–299. doi:10.1016/0304-3959(75)90044-5
17. Prinsen CAC, Mookkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018;27(5):1147–1157. doi:10.1007/s11136-018-1798-3
18. Ford L. Garbage-in, garbage-out: item generation as a threat to construct validity. 2019.
19. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4:1. doi:10.1186/2046-4053-4-1
20. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res.* 2018;27(5):1159–1170. doi:10.1007/s11136-018-1829-0
21. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* 2009;18(8):1115–1123. doi:10.1007/s11136-009-9528-5
22. Thomas RJ, McEwen J, Asbury AJ. The Glasgow pain questionnaire: a new generic measure of pain; development and testing. *Int J Epidemiol.* 1996;25:1060–1067. doi:10.1093/ije/25.5.1060
23. Gentile DA, Woodhouse J, Lynch P, Maier J, McJunkin T. Reliability and validity of the global pain scale with chronic pain sufferers. *Pain Physician.* 2011;14:61–70. doi:10.36076/ppj.2011/14/61
24. Mehack R, Torgerson WS. On the language of pain. *Anesthesiology.* 1971;34(1):50–59. doi:10.1097/0000542-197101000-00017
25. Clark WC, Kuhl JP, Keohan ML, Knotkova H, Winer RT, Griswold GA. Factor analysis validates the cluster structure of the dendrogram underlying the Multidimensional Affect and Pain Survey (MAPS) and challenges the a priori classification of the descriptors in the McGill Pain Questionnaire (MPQ). *Pain.* 2003;106(3):357–363. doi:10.1016/j.pain.2003.08.005
26. Melzack R, Wall PD. Pain mechanisms: a new theory. *Science.* 1965;150(3699):971–979. doi:10.1126/science.150.3699.971
27. Leavitt F, Garron DC, Whisler WW, Sheinkop MB. Affective and sensory dimensions of back pain. *Pain.* 1978;4(3):273–281. doi:10.1016/0304-3959(77)90139-7
28. Brennan AF, Barrett CL, Garretson HD. The utility of McGill pain questionnaire subscales for discriminating psychological disorder in chronic pain patients. *Psychol Health.* 1987;1(3):257–272. doi:10.1080/08870448708400329
29. Lowe NK, Walker SN, MacCallum RC. Confirming the theoretical structure of the McGill pain questionnaire in acute clinical pain. *Pain.* 1991;46(1):53–60. doi:10.1016/0304-3959(91)90033-T
30. Holroyd KA, Holm JE, Keefe FJ, et al. A multi-center evaluation of the McGill pain questionnaire: results from more than 1700 chronic pain patients. *Pain.* 1992;48(3):301–311. doi:10.1016/0304-3959(92)90077-O
31. Rucker KS, Metzler HM, Kregel J. Standardization of chronic pain assessment: a multiperspective approach. *Clin J Pain.* 1996;12:94–110. doi:10.1097/00002508-199606000-00004
32. Kerns RD, Turk DC, Rudy TE. The west haven-yale multidimensional pain inventory (WHYMPI). *Pain.* 1985;23(4):345–356. doi:10.1016/0304-3959(85)90004-1
33. Davidson MA, Tripp DA, Fabrigar LR, Davidson PR. Chronic pain assessment: a seven-factor model. *Pain Res Manag.* 2008;13:299–308. doi:10.1155/2008/976341
34. Hopwood CJ, Creech SK, Clark TS, Meagher MW, Morey LC. The convergence and predictive validity of the multidimensional pain inventory and the personality assessment inventory among individuals with chronic pain. *Rehabil Psychol.* 2007;52:443–450. doi:10.1037/0090-5550.52.4.443
35. McKillop JM, Nielson WR. Improving the usefulness of the multidimensional pain inventory. *Pain Res Manag.* 2011;16:239–244. doi:10.1155/2011/873424
36. Mikail SF, DuBreuil SC, D'Eon JL. A comparative analysis of measures used in the assessment of chronic pain patients. *Psychol Assess.* 1993;5(1):117–120. doi:10.1037/1040-3590.5.1.117
37. De Gagné TA, Mikail SF, D'Eon JL. Confirmatory factor analysis of a 4-factor model of chronic pain evaluation. *Pain.* 1995;60(2):195–202. doi:10.1016/0304-3959(94)00114-T
38. Deisinger JA, Cassisi JE, Lofland KR, Cole P, Bruehl S. An examination of the psychometric structure of the multidimensional pain inventory. *J Clin Psychol.* 2001;57(6):765–783. doi:10.1002/jclp.1048
39. Riley JL, Zawacki TM, Robinson ME, Geisser ME. Empirical test of the factor structure of the west haven-yale multidimensional pain inventory. *Clin J Pain.* 1999;15(1):24–30. doi:10.1097/00002508-199903000-00005
40. Wittink H, Turk DC, Carr DB, Sukiennik A, Rogers W. Comparison of the redundancy, reliability, and responsiveness to change among SF-36, oswestry disability index, and multidimensional pain inventory. *Clin J Pain.* 2004;20:133–142. doi:10.1097/00002508-200405000-00002
41. DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care.* 2007;45(5,Suppl1):S12–S21. doi:10.1097/01.mlr.0000254567.79743.e2
42. Revicki DA, Cook KF, Amtmann D, Harnam N, Chen W-H, Keefe FJ. Exploratory and confirmatory factor analysis of the PROMIS pain quality item bank. *Qual Life Res.* 2014;23:245–255. doi:10.1007/s11136-013-0467-9
43. Karoly P, Harris P. Assessing the affective component of chronic pain: development of the pain discomfort scale. *J Psychosom Res.* 1991;35:149–154. doi:10.1016/0022-3999(91)90069-Z
44. Ruehlman LS, Karoly P, Newton C, Aiken LS. The development and preliminary validation of a brief measure of chronic pain impact for use in the general population. *Pain.* 2005;113(1):82–90. doi:10.1016/j.pain.2004.09.037
45. Karoly P, Ruehlman LS, Aiken LS, Todd M, Newton C. Evaluating chronic pain impact among patients in primary care: further validation of a brief assessment instrument. *Pain Med.* 2006;7(4):289–298. doi:10.1111/j.1526-4637.2006.00182.x
46. Dargie E, Holden RR, Pukall CF. The Vulvar Pain Assessment Questionnaire inventory. *Pain.* 2016;157:2672–2686. doi:10.1097/j.pain.0000000000000682
47. Dargie E, Holden RR, Pukall CF. The vulvar pain assessment questionnaire: factor structure, preliminary norms, internal consistency, and test-retest reliability. *J Sex Med.* 2017;14(12):1585–1596. doi:10.1016/j.jsxm.2017.10.072
48. Prinsen CAC, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a “core outcome set” – a practical guideline. *Trials.* 2016;17(1):449. doi:10.1186/s13063-016-1555-2
49. Brodersen J, Meads D, Kreiner S, Thorsen H, Doward L, McKenna S. Methodological aspects of differential item functioning in the Rasch model. *J Med Econ.* 2007;10(3):309–324. doi:10.3111/13696990701557048
50. Salaffi F, Sarzi-Puttini P, Ciapetti A, Atzeni F. Clinimetric evaluations of patients with chronic widespread pain. *Best Pract Res Clin Rheumatol.* 2011;25:249–270. doi:10.1016/j.berh.2011.01.004
51. Deshaies K, Akhtar-Danesh N, Kaasalainen S. An evaluation of chronic pain questionnaires in the adult population. *J Nurs Meas.* 2015;23:22–39. doi:10.1891/1061-3749.23.1.22

52. Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. *J Clin Epidemiol.* 2018;95:73–93. doi:10.1016/j.jclinepi.2017.11.005
53. Treede RD, Rief BW, Barke BA, et al. A classification of chronic pain for ICD-11. *PAIN.* 2015;156(6):1003–1007. doi:10.1097/j.pain.000000000000160
54. Rasch G. An informal report on a theory of objectivity in comparisons. 1966.
55. Andersen E. Sufficient statistics and latent trait models. *Psychometrika.* 1977;42(1):69–81. doi:10.1007/BF02293746
56. Scott N, Fayers P, Aaronson N, et al. The practical impact of differential item functioning analyses in a health-related quality of life instrument. *Qual Life Res.* 2009;18(8):1125–1130. doi:10.1007/s11136-009-9521-z
57. Leder D. The experiential paradoxes of pain. *J Med Philos.* 2016;41(5):444–460. doi:10.1093/jmp/jhw020

Patient Related Outcome Measures

Dovepress

Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups.

The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>