

Review Article

Measurement properties of patient-reported outcome measures used in rehabilitation of adults with chronic musculoskeletal pain: A mapping review

A.J.A. Köke^{a,b,c,d,*}, C.H.G. Bastiaenen^e, J. Kleijnen^f, I. Telgenkamp^a, R.J.E.M. Smeets^{a,d,g} and L.W.M.E. Beckers^a

^aDepartment of Rehabilitation Medicine, Care and Public Health Research Institute (CAPHRI), Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

^bAdelante Centre of Expertise in Rehabilitation and Audiology, Hoensbroek, The Netherlands

^cFaculty Health and Technology, Zuyd University for Applied Sciences, Heerlen, The Netherlands

^dPain in Motion International Research Group (PiM), Brussels, Belgium

^eDepartment of Epidemiology, Care and Public Health Research Institute (CAPHRI), Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

^fKleijnen Systematic Reviews Ltd, York, UK

^gCentre for Integral Rehabilitation (CIR), Eindhoven, The Netherlands

Received 29 April 2022

Accepted 1 November 2022

Abstract.

BACKGROUND: Choosing measurement tools for diagnostic, prognostic, or evaluative purposes in a chronic musculoskeletal pain (CMP) population is challenging for rehabilitation practice. Implementation of measurement tools for clinical practice is impaired by gaps in knowledge about measurement properties.

OBJECTIVE: Identifying evidence about the measurement properties of tools frequently used in Dutch pain rehabilitation practice.

METHODS: A mapping review was conducted of eligible studies that investigated *reliability*, *validity*, or *responsiveness*, and *interpretability*, as defined by the COSMIN taxonomy, of original versions or Dutch translations of predefined Patient-Reported Outcome Measures (PROMs) in a CMP population. MEDLINE, PsycINFO, EMBASE, and CINAHL were searched in March 2021. Results were visually mapped.

RESULTS: Thirty-five studies were included. The results show many knowledge gaps in both original and translated versions. In general, aspects of validity were most frequently reported. The Pain Disability Index, Pain Catastrophizing Scale, and the 12-Item Short Form Health Survey were the most studied measurement tools. No results were found for the Checklist Individual Strength, Illness Perception Questionnaire, and Utrecht Coping List.

CONCLUSION: Little evidence of the measurement properties of PROMs used in rehabilitation of patients with CMP in the Netherlands was found. PROMs need to be used and interpreted with caution in daily practice.

Keywords: Validity, reliability, responsiveness, measurement tool, translation

*Corresponding author: A.J.A. Köke, Department of Rehabilitation Medicine, Care and Public Health Research Institute (CAPHRI), Faculty of Health Medicine and Life Sciences, Maastricht University,

P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail: albere.koke@maastrichtuniversity.nl.

1. Introduction

Rehabilitation in the field of chronic musculoskeletal pain (CMP) is based theoretically on the biopsychosocial model [1]. The main aim is to enable the patient to deal better with pain and pain-related disabilities in order to improve daily functioning and participation. The selection of specific treatment modules is preceded by the assessment of relevant factors that maintain chronic pain and associated disabilities. Subsequently, when conducting a rehabilitation programme, proper measurement of factors maintaining pain and disability considered relevant, as well as of outcomes, is extremely important [1]. The purposes of such measurements can be diagnostic, including setting treatment goals (focusing on problems that need to be targeted to achieve the patient's treatment goals), prognostic (predicting outcomes), or evaluative (evaluation of treatment goals) [2]. The measurement tools used could be a part of history taking, physical examination, neurophysiological testing, or imaging, and could include biomarker measurements, performance tests, or a wide array of patient-reported outcome measures (PROMs). PROMs are identified through self-completed questionnaires, with the goal of enabling patients to rate their own perceived physical, psychological and/or social functioning, participation, and/or quality of life [1,3]. The use of PROMs, as part of the overall assessment, continues to expand beyond clinical research, in recognition of its potential for daily care by supporting patient-centred approaches and contributing to shared decision-making.

Choosing from the large variety of available PROMs, identifying the most appropriate measure(s) for predefined diagnostic, prognostic or, most relevant, evaluative purposes within the population being treated, is a real challenge for clinical practice [4]. In the field of pain rehabilitation, several attempts and initiatives have been described for developing a minimum set of tools. This set should be used in clinical practice to support informed clinical decision-making, for prognosis, to monitor treatment progress, to evaluate outcomes, and also to facilitate the comparison and pooling of data for scientific purposes. The Dutch Dataset Pain Rehabilitation (DDPR), in addition to others like the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) and Outcome Measures in Rheumatoid Arthritis Clinical Trial (OMERACT), is an example of such a set and has been chosen as the core outcome set of PROMs for clinical practices providing interdisciplinary multimodal pain programmes in the

Netherlands [1]. The set has been implemented in part of Dutch chronic pain rehabilitation facilities but there are, however, important drawbacks to this implementation, such as the lack of a systematic overview of the research literature on the measurement properties of the tools. This also applies to PROMs that health care professionals use in addition to the DDPR. An important first step to aid further research and eventually for adopting tools in daily practice is an overview of which measurement properties have already been studied, and which have not.

The Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) advocates a scientifically sound taxonomy, which groups measurement properties for health measurement tools, including PROMs, into the domains of *reliability*, *validity*, and *responsiveness*. The COSMIN taxonomy is a recommended resource in the identification of knowledge gaps in clinimetric evaluations of PROMs [5]. An additional characteristic is the *interpretability* of the outcome scores themselves. Within the taxonomy of the COSMIN, interpretability is considered the assignment of clinical or commonly understood connotations to a measure's quantitative collected outcome scores or change in scores [2]. As such, interpretability cannot be considered a measurement property like reliability, validity, and responsiveness because it does not refer to the clinimetric quality of a measure itself. However, interpretability stands for the translation from quantitative scores into a clinically meaningful message to the patient and caregiver, which clearly demonstrates its great importance for clinical practice [2]. Unfortunately, interpretability often receives little attention in science or clinical practice and, therefore, may be an insufficiently studied aspect of the measurement tools, along with measurement properties [2].

As indicated above, a mapping review is a significant first step in increasing the evidence about and application of measurement tools. The goal of a mapping review is to identify and categorize existing literature to define research priorities and make informed decisions for systematic reviews as well as primary studies. The research field is described through a systematic search in a broad field and presentation of the results in a user-friendly format, often a visual figure or graph [6–8]. The scope for this mapping review is defined as follows:

1. To systematically map out existing literature about the measurement property domains *reliability*, *validity*, and *responsiveness* (as defined in the COSMIN taxonomy) for the tools proposed by the DDPR, expanded with additional frequently used tools, within pain rehabilitation clinical practice.

2. To systematically map out existing literature about the *interpretability* (as defined in the COSMIN taxonomy) of these tools.

2. Materials and methods

A mapping review was conducted to provide a broad overview of existing studies on measurement properties for commonly used PROMs in clinical pain rehabilitation programmes in the Netherlands.

2.1. Eligibility criteria

To be included, full-text articles and conference abstracts had to present original data. Any type of study that used PROMs only as outcome measurements was excluded, as these provide only indirect evidence about measurement properties of PROMs. No restriction was imposed as to publication date. The languages of the records were restricted to English, German, and Dutch. Inclusion criteria for PROMs, measurement properties, and the chronic pain population were specified as follows.

2.1.1. PROMs

Eligible studies had to present data on measurement properties of at least one of the predefined PROMs commonly used in pain rehabilitation centres in the Netherlands: the Hospital Anxiety and Depression Scale (HADS), Pain Catastrophizing Scale (PCS), Pain Disability Index (PDI), and Patient Specific Complaint (PSC), all part of the DDP, and the Checklist Individual Strength (CIS), Psychological Inflexibility in Pain Scale (PIPS), Illness Perception Questionnaire (IPQ), Pain Self-Efficacy Questionnaire (PSEQ), 12-Item Short Form Health Survey (SF-12), Symptom Checklist-90 (SCL-90), and Utrecht Coping List (UCL). The latter six PROMs are highly relevant for clinical practice in addition to the DDP and were selected based on the clinical expertise of the research team. Derivatives of PROMs were included as well. Studies were included if the measurement properties of the original PROM were reported, as were studies in which the measurement properties of the Dutch version of an original PROM were reported.

An overview of the list of PROMs, their intended constructs, and their purpose is presented in Table 1 [9].

2.1.2. Measurement properties

Eligible studies had to present data on at least one

measurement property in the domains reliability, validity, responsiveness, or on interpretability [5,9,10]. Definitions by the COSMIN of the domains and the measurement properties thereof were used to assess eligibility [5].

The domain *reliability* comprises reliability, measurement error, including test-retest and absolute agreement, and internal consistency;

The domain *validity* includes content validity (i.e., face validity), criterion validity (i.e., concurrent and predictive validity), and construct validity (i.e., structural validity, cross-cultural validity, and hypotheses testing);

Responsiveness contains measurement properties that refer to the ability of an outcome measure to detect whether change over time in the construct being measured has indeed occurred, or not. Responsiveness is also considered longitudinal validity, but separately described in the COSMIN taxonomy for reasons of the timing of the measurement (validity is only cross-sectional while responsiveness makes use of two measurements over time);

Interpretability represents the degree to which one can assign qualitative meaning – that is, clinically or commonly understood connotations – to a PROM's quantitative scores or change in scores. Several aspects can be used to provide additional quantitative information about interpretability: distribution scores, floor and ceiling effects, and minimally important change values.

2.1.3. Population

Studies were only included if they reported data on measurement properties of listed PROMs in a population with chronic primary pain (e.g., chronic primary musculoskeletal pain and chronic widespread pain (e.g., fibromyalgia)), chronic secondary pain (e.g., chronic post-surgical or post-traumatic pain (e.g., chronic pain whiplash injury)), chronic secondary musculoskeletal pain (e.g., osteoarthritis), or chronic neuropathic pain, according to the International Classification of Diseases ICD-11 classification of chronic pain [11]. Studies with a population of patients with chronic pain aged 16 years or older were included, since, as of 2018, the age of 16 is considered legal for independent medical-related decision-making in the Netherlands. Studies including mixed populations, other than chronic pain, were only included if the results were reported separately for the eligible population.

2.2. Data sources and searches

Multiple electronic data sources were searched in

Table 1
Construct and purpose (i.e., discriminative (D), evaluative (E), and predictive (P)) depicted for the original developed PROM version and the translated Dutch version

PROM	Language	First author	Year	Construct	Purpose PROM		
					D	E	P
HADS	English [†]	Zigmond [12]	1983	"To detect anxiety and depression among patients in medical settings."	x	x	
	Dutch	Spinhoven [13]	1997	"To provide a screening measure for the presence of anxiety and depression specifically and not for a global psychiatric disorder in general."	x		
PCS	English [†]	Sullivan [14]	1995	"To reflect on past painful experiences and to indicate the degree to which they experienced each of 13 thoughts or feelings when experiencing pain."	x	x	x
PDI	Dutch	Crombez [15]	1996	"To measure the extent of catastrophizing within pain."	x	x	
	English [†]	Pollard [16]	1984	"To measure the extent to which chronic pain interferes with a person's disability to engage in various life activities."	x	x	
PSC	Dutch	Soer [17]	2013	"To measure and evaluate disability associated with pain."	x	x	
	Dutch [†]	Beurskens [18]	1999	"To assess functional status in patients with low back pain."			x
CIS	Dutch [†]	Vercoulen [19]	1994	"To measure several aspects of fatigue."	x	x	x
PIPS	Swedish [†]	Wicksell [20]	2008	"To assess relevant aspects of psychological in/flexibility, such as avoidance and cognitive fusion."	x	x	x
	Dutch	Trompeter [21]	2014	"To measure psychological inflexibility."	x	x	
IPQ	English [†]	Weinman [22]	1996	"To assess cognitive representations of illness."	x	x	x
	Dutch	de Raaij [23]	2012	"To assess illness perceptions."	x	x	
PSEQ	English [†]	Nicholas [24]	2007	"To take pain into account when rating their self-efficacy beliefs."	x	x	x
	Dutch	van der Maas [25]	2012	"To assess participants' confidence in their ability to perform activities of daily living despite the pain."	x	x	x
SF-12	English [†]	Ware [26]	1996	"To measure health status."	x	x	
SF-12	Dutch	Mols [27]	2009	"To compare health status between groups of patients, to identify predictors of health status."	x	x	
SCL-90	English [†]	Derogatis [28]	1973	"To rate multidimensional symptom distress."	x	x	
SCL-90	Dutch	Arrindel [29]	2003	"To measure psychological distress." ^{**}	x	x	
UCL	Dutch [†]	Schreurs [30]	1984	"To measure coping strategies in problematic or adjustment demanding situations."	x	x	x

[†] Language of original developed PROM. ^{**} Interpretation of the reviewers based on ambiguous information. Abbreviations: CIS: Checklist Individual Strength; HADS: Hospital Anxiety and Depression Scale; IPQ: Illness Perception Questionnaire; PCS: Pain Catastrophizing Scale; PDI: Pain Disability Index; PIPS: Psychological Inflexibility in Pain Scale; PSEQ: Pain Self-Efficacy Questionnaire; PSFS: Patient-Specific Functional Scale; SCL-90: Symptom Checklist-90, SF-12: 12-Item Short Form Health Survey; UCL: Utrecht Coping List.

March 2021: MEDLINE, including in-process and other non-indexed citations, Epub ahead of print, and daily update (Ovid), PsycINFO (Ovid), EMBASE (Ovid), and CINAHL (EBSCOhost). Specific search queries encompassed both controlled vocabulary (e.g., MESH terms) and text words in titles and abstracts. The search strategies, tailored for each electronic data source (*Supplement 1*), were developed by author IT and peer-reviewed by JK. An information specialist of Kleijnen Systematic Reviews Ltd conducted the searches in all databases. Potentially relevant records were collected and duplicates were removed, both using End-Note X9 and manually, using the web-based system Rayyan (based on publication year, journal, volume, and issue) [31]. Included records, excluded literature reviews, and reviewers' personal libraries were screened for relevant records for inclusion as well.

2.3. Study selection

During the first phase, two initial reviewers, IT and either AK, RS, or LB, independently screened the search results for eligible studies on title and abstract, using Rayyan. During the second phase, the same initial reviewers screened the selected records in duplicate based on full-text articles according to the predefined eligibility criteria. Disagreements regarding the inclusion of specific records during both phases were discussed by the two initial reviewers and, if necessary, mediated by a third reviewer.

2.4. Data charting

Specifics on included studies were charted: author(s), publication year, PROM of interest, language of the investigated PROM, setting in which the study was conducted, country, the population of interest, and the study sample's age, gender, and pain duration. In addition, an overview of studies for each measurement domain and its specific properties was presented in a table and additionally visually mapped. Data extraction for measurement properties was performed as described in the original studies, according to COSMIN taxonomy [5], by one author (IT). In the case of ambiguous reporting, allocation of measurement properties was performed by author IT. A second author (LB) randomly verified 20% of the data extraction. As mapping reviews focus on the quantity and key characteristics of literature, quality assessment of included studies and/or methodologies is not indicated and thus no part of this study.

Table 2
Overview included studies per PROM

	Original language	(E/S/D)	Dutch translation
HADS	2	E	1
PCS	4 [§]	E	3 [§]
PDI	10*	E	4*
PSC	2	D	
CIS	0	D	
PIPS	2	S	1
IPQ	0	E	0
PSEQ	5	E	1
SF-12	3	E	0
SCL-90	4	E	0
UCL	0	D	

Abbreviations: CIS: Checklist Individual Strength, HADS: Hospital Anxiety and Depression Scale; IPQ: Illness Perception Questionnaire; PCS: Pain Catastrophizing Scale; PDI: Pain Disability Index; PIPS: Psychological Inflexibility in Pain Scale; PSEQ: Pain Self-Efficacy Questionnaire; SCL-90: Symptom Checklist-90; SF-12: 12-Item Short Form Health Survey; UCL: Utrecht Coping List. E = English, S = Swedish, D = Dutch. [§]Including Delphi-study. *Including study described results of Dutch and English version together.

3. Results

A total of 24,476 records were identified, of which 24,402 were obtained through database searching and 74 through other sources (reference tracking and reviewers' personal libraries). After deduplication, 13,957 records were screened based on title and abstract (Fig. 1). Consequently, 106 full-text records were assessed for eligibility, of which 40 were included in the quantitative synthesis. Thirty-six full-text publications and four (conference) abstracts were distinguished. Finally, 35 unique studies were identified for inclusion.

The study populations of the majority of studies consisted of groups of patients with different diagnoses: chronic primary musculoskeletal pain (e.g., low back, neck, shoulder, and knee pain), chronic widespread pain (e.g., fibromyalgia), chronic post-traumatic pain (e.g., whiplash injury), and/or chronic secondary musculoskeletal pain (e.g., arthritis). In addition, studies were conducted in different settings (hospitals, rehabilitation centres, pain centres, primary care, or national surveys), or included patients from several settings. The mean reported duration of pain varied from 3 to 162 months. The average age of all populations was 46.5 years (SD 6.4, range 32.1–68.6).

Results for measurement properties were found for the HADS, PCS, PDI, PSC, PIPS, PSEQ (including a short form PSEQ-2 item version), SF-12 (including a derivative SF6D12), and SCL-90 (including SCL-90-R). No studies, in chronic pain populations, were found for the CIS, IPQ, and UCL.

Thirty-one studies included PROMs in the language of original development (English ($n = 27$), Dutch ($n =$

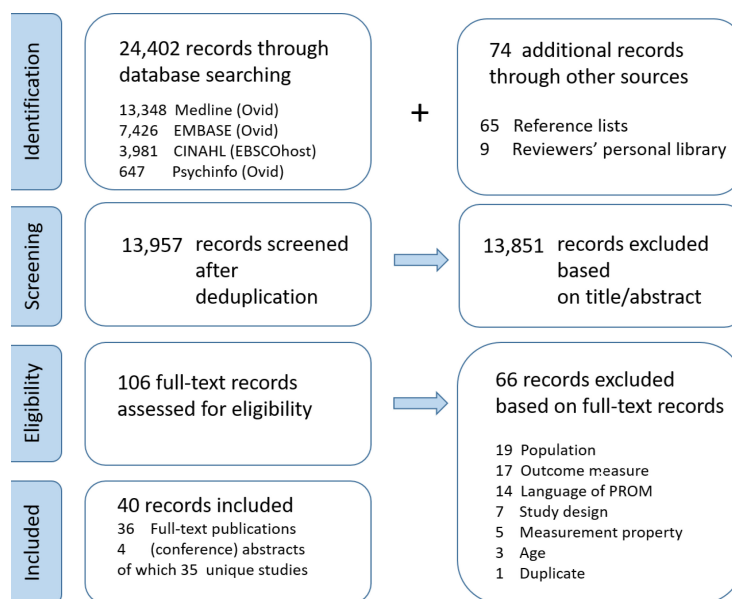


Fig. 1. The flowchart representing the identification, screening, and eligibility process.

2), Swedish ($n = 2$). In addition, one study evaluated a PROM (PDI) in both the original (English) and translated versions (Dutch), and eight studies investigated PROMs translated into Dutch. One included article reported the results of a Delphi study for consensus on the most appropriate PROM to assess psychosocial risk factors (PSEQ and PCS) in patients with chronic pain. These results were regarded as an outcome for the domain of validity (content validity). Four articles included results for two PROMs: HADS/PDI, PCS/PSEQ, SF-12/PDI, and SF-12/SCL-90.

The numbers of included studies for each PROM in the original version and translated version (Dutch) are presented in Table 2. The study characteristics of the included records for each PROM are presented in Supplement 2.

3.1. Overview of measurement properties for original PROMs

The data, as mapped in Fig. 2 and presented in Table 3, show several knowledge gaps for the original PROMs in the population of patients with CMP. Out of all the original language PROMs included, the PCS had the most measurement properties studied. Studies have been carried out for all COSMIN domains and every measurement property included in these, apart from face validity. Likewise, most measurement properties of the original PDI were studied, again covering all COSMIN domains, apart from reliability and mea-

surement error in the domain reliability and face validity in the domain validity. Despite the fact that studies reported results within all domains of the SF-12, results for reliability, measurement error, content and criterion validity, as well as structural validity, were lacking. For the PSEQ, no studies at all reported on interpretability, while, in the domains reliability and validity, there were several knowledge gaps. No studies reported on responsiveness and interpretability for the HADS and PIPS and there were gaps for both in the domains reliability and validity. For the SCL-90, only results for the domain validity were found. For the original Dutch PSC, only responsiveness was studied. Furthermore, no studies at all were found for the CIS, IPQ, and UCL.

3.2. Overview of measurement properties of translated PROMs

Of the PROMs translated into Dutch, cross-cultural validity was only reported on for the HADS and PDI. Most measurement properties were studied for the Dutch PDI, with the exception of face, criterion and structural validities. For the PCS, no studies were found for the domain reliability, and in the domain of validity, only structural validity was examined. In the domain reliability, internal consistency was tested for HADS, PIPS, and PSEQ, with reliability as well for the PSEQ. In the domain validity, structural validity was reported for all three and hypotheses testing for the PSEQ and PIPS. Only interpretability was studied for HADS, but

Table 3
Investigated measurement properties presented in included studies for each PROM

PROM	Author	Year	Reliability	Internal consistency	Measurement error	Content validity	Criterion validity	Structural validity	Hypotheses testing	Cross cultural validity	Responsiveness	Interpretability	Other
Dutch language PROMs	Giusti [32]	2020	✓	✓		✓/CFA		✓/CFA	✓	✓	✓/RV	✓/EFA	
	van Damme [33]	2000						✓/CFA			✓		
	van Damme [34]	2002						✓/CFA			✓		
	Pulles [35]	2020									✓		
	Soer [36,37]	2011, 2012									✓		
	Soer [17]	2013	✓	✓	✓				✓	✓			
	Soer [38]	2015											
	Beurskens [39]	1996									✓		
	Beurskens [18]	1999									✓		✓/FE
	Trompetter [21]	2014		✓				✓/CFA	✓				✓/EFA*
	van der Maas [25]	2012	✓	✓				✓/CFA	✓				✓/EFA*
	Pallant [40]	2005		✓				✓/CFA	✓				✓/EFA
	Rusu [41]	2016		✓			✓/CV	✓/CFA	✓				
	Osman [42]	2000		✓	✓		✓/CV	✓/CFA	✓				
	George [43]	2010		✓			✓/CV*		✓*				
	Prime [44]	2012											
	Sleijser [45]	2019			✓			✓* CFA	✓				
	Pollard [16]	1984							✓				
	Tait [46]	1987		✓					✓				
	Jerome [47]	1991							✓				
	Millard [48]	1991							✓				
	Strong [49]	1994		✓					✓				
	Crighton [50]	2014		✓				✓/CV	✓				
	Soer [38]	2015						✓/CV	✓		✓/RV		
	Morris [51]	2015											
	Rusu [41]	2016						✓/CV	✓				
	McKillop [52]	2018							✓				
	Wicksell [20]	2008		✓				✓/CV	✓				✓/EFA*
	Wicksell [53]	2010		✓				✓/CV	✓				✓/EFA*
	Nicholas [24]	2007		✓				✓/CFA	✓				
	Maughan [54]	2010	✓	✓					✓				
	Nicholas** [55]	2015	✓	✓				✓/CFA	✓	✓			✓/EFA
	Costa [56]	2017	✓	✓									
	Sleijser [45]	2019		✓		✓			✓				
	Luo** [57]	2012		✓					✓		✓*		
	Morris [51]	2015		✓					✓				
	Tawiah [58]	2018		✓					✓				
	Tawiah [59]	2019		✓					✓				
	Kroenke** [60]	2019		✓					✓				
	Kimney [61]	1991		✓					✓				
	Bernstein [62]	1994		✓				✓/CFA	✓				
	Peebles [63]	2001		✓					✓				
	Kroenke** [64]	2019		✓					✓				

* Interpretation of the reviewers based on ambiguous information. Abbreviations: HADS: Hospital Anxiety and Depression Scale; PCS: Pain Catastrophizing Scale; PDI: Pain Disability Index; PIPS: Psychological Inflexibility in Pain Scale; PSEQ: Pain Self-Efficacy Questionnaire; PSC: Patient Specific Complaints; SCL-90: Symptom Checklist-90; SF-12: 12-Item Short Form Health Survey; CV: Concurrent Validity; RV: reference values; FE: feasibility; CFA: Confirmatory Factor Analysis; EFA: Exploratory Factor Analysis. ** study of derivative or subscale of original PROM.

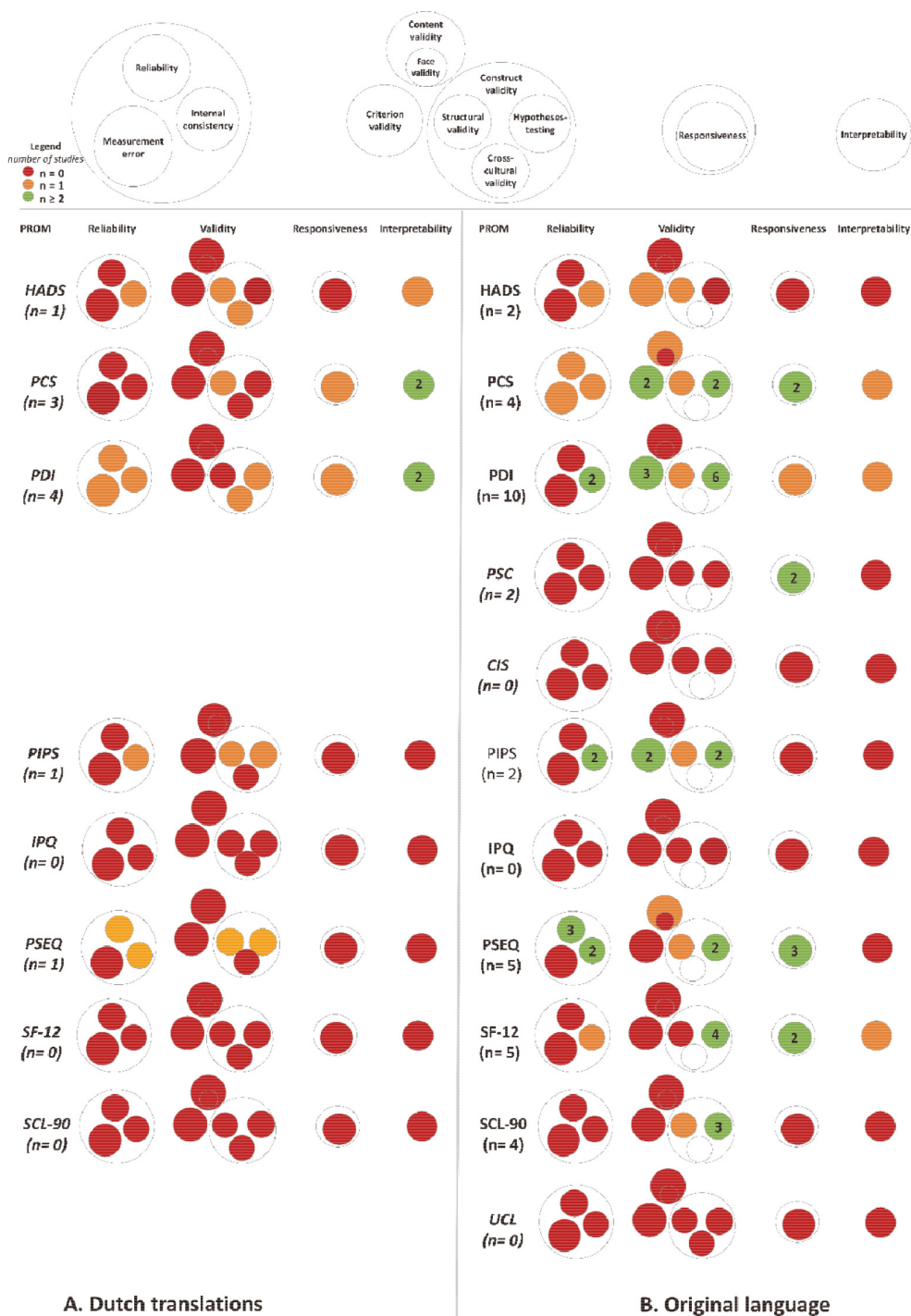


Fig. 2. Data on measurement properties mapped in ascending order of the number of studies presented for each PROM (PROMs in bold and italicized = Dutch, bold = English, normal font = Swedish). The shapes of the COSMIN domains are according to the taxonomy developed by COSMIN [5]. Red indicates no studies, orange one study, and green two or more studies, blank does not apply (cross-cultural validation). Abbreviations: HADS: Hospital Anxiety and Depression Scale; PCS: Pain Catastrophizing Scale; PDI: Pain Disability Index; PSC: Patient Specific Complaints; CIS: Checklist Individual Strength; PIPS: Psychological Inflexibility in Pain Scale; IPQ: Illness Perception Questionnaire; PSEQ: Pain Self-Efficacy Questionnaire; SF-12: 12-Item Short Form Health Survey; SCL-90: Symptom Checklist-90; UCL: Utrecht Coping List.

not responsiveness; for PIPS and PSEQ, responsiveness and interpretability were not studied. No studies on the Dutch SF-12 and SCL-90 were found in populations of patients with chronic pain.

4. Discussion

The objective of this study was to map out the literature available on the reliability, validity, responsiveness, and interpretability of PROMs commonly used in Dutch pain rehabilitation, including measurement tools proposed by the DDPR. The results show many knowledge gaps, although more studies were found for the PROMs in their original languages than for translated versions. Information has been published for all domains, but not all measurement properties, of both the original and Dutch version of the PDI and the original versions of the PCS and SF-12. For the original Dutch CIS, IPQ, and UCL, as well as for the Dutch translations of the SF-12 and SCL-90, no information was found on any measurement property. Overall, for the other tools, aspects of validity were most frequently reported, followed by reliability, responsiveness, and interpretability.

The results of this study complement a previous evidence review for a population with CMP evaluating measurement properties of PROMs for pain severity and pain-related functional impairment not included in the current review. For only three of the 13 multi-item tools (Oswestry Disability Index (ODI); Roland-Morris Disability Questionnaire (RMDQ); SF-36 Bodily Pain Scale (SF-36 BPS)), were data found for all COSMIN domains [65]. Interestingly, the PROMs investigated in the present study had already been developed in the '70s, '80s, or '90s, apart from the PSEQ (2007) and PIPS (2008). Since their development, PROMs have been widely used both clinically and in research, and are nowadays recommended as core outcome measures by the DDPR. The limited number of clinimetric studies performed in the CMP population in recent decades, despite the prominent role of these questionnaires, clearly points out a shortcoming in prior and current practices. At the same time, the paucity of research found by this review suggests that both researchers and funders have insufficiently acknowledged the importance of clinimetric research. An additional illustration of this point is the fact that a selective group of researchers is responsible for the majority of studies concerning specific PROMs, such as the PDI, which is unfavourable considering research biases.

The finding that responsiveness has received less attention than reliability and validity may be explained by the challenges in executing and interpreting research on this, including the necessity of multiple repeated measurements, as well as the fact that not all PROMs have an evaluative purpose. Interpretability, of high importance as it gives clinical meaning to (change in) scores, has been unjustifiably ignored. The importance of interpretation of scores was underlined during the implementation of the DDPR, as the lack of information on interpretability was a frequently mentioned reason for the PROMs not being feasible or relevant to use in practice [66]. It is also striking that only a single study reported on the content validity of a measure, while COSMIN considers it the most important measurement property. This is because the content of a measurement instrument should be relevant, comprehensive, and comprehensible with respect to the construct of interest and to the target population [5].

A major strength of this review is that it was performed completely within the theoretical framework of the COSMIN taxonomy, which is based on international consensus [5]. However, as the extraction of measurement properties was determined by the authors' reporting instead of reviewer assessment, the allocation may not entirely be in accordance with the taxonomy's definition. This limitation is a particular concern for PROMs developed before the publication of the taxonomy, and where the terminology used in a study is ambiguous. Content validity and pilot testing (not included in this review), for example, can involve overlapping aspects, and factor analysis can be part of both construct validity and field testing (not included) [2]. Similarly, potentially relevant studies may not have been identified due to discrepancies between the terms in our comprehensive search strategy and the terminology of measurement properties used by authors of such studies, resulting in selection bias. This may also explain the finding that nine studies from reviewers' personal libraries were not identified by our search strategy.

Another strength was the broad range of PROMs examined, covering many aspects of the biopsychosocial model. They were chosen pragmatically, emerging from the DDPR and the clinical experience of the research team. Nevertheless, a limitation of the selection's being made by a narrow group is that other relevant questionnaires and corresponding studies may have been missed. The Global Perceived Effect (GPE), Numerical Rating Scale (NRS), and Visual Analogue Scale (VAS), although part of the DDPR, were considered nonspecific response scales: they are used to measure a variety

of constructs and, correspondingly, adopt diverse question wordings, anchors at either end of the scale, time spans, and presentations. Given the limited comparability of the different variants, single-item GPE, NRS, and VAS scales were not included in this review. Last, the objective was restricted to the language of the original PROMs and their Dutch translations. Therefore, it is not possible to generalize our results to other languages and cultures, although similar trends, or even fewer studies, could be expected.

For all PROMs, more research is needed to fill the knowledge gaps about their measurement properties, with particular attention to content validity and interpretability. At the same time, for the Dutch versions of the PDI, PCS, and PSC as well as for the original language versions of the PDI, PCS, PSEQ, PIPS, SF-12, and SCL-90, sufficient data seem available to perform systematic reviews to describe their measurement properties, synthesize the data, and perform quality assessments of studies. Specific attention is required as to whether the measurement properties of investigated tools are in line with their intended purposes (diagnostic, predictive, evaluative). The focus of further work could be extended to translations other than into Dutch and to other PROMs. Several research teams performing (replication) studies would likely produce more valid and generalizable conclusions. While the COSMIN taxonomy and the other COSMIN tools are recommended for future studies, their applicability or otherwise to the specific field of rehabilitation medicine should be considered. For instance, the lack of confirmed clinimetrically sound comparator measurement instruments can make testing hypotheses imprecise. Moreover, adequate sample sizes, as indicated in the COSMIN study design checklist, can be challenging to accomplish within the setting of CMP, especially if one aims to examine measurement properties in subpopulations.

While awaiting the results of future clinimetric studies, researchers need to be critical when using PROMs and transparently acknowledge any limitations. Policy- and decision-makers, including health insurance companies, should not overestimate the impact of either individual or group-based PROM outcomes and be careful when drawing conclusions from them. Likewise, health care professionals in rehabilitation should be aware of the limited evidence of PROMs' clinimetric qualities. Given the heterogeneity of the CMP population and the complexity of rehabilitation interventions, it is in any case of the utmost importance to guide the rehabilitation trajectory, based on combinations of measurement instruments within the biopsychosocial model, together

with the clinical expertise of the interdisciplinary team and the patient's perspective. However, this necessity to integrate outcomes of multiple PROMs covering the total biopsychosocial spectrum when assessing CMP patients, can lead to extensive sets of questionnaires to be completed by the patient, not only in the diagnostic phase but also during and after treatment.

Given the associated burden of this for patients, item banks and computer adaptive testing (such as the Patient-Reported Outcomes Measurement Information System (PROMIS)) that have been validated could be considered as alternatives to traditional patient-reported outcome measurements. For the sake of comparability, a collective approach based on expert consensus between representatives of health care professionals, researchers, and patients is desirable, and an important focus for future work.

5. Conclusion

The studies included in this mapping review demonstrate a paucity of evidence for the reliability, validity, responsiveness, and interpretability of most PROMs frequently used in rehabilitation of patients with CMP in the Netherlands. The main implication is that these PROMs need to be used and interpreted with caution in daily practice.

Ethical approval

Not applicable.

Funding

This work was supported by the Centre for Integral Rehabilitation (CIR). The public partner, Maastricht University, is responsible for the study design, data collection and analysis, decision to publish, and preparation of the manuscript.

Informed consent

Not applicable.

Conflict of interest

None of the authors have any conflicts of interest to declare.

Acknowledgments

The authors thank Shelley de Kock for performing the literature searches and Les Hearn for proofreading the manuscript.

Author contributions

Study design and conceptualizations: S.R., B.L., Search strategy: K.J., I.T., B.L., Databases searching: I.T. B.L., data screening K.A., I.T., S.R., B.L., Data extraction: K.A., I.T., B.L., Data synthesis and interpretation: K.A., B.C., S.R., B.L., Manuscript drafting B.L., Writing review K.A. B.C., K.J., S.R., B.L., Editing review and visualization K.A., Supervision S.R., B.L. All authors read and approved the final version of the manuscript.

Supplementary data

The supplementary files are available to download from <http://dx.doi.org/10.3233/BMR-220133>.

References

- [1] Köke AJ, Smeets RJ, Schreurs KM, van Baalen B, de Haan P, Remerie SC, Schiphorst Preuper HR, Reneman MF. Dutch Dataset Pain Rehabilitation in daily practice: Content, patient characteristics and reference data. *European Journal of Pain*. 2017; 21(3): 434-444. doi: 10.1002/ejp.937.
- [2] de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*: Cambridge University Press. 2011.
- [3] Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, Cleeland C, Dionne R, Farrar JT, Galer BS, Hewitt DJ, Jadad AR, Katz NP, Kramer LD, Manning DC, McCormick CG, McDermott MP, McGrath P, Quessy S, Rappaport BA, Robinson JP, Royal MA, Simon L, Stauffer JW, Stein W, Tollett J, Witter J. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2003; 106(3): 337-345. doi: 10.1016/j.pain.2003.08.001.
- [4] Prinsen CAC, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, Williamson PR, Terwee CB. 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.
- [5] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*. 2010; 63(7): 737-745. doi: 10.1016/j.jclinepi.2010.02.006.
- [6] Hetrick SE, Parker AG, Callahan P, Purcell R. Evidence mapping: Illustrating an emerging methodology to improve evidence-based practice in youth mental health. *Journal of Evaluation in Clinical Practice*. 2010; 16(6): 1025-1030. doi: 10.1111/j.1365-2753.2008.01112.x.
- [7] Miake-Lye IM, Hempel S, Shanman R, Shekelle PG. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. *Systematic Reviews*. 2016; 5: 28. doi: 10.1186/s13643-016-0204-x.
- [8] Cooper ID. What is a "mapping study"? *Journal of the Medical Library Association*. 2016; 104(1): 76-78. doi: 10.3163/1536-5050.104.1.013.
- [9] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*. 2018; 27(5): 1147-1157. doi: 10.1007/s11136-018-1798-3.
- [10] de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine*. New York, NY: Cambridge University Press. 2011.
- [11] Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoit R, Cohen M, Evers S, Finnerup NB, First MB, Giamberardino MA, Kaasa S, Kosek E, Lavand'homme P, Nicholas M, Perrot S, Scholz J, Schug S, Smith BH, Svensson P, Vlaeyen JWS, Wang SJ. A classification of chronic pain for ICD-11. *Pain*. 2015; 156(6): 1003-1007. doi: 10.1097/j.pain.000000000000160.
- [12] Zigmund AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*. 1983; 67(6): 361-370. doi: 10.1111/j.1600-0447.1983.tb09716.x.
- [13] Spinoven P, Ormel J, Sloekers PP, Kempen GI, Speckens AE, Van Hemert AM. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychological Medicine*. 1997; 27(2): 363-370. doi: 10.1017/s0033291796004382.
- [14] Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. *Psychological Assessment*. 1995; 7(4): 524-532. doi: 10.1037/1040-3590.7.4.524.
- [15] Crombez G, Vlaeyen JWS. De Pain Catastrophizing Scale (PCS). Ungepubliceerde geautoriseerde Nederlandstalige bewerking (Unpublished authorized Dutch version). 1996.
- [16] Pollard CA. Preliminary validity study of the pain disability index. *Perceptual and Motor Skills*. 1984; 59(3): 974. doi: 10.2466/pms.1984.59.3.974.
- [17] Soer R, Köke AJ, Vroomen PC, Stegeman P, Smeets RJ, Coppes MH, Reneman MF. Extensive validation of the pain disability index in 3 groups of patients with musculoskeletal pain. *Spine*. 2013; 38(9): E562-568. doi: 10.1097/BRS.0b013e31828af21f.
- [18] Beurskens AJ, de Vet HC, Köke AJ, Lindeman E, van der Heijden GJ, Regtop W, Knipschild PG. A patient-specific approach for measuring functional status in low back pain. *Journal of Manipulative and Physiological Therapeutics*. 1999; 22(3): 144-148. doi: 10.1016/s0161-4754(99)70127-2.
- [19] Vercoulen JHMM, Swanink CMA, Fennis JFM, Galama JMD, van der Meer JWM, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. *Journal of Psychosomatic Research*. 1994; 38(5): 383-392. doi: 10.1016/0022-3999(94)90099-X.
- [20] Wicksell RK, Renöfält J, Olsson GL, Bond FW, Melin L. Avoidance and cognitive fusion – Central components in pain related disability? Development and preliminary validation of the Psychological Inflexibility in Pain Scale (PIPS). *European Journal of Pain*. 2008; 12(4): 491-500. doi: 10.1016/j.ejpain.2007.08.003.
- [21] Trompetter HR, Bohlmeijer ET, van Baalen B, Kleen M, Köke

- A, Reneman M, Schreurs KMG. The Psychological Inflexibility in Pain Scale (PIPS). *European Journal of Psychological Assessment*. 2014; 30(4): 289-295. doi: 10.1027/1015-5759/a000191.
- [22] Weinman J, Petrie KJ, Moss-morris R, Horne R. The illness perception questionnaire: A new method for assessing the cognitive representation of illness. *Psychology & Health*. 1996; 11(3): 431-445. doi: 10.1080/08870449608400270.
- [23] de Raaij EJ, Schröder C, Maissan FJ, Pool JJ, Wittink H. Cross-cultural adaptation and measurement properties of the Brief Illness Perception Questionnaire-Dutch Language Version. *Manual Therapy*. 2012; 17(4): 330-335. doi: 10.1016/j.math.2012.03.001.
- [24] Nicholas MK. The pain self-efficacy questionnaire: Taking pain into account. *European Journal of Pain*. 2007; 11(2): 153-163. doi: 10.1016/j.ejpain.2005.12.008.
- [25] van der Maas LCC, de Vet HCW, Köke A, Bosscher RJ, Peters ML. Psychometric properties of the pain self-efficacy questionnaire (PSEQ). *European Journal of Psychological Assessment*. 2012; 28(1): 68-75. doi: 10.1027/1015-5759/a000092.
- [26] Ware J, Jr, Kosinski M, Keller SD. A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*. 1996; 34(3): 220-233. doi: 10.1097/00005650-199603000-00003.
- [27] Mols F, Pelle AJ, Kupper N. Normative data of the SF-12 health survey with validation using postmyocardial infarction patients in the Dutch population. *Quality of Life Research*. 2009; 18(4): 403-414. doi: 10.1007/s11136-009-9455-5.
- [28] Derogatis LR, Lipman RS, Covi L. SCL-90: An outpatient psychiatric rating scale-preliminary report. *Psychopharmacology Bulletin*. 1973; 9(1): 13-28.
- [29] Arrindell WAEJHM. *Symptom checklist SCL-90: handleiding bij een multidimensionele psychopathologie-indicator*. Lisse; Amsterdam: Swets Test Publishers; Harcourt Test Publ. 2003.
- [30] Schreurs PG, Tellegen B, Van de Willige G. Gezondheid, stress en coping: De ontwikkeling van de Utrechtse Coping Lijst. *Gedrag: Tijdschrift Voor Psychologie*. 1984; 12: 101-117.
- [31] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan – a web and mobile app for systematic reviews. *Systematic Reviews*. 2016; 5(1): 210. doi: 10.1186/s13643-016-0384-4.
- [32] Giusti EM, Jonkman A, Manzoni GM, Castelnuovo G, Terwee CB, Roorda LD, Chiarotto A. Proposal for improvement of the hospital anxiety and depression scale for the assessment of emotional distress in patients with chronic musculoskeletal pain: A bifactor and item response theory analysis. *Journal of Pain*. 2020; 21(3-4): 375-389. doi: 10.1016/j.jpain.2019.08.003.
- [33] Van Damme S, Crombez G, Vlaeyen J, Goubet L, Van den Broeck A, Van Houdenhove B. De Pain Catastrophizing Scale: Psychometrische karakteristieken en normering. *Gedragstherapie*. 2000; 33(3): 209-220.
- [34] Van Damme S, Crombez G, Bijttebier P, Goubert L, Van Houdenhove B. A confirmatory factor analysis of the Pain Catastrophizing Scale: Invariant factor structure across clinical and non-clinical populations. *Pain*. 2002; 96(3): 319-324. doi: 10.1016/S0304-3959(01)00463-8.
- [35] Pulles A, Köke AJA, Strackke RP, Smeets R. The responsiveness and interpretability of psychosocial patient-reported outcome measures in chronic musculoskeletal pain rehabilitation. *European Journal of Pain*. 2020; 24(1): 134-144. doi: 10.1002/ejp.1470.
- [36] Soer R, Reneman MF, Vroomen PC, Stegeman P, Coppes MH. Responsiveness and minimal clinically important change of the pain disability index in patients with chronic back pain. *Spine*. 2012; 37(8): 711-715. doi: 10.1097/BRS.0b013e31822c8a7a.
- [37] Soer R, Reneman MF, Stegeman P, Vroomen PC, Coppes MH. T523 responsiveness and minimal clinically important change of the pain disability index in patients with chronic back pain. *European Journal of Pain Supplements*. 2011; 5(S1): 80-80. doi: 10.1016/s1754-3207(11)70271-1.
- [38] Soer R, Köke AJ, Speijer BL, Vroomen PC, Smeets RJ, Coppes MH, Reneman MF, Gross DP, Groningen Spine Study G. Reference values of the pain disability index in patients with painful musculoskeletal and spinal disorders: A cross-national study. *Spine*. 2015; 40(9): E545-551. doi: 10.1097/BRS.0000000000000827.
- [39] Beurskens AJ, de Vet HC, Köke AJ. Responsiveness of functional status in low back pain: A comparison of different instruments. *Pain*. 1996; 65(1): 71-76. doi: 10.1016/0304-3959(95)00149-2.
- [40] Pallant JF, Bailey CM. Assessment of the structure of the Hospital Anxiety and Depression Scale in musculoskeletal patients. *Health & Quality of Life Outcomes*. 2005; 3: 82. doi: 10.1186/1477-7525-3-82.
- [41] Rusu AC, Santos R, Pincus T. Pain-related distress and clinical depression in chronic pain: A comparison between two measures. *Scandinavian Journal of Pain*. 2016; 12: 62-67. doi: 10.1016/j.sjpain.2016.04.001.
- [42] Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. The Pain Catastrophizing Scale: Further psychometric evaluation with adult samples. *Journal of Behavioral Medicine*. 2000; 23(4): 351-365. doi: 10.1023/a:1005548801037.
- [43] George SZ, Valencia C, Beneciuk JM. A Psychometric investigation of fear-avoidance model measures in patients with chronic low back pain. *Journal of Orthopaedic and Sports Physical Therapy*. 2010; 40(4): 197-205. doi: 10.2519/jospt.2010.3298.
- [44] Prime H, Sullivan M, Thibault P. Functional distinctiveness of the subscales of the pain catastrophizing scale. *Pain Research and Management*. 2012; 17(3): 220-221.
- [45] Sleijser-Koehorst MLS, Bijker L, Cuijpers P, Scholten-Peeters GGM, Coppieters MW. Preferred self-administered questionnaires to assess fear of movement, coping, self-efficacy, and catastrophizing in patients with musculoskeletal pain – A modified Delphi study. *Pain*. 2019; 160(3): 600-606. doi: 10.1097/j.pain.0000000000001441.
- [46] Tait RC, Pollard CA, Margolis RB, Duckro PN, Krause SJ. The Pain Disability Index: Psychometric and validity data. *Archives of Physical Medicine and Rehabilitation*. 1987; 68(7): 438-441.
- [47] Jerome A, Gross RT. Pain disability index: Construct and discriminant validity. *Archives of Physical Medicine and Rehabilitation*. 1991; 72(11): 920-922. doi: 10.1016/0003-9993(91)90012-8.
- [48] Millard RW, Jones RH. Construct validity of practical questionnaires for assessing disability of low-back pain. *Spine*. 1991; 16(7): 835-838. doi: 10.1097/00007632-199107000-00026.
- [49] Strong J, Ashton R, Large RG. Function and the patient with chronic low back pain. *Clinical Journal of Pain*. 1994; 10(3): 191-196. doi: 10.1097/00002508-199409000-00004.
- [50] Crighton AH, Wygant DB, Applegate KC, Umlauf RL, Granacher RP. Can brief measures effectively screen for pain and somatic malingering? Examination of the modified somatic perception questionnaire and pain disability index. *Spine Journal*. 2014; 14(9): 2042-2050. doi: 10.1016/j.spinee.

- 2014.04.012.
- [51] Morris T, Hee SW, Stallard N, Underwood M, Patel S. Can we convert between outcome measures of disability for chronic low back pain? *Spine*. 2015; 40(10): 734-739. doi: 10.1097/BRS.0000000000000866.
- [52] McKillop AB, Carroll LJ, Dick BD, Battie MC. Measuring participation in patients with chronic back pain-the 5-Item Pain Disability Index. *Spine Journal*. 2018; 18(2): 307-313. doi: 10.1016/j.spinee.2017.07.172.
- [53] Wicksell RK, Lekander M, Sorjonen K, Olsson GL. The psychological inflexibility in pain scale (PIPS) – statistical properties and model fit of an instrument to assess change processes in pain related disability. *European Journal of Pain*. 2010; 14(7): 771 e771-714. doi: 10.1016/j.ejpain.2009.11.015.
- [54] Maughan EF, Lewis JS. Outcome measures in chronic low back pain. *European Spine Journal*. 2010; 19(9): 1484-1494. doi: 10.1007/s00586-010-1353-6.
- [55] Nicholas MK, McGuire BE, Asghari A. A 2-item short form of the Pain Self-efficacy Questionnaire: Development and psychometric evaluation of PSEQ-2. *Journal of Pain*. 2015; 16(2): 153-163. doi: 10.1016/j.jpain.2014.11.002.
- [56] Costa DSJ, Asghari A, Nicholas MK. Item response theory analysis of the Pain Self-Efficacy Questionnaire. *Scandinavian Journal of Pain*. 2017; 14: 113-117. doi: 10.1016/j.sjpain.2016.08.001.
- [57] Luo N, Wang P, Fu AZ, Johnson JA, Coons SJ. Preference-based SF-6D scores derived from the SF-36 and SF-12 have different discriminative power in a population health survey. *Medical Care*. 2012; 50(7): 627-632. doi: 10.1097/MLR.0b013e31824d7471.
- [58] Tawiah A, Al Sayah F, Ohinmaa A, Johnson JA. PRM173 – discriminative validity of the EQ-5D-5L and SF-12 in older adults with arthritis. *Value in Health*. 2018; 21(Suppl 3): S386. doi: 10.1016/j.jval.2018.09.2292.
- [59] Tawiah AK, Al Sayah F, Ohinmaa A, Johnson JA. Discriminative validity of the EQ-5D-5L and SF-12 in older adults with arthritis. *Health & Quality of Life Outcomes*. 2019; 17(1): 68. doi: 10.1186/s12955-019-1129-6.
- [60] Kroenke K, Baye F, Lourens SG. Comparative validity and responsiveness of PHQ-ADS and other composite anxiety-depression measures. *Journal of Affective Disorders*. 2019; 246: 437-443. doi: 10.1016/j.jad.2018.12.098.
- [61] Kinney RK, Gatchel RJ, Mayer TG. The SCL-90R evaluated as an alternative to the MMPI for psychological screening of chronic low-back pain patients. *Spine*. 1991; 16(8): 940-942. doi: 10.1097/00007632-199108000-00013.
- [62] Bernstein IH, Jaremko ME, Hinkley BS. On the utility of the SCL-90-R with low-back pain patients. *Spine*. 1994; 19(1): 42-48. doi: 10.1097/00007632-199401000-00008.
- [63] Peebles JE, McWilliams LA, MacLennan R. A comparison of symptom checklist 90-revised profiles from patients with chronic pain from whiplash and patients with other musculoskeletal injuries. *Spine*. 2001; 26(7): 766-770. doi: 10.1097/00007632-200104010-00014.
- [64] Kroenke K, Baye F, Lourens SG. Comparative responsiveness and minimally important difference of common anxiety measures. *Medical Care*. 2019; 57(11): 890-897. doi: 10.1097/mlr.0000000000001185.
- [65] Goldsmith ES, Taylor BC, Greer N, Murdoch M, MacDonald R, McKenzie L, Rosebush CE, Wilt TJ. Focused evidence review: Psychometric properties of patient-reported outcome measures for chronic musculoskeletal pain. *Journal of General Internal Medicine*. 2018; 33(Suppl 1): 61-70. doi: 10.1007/s11606-018-4327-8.
- [66] Köke AJA. Eindrapportage Revalidatie Nederland implementatieproject IPR2010-01. Hoensbroek: Samenwerkende Ontwikkelcentra Pijnrevalidatie Nederland/Adelante Kenniscentrum. 2012.