



# Rapid Evidence Review: Measures for Patients with Chronic Musculoskeletal Pain

August 2017

## Prepared for:

Department of Veterans Affairs  
Veterans Health Administration  
Quality Enhancement Research Initiative  
Health Services Research & Development Service  
Washington, DC 20420

## Prepared by:

Evidence-based Synthesis Program (ESP) Center  
Minneapolis VA Health Care System  
Minneapolis, MN  
Timothy J. Wilt, MD, MPH, Director  
Nancy Greer, PhD, Program Manager

## Investigators:

Principal Investigator:  
Elizabeth S. Goldsmith, MD

Co-Investigators:  
Maureen Murdoch, MD, MPH  
Brent Taylor, PhD

Research Associates:  
Nancy Greer, PhD  
Roderick MacDonald, MS  
Lauren G. McKenzie, MPH  
Christina Rosebush, MPH



VA  
HEALTH  
CARE | Defining  
EXCELLENCE  
in the 21st Century

## PREFACE

The VA Evidence-based Synthesis Program (ESP) was established in 2007 to provide timely and accurate syntheses of targeted healthcare topics of particular importance to clinicians, managers, and policymakers as they work to improve the health and healthcare of Veterans. QUERI provides funding for four ESP Centers, and each Center has an active University affiliation. Center Directors are recognized leaders in the field of evidence synthesis with close ties to the AHRQ Evidence-based Practice Centers. The ESP is governed by a Steering Committee comprised of participants from VHA Policy, Program, and Operations Offices, VISN leadership, field-based investigators, and others as designated appropriate by QUERI/HSR&D.

The ESP Centers generate evidence syntheses on important clinical practice topics. These reports help:

- Develop clinical policies informed by evidence;
- Implement effective services to improve patient outcomes and to support VA clinical practice guidelines and performance measures; and
- Set the direction for future research to address gaps in clinical knowledge.

The ESP disseminates these reports throughout VA and in the published literature; some evidence syntheses have informed the clinical guidelines of large professional organizations.

The ESP Coordinating Center (ESP CC), located in Portland, Oregon, was created in 2009 to expand the capacity of QUERI/HSR&D and is charged with oversight of national ESP program operations, program development and evaluation, and dissemination efforts. The ESP CC establishes standard operating procedures for the production of evidence synthesis reports; facilitates a national topic nomination, prioritization, and selection process; manages the research portfolio of each Center; facilitates editorial review processes; ensures methodological consistency and quality of products; produces “rapid response evidence briefs” at the request of VHA senior leadership; collaborates with HSR&D Center for Information Dissemination and Education Resources (CIDER) to develop a national dissemination strategy for all ESP products; and interfaces with stakeholders to effectively engage the program.

Comments on this evidence report are welcome and can be sent to Nicole Floyd, ESP CC Program Manager, at [Nicole.Floyd@va.gov](mailto:Nicole.Floyd@va.gov).

**Recommended citation:** Goldsmith ES, Murdoch M, Taylor B, Greer N, MacDonald R, McKenzie LG, Rosebush C, Wilt TJ. Rapid Evidence Review: Measures for Patients with Chronic Musculoskeletal Pain. VA ESP Project #09-009; 2017.

This report is based on research conducted by the Evidence-based Synthesis Program (ESP) Center located at the **Minneapolis VA Health Care System, Minneapolis, MN**, funded by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Quality Enhancement Research Initiative. The findings and conclusions in this document are those of the author(s) who are responsible for its contents; the findings and conclusions do not necessarily represent the views of the Department of Veterans Affairs or the United States government. Therefore, no statement in this article should be construed as an official position of the Department of Veterans Affairs. No investigators have any affiliations or financial involvement (eg, employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties) that conflict with material presented in the report.

## TABLE OF CONTENTS

### ABSTRACT

<b>Objective</b> .....	<b>1</b>
<b>Methods</b> .....	<b>1</b>
<b>Results</b> .....	<b>1</b>
<b>Conclusions</b> .....	<b>2</b>
<b>Impacts</b> .....	<b>2</b>

### EVIDENCE REPORT

<b>Introduction</b> .....	<b>3</b>
<b>Key Question</b> .....	<b>4</b>
<b>Included Pain Measures</b> .....	<b>4</b>
<b>Methods</b> .....	<b>6</b>
Study Selection .....	6
Data Abstraction .....	7
Quality Assessment.....	7
Data Synthesis.....	7
Rating the Body of Evidence .....	7
Peer Review .....	7
<b>Results</b> .....	<b>8</b>
Literature Flow.....	8
Overview of Pain Measures and Included studies .....	9
Overview of Included Studies.....	9
Characteristics of Included Studies for Each Pain Measure .....	14
Outcomes .....	18
Minimally Important Difference.....	22
Test-retest Reliability .....	23
Inter-rater Reliability .....	23
Internal Consistency.....	23
Concurrent and/or Criterion Validity.....	23
Discriminant Validity.....	24
Responsiveness .....	24
Feasibility.....	26

**Summary and Discussion .....28**

Key Messages ..... 28

Discussion ..... 28

Challenges in Assessment of Psychometric Properties ..... 30

    Minimally Important Different (MID)..... 30

    Validity ..... 30

    Responsiveness ..... 31

    Test-retest Reliability..... 31

Limitations and Future Research ..... 31

    Limitations of a Rapid Evidence Review ..... 31

    Chronic Musculoskeletal Pain Definition and Reporting ..... 32

    Study Populations ..... 33

Applicability to VHA research ..... 33

Conclusions..... 33

**References .....34**

**TABLES**

Table 1. Overview of Pain Measures ..... 9

Table 2. Overview of Included Studies..... 11

Table 3. Summary of Results: Studies Assessing Psychometric Properties of Self-Report Measures of Pain Severity (S) and Functional Interference (I) in Chronic Musculoskeletal Pain Populations ..... 19

Table 4. Comparative External Responsiveness based on (AUC) Values for Detecting Any Improvement..... 26

**FIGURES**

Figure 1. Literature Flow Chart ..... 8



## ABSTRACT

### OBJECTIVE

Developing successful interventions for chronic musculoskeletal pain requires valid, responsive, and reliable outcome measures. By request of the 2016 State of the Art Conference on non-pharmacological approaches to chronic musculoskeletal pain, the Minneapolis VA Evidence-based Synthesis Program completed a rapid evidence review. We addressed a key question regarding psychometric properties of selected self-report pain measures to assist in adoption of these measures as core outcomes in clinical trials and other research of nonpharmacological approaches to chronic musculoskeletal pain.

### METHODS

With input from operational partners, we identified 17 English-language candidate measures. All measures assessed pain severity or intensity or pain-related functional impairment. Our primary outcome was the measure's minimally important difference (MID); secondary outcomes included the measure's reliability, validity, and responsiveness to change. We searched MEDLINE (Ovid) from January 2000 to January 2017 for English language publications. We also searched reference lists of relevant studies and systematic reviews and websites specific to pain measures of interest, with no publication date restrictions for these searches. We included studies that 1) evaluated at least one of the 17 pain measures; 2) included adults with chronic musculoskeletal pain of at least 3 months duration or adults with musculoskeletal pain described as "chronic" by the study authors; and 3) reported on at least one of the 4 psychometric outcomes listed above. We excluded 1) studies that used non-English language versions of the pain measures; 2) studies of acute musculoskeletal pain or studies of musculoskeletal conditions often associated with chronic pain that did not specify the presence or duration of their participants' pain; 3) intervention trials, unless the trial also assessed the psychometric properties of their measures and noted this in the abstract; and 4) studies of patients with rheumatoid arthritis, orofacial pain other than temporomandibular disorder, or headache. Abstracts and full text of articles meeting inclusion criteria were reviewed by trained staff, who extracted study/population characteristics and psychometric outcomes. Results were qualitatively synthesized. Our protocol was registered in PROSPERO (CRD42017056610).

### RESULTS

Of 1635 abstracts identified, 318 articles underwent full-text review, and 43 met inclusion criteria. Six of the 43 studies included Veteran populations. Eight studies provided MID estimates for 8 of the 17 measures. MIDs for individual measures differed considerably based on study design and analysis methods. Four measures – Oswestry Disability Index (ODI), Roland-Morris Disability Questionnaire (RMDQ), Numeric Rating Scale (NRS), and Visual Analog Scale (VAS) – had data reported on all 4 psychometric outcomes. However, the NRS and VAS, both single-item measures, were often modified across different studies; results from one study might therefore not apply to others using a different version. MIDs, responsiveness, and validity were reported for the Brief Pain Inventory (BPI), Global Chronic Pain Scale (GCPS), PEG, and Short Form 36 Bodily Pain Scale (SF-36 BPS). Responsiveness, validity, and test-retest

reliability estimates were reported for the McGill Pain Questionnaire (MPQ), PROMIS Pain Interference (PROMIS-PI), West Haven-Yale Multidimensional Pain Inventory (WHYMPI), and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC).

## CONCLUSIONS

Among the multi-item pain measures we assessed, the ODI, RMDQ, and SF-36 BPS had the most complete psychometric evidence within chronic musculoskeletal pain populations. Several additional measures had at least some evidence for psychometric reliability, validity, and responsiveness. Research into pain measurement would be considerably strengthened if future investigators use consistent definitions of chronic musculoskeletal pain, standardized methods for assessing psychometric outcomes, and comprehensive descriptions of their patient populations.

## IMPACTS

Findings from this review can inform recommendations on specific core outcome measures for clinical research on chronic musculoskeletal pain interventions. Further methods research is needed to validate patient-reported pain outcome measures in populations with chronic musculoskeletal pain and develop a framework for determining outcome measurement selection that incorporates feasibility and applicability.

# EVIDENCE REPORT

## INTRODUCTION

Chronic musculoskeletal pain is a major source of disability and morbidity for Veterans in the US, affecting approximately 60% of Veterans with chronic health conditions in Veterans Health Administration (VHA) primary care.<sup>1</sup> Management of chronic musculoskeletal pain remains challenging, and groups ranging from pain expert coalitions to the National Institutes of Health and the Institute of Medicine have recently called for more focused and strategic pain therapy research.<sup>2,3</sup> As these groups note, successful development and testing of interventions to improve chronic musculoskeletal pain depends on the use of valid, reliable, and responsive measures of pain and pain-related outcomes domains.

Pain-related measures span multiple physical, emotional, and social domains that are affected by chronic musculoskeletal pain. To guide development and use of these measures, experts and stakeholders have formed such initiatives as Outcome Measures in Rheumatology (OMERACT), the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks (ACTTION), public-private partnership with the United States Food and Drug Administration (FDA), and the associated Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (IMMPACT). These groups have published several reviews and compiled recommendations suggesting that pain outcome studies measure multiple domains via multiple modes of assessment.<sup>4-8</sup>

Such expert groups have identified both pain intensity or severity (hereafter “severity”) and pain-related impairment of physical function (hereafter “functional impairment”) as key domains for study, as these reflect both pain symptoms and pain’s impact on people’s daily lives.<sup>4,6</sup> Functional impairment in particular has been identified as a priority concern for patients,<sup>9</sup> and is an increasingly common primary outcome domain alongside pain severity. Self-report measures remain the gold-standard mode of assessing core pain outcomes, as they reflect the subjective pain experience, and as existing observer- and laboratory-based pain measures do not consistently reflect clinically meaningful changes in key pain domains.<sup>4,5,10</sup>

Researchers who wish to select appropriate self-report pain outcome measures for these key domains still face challenging evidence limitations. There is particular need for measures appropriate for non-pharmacological interventions. While available measures have been developed and adapted for multiple pain conditions and bodily locations, and have been studied in populations with a wide range of demographic traits, existing psychometric property and feasibility evidence is difficult to locate and compare across measures. Additionally, a consensus on ideal measures has not yet been achieved.

Therefore, it would be advantageous to have a core set of measures across intervention studies. This would make it easier to synthesize, disseminate, and provide recommendations to the VHA about the effectiveness and harms of different interventions. Even if evidence does not clearly demonstrate a single best measure or core set, identification of existing evidence would be informative.

As such, the 2016 State of the Art (SOTA) Conference on non-pharmacological approaches to chronic musculoskeletal pain management recognized the potential value of adopting a core set of measures and recommended that VA Health Services Research and Development (HSR&D) convene a small group of researchers to develop a short set of core outcome measures for prospective pain research. The set of measures should cover 2 core patient-reported outcomes: pain intensity and pain-related functioning. The group plans to consider many factors in selecting the core measures, choosing from among measures that have demonstrated suitable psychometric properties in the target population. The group requested a rapid evidence review to describe and compare the key psychometric qualities of commonly used measures, particularly those that might be suitable for clinical trials of nonpharmacological approaches to chronic pain management. These qualities would not be the only criterion for selecting core measures, but could serve as a basic requirement of measures considered candidates for wide implementation.

In conjunction with the topic nominators we identified the population of interest, pain measures to be reviewed, study inclusion and exclusion criteria, and primary and secondary outcomes and developed a protocol (registered in PROSPERO - CRD42017056610).

## KEY QUESTION

We addressed the following key question:

What specific self-report measures of pain (intensity, severity) and pain-related functional impairment (activity limitations, participation, physical functioning, social role functioning, pain impact, pain interference, pain-related disability) have sufficient information on psychometric properties (*eg*, minimally important differences, validity, responsiveness, reliability) to consider their adoption for use as core outcome measures in prospective observational research and clinical trials of nonpharmacological approaches to care for persons (including Veterans) with chronic ( $\geq 3$  months) musculoskeletal pain (*eg*, low back pain, osteoarthritis, and non-traumatic joint pain)?

## INCLUDED PAIN MEASURES

Our review focused on the following measures of pain intensity/severity, pain-related interference, or pain global change for persons with chronic musculoskeletal pain (as identified by the Operational Partners for the review and the SOTA Planning Committee):

Brief Pain Inventory (BPI)

Defense & Veterans Pain Rating Scale (DVPRS)

Graded Chronic Pain Scale (GCPS)

Hip Osteoarthritis Outcomes Scale (HOOS)

Knee Osteoarthritis Outcomes Scale (KOOS)

McGill Pain Questionnaire (MPQ)

Multidimensional Pain Inventory (MPI, WHYMPI)

Numeric Rating Scale (NRS)

Oswestry Disability Index (ODI)

Patient Global Impression of Change (PGIC)

PEG (assesses [P] pain intensity, [E] enjoyment of life, and [G] general activity)

Patient-Reported Outcomes Measurement Information System - Pain Interference (PROMIS-PI)  
Roland-Morris Disability Questionnaire (RMDQ)  
SF-36 Bodily Pain Scale (SF-36 BPS)  
Visual Analogue Scale (VAS)  
Western Ontario and McMaster Universities Arthritis Index (WOMAC)  
Wong Faces Scale

## METHODS

We searched MEDLINE (Ovid) for English-language articles published from 2000 to January 2017. Our search strategy, developed with input from a medical librarian, included Medical Subject Heading (MeSH) terms for Pain Measurement and specific locations/types of pain (*eg*, Low Back, Shoulder, Chronic) along with title and abstract words. The search was designed to include all study designs, including systematic reviews. The full search strategy is presented in Supplemental Content, Table 1. At the request of peer reviewers, we repeated the search with MeSH and title/abstract terms for fibromyalgia.

We used Google Scholar, the National Center for Biotechnology Information (NCBI), and PubMed to search for Web sites associated with each pain measure and publications not retrieved by our MEDLINE search. Additional articles were obtained by reviewing reference lists of relevant systematic reviews identified in our MEDLINE search and reference lists of included studies. We also reviewed studies suggested by content experts. *For these sources, there were no limits on publication date.*

## STUDY SELECTION

Abstracts of studies identified in our MEDLINE search were reviewed by a single investigator or research associate. The full text of potentially eligible articles from the abstract review and all articles identified from reference list searching or suggested by content experts were reviewed by 2 investigators or research associates.

At the abstract and full-text review levels, we included studies that:

- 1) Evaluated pain measures in adults with chronic musculoskeletal pain of at least 3 months duration (or was described as “chronic pain” by the study authors); if the study included multiple types of pain, at least 75% of the population must have had chronic musculoskeletal pain unless results were reported separately for the chronic musculoskeletal pain group,
- 2) Reported on self-reported measures of pain or pain-related functioning (17 measures as determined by Operational Partners and SOTA Planning Committee),
- 3) Reported outcomes of interest: minimally important difference (MID) (primary outcome), test-retest reliability, validity, feasibility (*ie*, number of items, public domain vs proprietary, self-report vs interviewer-administered), responsiveness, and generalizability.

Our exclusion criteria were as follows:

- 1) Studies that specified that they used non-English-language versions of the pain measures,
- 2) Studies of patients with chronic musculoskeletal conditions commonly associated with pain but without specifying that enrolled patients had chronic musculoskeletal pain (*eg*, osteoarthritis),
- 3) Trials of interventions for pain that did not note assessment of psychometric properties in the abstract,

4) Studies of patients with rheumatoid arthritis, orofacial pain (other than temporomandibular joint pain – a musculoskeletal condition), or headache.

## DATA ABSTRACTION

From each eligible study, we abstracted the following:

1) Study/population characteristics: location of study, funding source, pain measures evaluated, time period of assessment (*eg*, reporting pain over past week, past month, *etc*), mode of administration, setting, chronic pain condition, study inclusion/exclusion criteria, baseline pain characteristics, sample size, age, gender, and race/ethnicity,

2) Outcomes: MID, reliability, validity, responsiveness, and other psychometric properties.

## QUALITY ASSESSMENT

We included only studies that discussed psychometric properties of the pain measures. Trials that used the measures but did not comment on how well the measures performed were not included.

## DATA SYNTHESIS

We narratively summarized included studies by pain measure to provide an overview of the populations and pain conditions for which the psychometric properties of the measure have been evaluated. We narratively summarized outcomes by psychometric properties. We focused on MID, responsiveness, validity, and test-retest reliability and highlighted comparative effectiveness when reported.

## RATING THE BODY OF EVIDENCE

We did not rate the overall body of evidence.

## PEER REVIEW

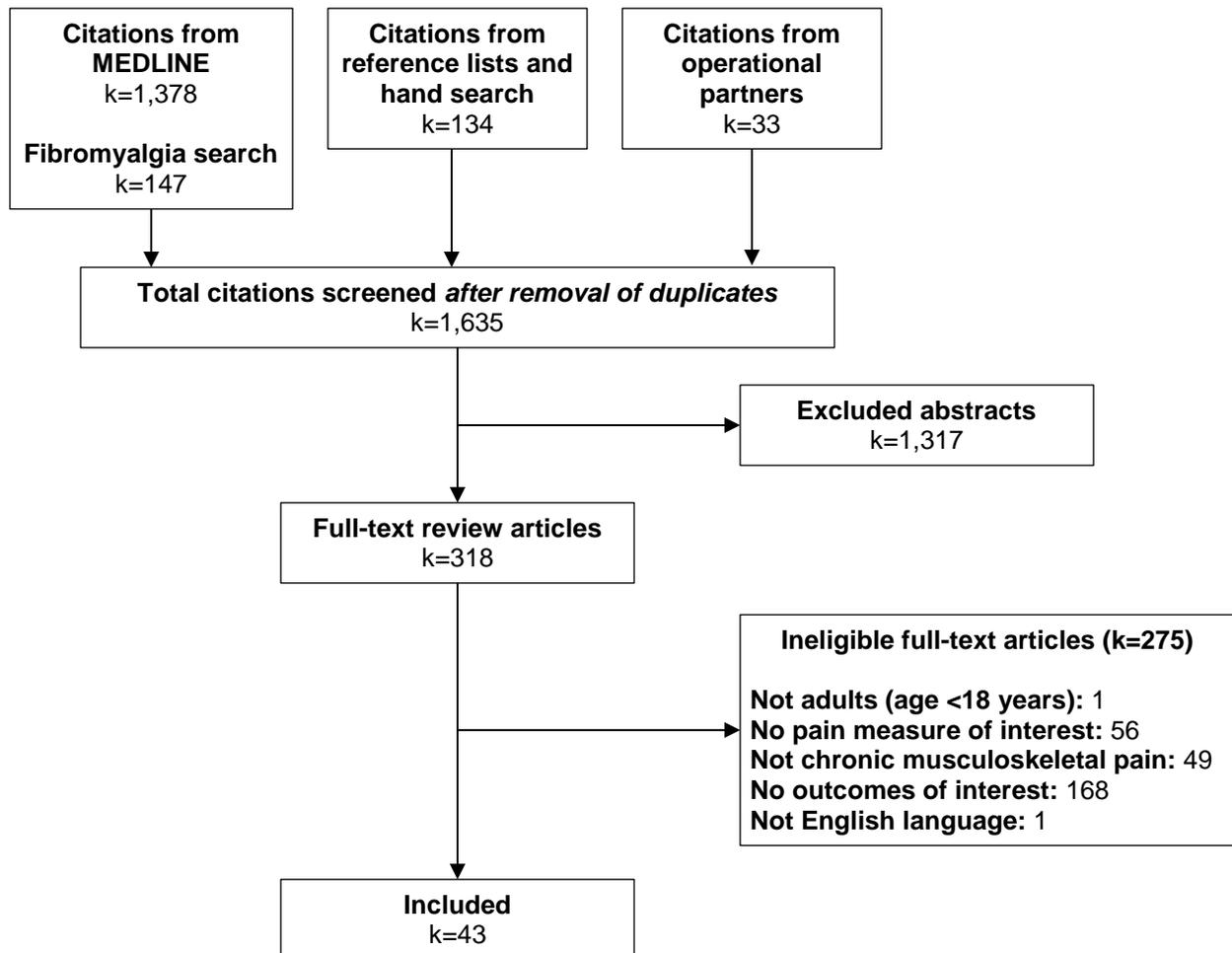
A draft version of this report was reviewed by content experts and clinical leadership, and the report was modified in response to reviewers' input. Reviewer comments and our responses are presented in Supplemental Content, Table 2.

## RESULTS

### LITERATURE FLOW

After removing duplicate citations, we reviewed 1,635 abstracts and excluded 1,317. Of 318 articles reviewed at the full text level, 275 were excluded (Figure 1). Over 60% were excluded because they did not report outcomes of interest. Other reasons for exclusion were not including a pain measure of interest, using a non-English version of the pain measure, and not defining the study population as having chronic musculoskeletal pain.

**Figure 1. Literature Flow Chart**



## OVERVIEW OF PAIN MEASURES AND INCLUDED STUDIES

Table 1 below summarizes the characteristics of the pain measures included in the review. Additional information about each pain measure is included in Supplemental Content, Table 3.

**Table 1. Overview of Pain Measures**

Pain Measure	Development Pain Type				Pain Domain		Length	Restrictions on Use
	General	LBP	Knee/hip	Other	Severity/Intensity	Function/Interference	Number of Items	
BPI <sup>11</sup>				X	X	X	11	Yes
DVPRS <sup>12</sup>	X				X	X	5	No
GCPS <sup>13</sup>	X				X	X	7	No
HOOS <sup>14</sup>			X		X	X	40	No
KOOS <sup>15</sup>			X		X	X	42	No
MPQ <sup>16</sup>	X				X		78 <sup>a</sup>	No
MPI/WHYMPI <sup>17</sup>	X	X			X	X	52	No
NRS for Pain <sup>16</sup>	X				X	X	1	No
ODI <sup>18</sup>		X			X	X	10	Yes
PGIC <sup>19</sup>	?				?	?	1	No
PEG <sup>20)</sup>	X				X	X	3	No
PROMIS-PI <sup>21</sup>	X					X	41	Yes
RMDQ <sup>22</sup>		X				X	24	No
SF-36 BPS <sup>23</sup>	X				X	X	2	Yes
VAS for Pain <sup>24</sup>				X	X	X	1	No
WOMAC <sup>25</sup>			X		X	X	24	Yes
Wong Faces Scale <sup>26</sup>	X			X	X		1	Yes

?=not identified

S=pain severity; I=pain interference

<sup>a</sup>Pain Rating Index (PRI) based on 78 pain descriptors; Present Pain Intensity (PPI) based on 6 additional items

### Overview of Included Studies

We included 43 studies: 23 from the US,<sup>17,20,27-47</sup> 3 from Canada,<sup>48-50</sup> one from South America,<sup>51</sup> 5 from Australia,<sup>52-56</sup> and 11 from Europe.<sup>57-67</sup> Of the US studies, 4 enrolled exclusively Veterans<sup>17,35,39,44</sup> and 2 enrolled both Veterans and non-Veterans.<sup>20,37</sup> Study characteristics are presented on Table 2 with additional detail in Supplemental Content, Table 4.



Study enrollments ranged from 30<sup>62</sup> to 998,<sup>43</sup> with 29 enrolling more than 100 and 3 enrolling more than 500.<sup>29,34,43</sup> The most common chronic musculoskeletal pain condition was low back pain (LBP) with 16 studies enrolling only LBP patients.<sup>28,29,33,34,36,39,41,45,52,55,57-59,61,63,66</sup> Another 13 studies included patients with any chronic musculoskeletal pain.<sup>27 17,30,35,38,43,44,47,49,51,53,62,65</sup> One study reported that 62% of participants were over age 50 years.<sup>28</sup> In the remaining 40 studies that reported mean age, values ranged from 32 years<sup>67</sup> to 80 years.<sup>33</sup> The mean age was less than 50 years in 18 studies, 50 to 59 years in 15 studies, and 60 years and older in 7 studies. In the studies that enrolled exclusively US Veterans, the percentage of women ranged from 8% to 19%. In the remaining studies, 5 studies enrolled fewer than 50% women,<sup>32,42,52,62,64</sup> 29 enrolled 50% or more, and 5 did not report the percentage of women enrolled. Race/ethnicity was reported in 18 of the studies, all but one from the US. The percentage of white enrollees was 75% or higher for 11 of the 18 studies.

No studies meeting eligibility criteria evaluated psychometric properties of the DVPRS or KOOS. DVPRS studies intermixed patients with chronic and acute pain and either had fewer than 75% of patients with chronic pain<sup>68</sup> or did not specify the percentage with chronic pain.<sup>12</sup> Studies of the KOOS used non-English versions.<sup>15,69</sup>

**Table 2. Overview of Included Studies**

Author Year	Pain Measures Evaluated	Study Characteristics				
		Sample Size	Pain Condition	Mean Age (years)	Women (%)	Race/Ethnicity (%)
<b>Anagnostis 2004</b> <sup>27</sup>	ODI	230	CDMD	43	53	White: 60 African American/Black: 29 Hispanic: 11 Other: 0.1
<b>Askew 2016</b> <sup>28</sup>	PROMIS-PI	218	LBP	NR	56	White: 84 African American/Black: 4 Other: 12
<b>Burnham 2012</b> <sup>48</sup>	MPQ, ODI	60	Spine	60	67	NR
<b>Changulani 2009</b> <sup>57</sup>	ODI, VAS	107	LBP	58	58	NR
<b>Chansirinukor 2005</b> <sup>52</sup>	RMDQ	143	LBP	38	26	NR
<b>Chien 2013</b> <sup>53</sup>	BPI	254	General MSP	51	50	NR
<b>Cook 2008</b> <sup>29</sup>	RMDQ	875	LBP	47	NR	White: 85 African American/Black: 9 Other: 6
<b>de Vet 2007</b> <sup>58</sup>	NRS	438	LBP	NR	NR	NR
<b>Deyo 2016</b> <sup>30</sup>	PROMIS-PI-SF	198	General MSP	67	62	White: 92 Hispanic: 4 Other: 4
<b>Driban 2015</b> <sup>31</sup>	PROMIS-PI, SF-36 BPS, WOMAC	204	Knee (OA)	60	70	White: 53 African American/Black: 36 Other: 12
<b>Fisher 1997</b> <sup>59</sup>	MPQ, ODI	54	LBP	41	63	NR
<b>Gallasch 2007</b> <sup>51</sup>	Wong Faces, NRS, VAS	32	General MSP	51	NR	NR
<b>Gentelle-Bonnassies 2000</b> <sup>60</sup>	VAS, WOMAC	80	Knee (OA)	62	70	NR
<b>Godil 2015</b> <sup>32</sup>	NRS	88	Neck and arm	52	44	NR
<b>Gronblad 1993</b> <sup>61</sup>	ODQ, VAS	94	LBP	43	51	NR

Author Year	Pain Measures Evaluated	Study Characteristics				
		Sample Size	Pain Condition	Mean Age (years)	Women (%)	Race/Ethnicity (%)
Hicks 2009 <sup>33</sup>	ODI, SF-36 BPS	107	LBP	80	72	White: 100
Jensen 2012 <sup>34</sup>	VAS	639	LBP	52	62	White: 90 African American/Black: 5 Other: 5
Kamper 2015 <sup>54</sup>	NRS, SF-36 BPS	280	Whiplash	44	65	NR
Kean 2016 <sup>a35</sup>	BPI, PEG, PROMIS-PI-SF, SF-36 BPS	244	MSP	55	17	White: 77 African American/Black: 19 Other: 4
Keller 2004 <sup>36</sup>	BPI, GCPS, SF-36 BPS, RMDQ	131	LBP	46	NR	NR
Kerns 1985 <sup>a17</sup>	MPI (WHYMPI), MPQ	120	Chronic MSP	51	19	NR
Krebs 2010 <sup>b37</sup>	BPI, GCPS, PEG, RMDQ, SF-36 BPS	427	Back, hip, knee	59	53	White: 58 African American/Black: 38 Other: 4
Krebs 2009 <sup>b20</sup>	BPI, GCPS, PEG, PGIC, RMDQ, SF- 36 BPS	500	Back, hip, knee	59	52	White: 58 African American/Black: 38 Other: 4
Krebs 2007 <sup>38</sup>	NRS	275	General MSP	59	59	White: 70 African American/Black: 24 Other: 6
Lovejoy 2012 <sup>a39</sup>	MPI, MPQ-2-SF, MPQ	186	LBP, neck, joint	54	8	White: 75 Other: 15
Lund 2005 <sup>62</sup>	VAS	30	MSP	43	43	NR
Macedo 2011 <sup>55</sup>	RMDQ	461	LBP	53	61	NR
Maughan 2010 <sup>63</sup>	NRS, ODI, RMDQ	48	LBP	52	67	NR
Merriwether 2016 <sup>40</sup>	PROMIS-PI	106	Fibromyalgia	49	100	White: 96 Other: 4

Author Year	Pain Measures Evaluated	Study Characteristics				
		Sample Size	Pain Condition	Mean Age (years)	Women (%)	Race/Ethnicity (%)
Mikail 1993 <sup>49</sup>	MPI, ODI	315	General MSP	44	53	NR
Nilsson 2003 <sup>64</sup>	HOOS, WOMAC, SF-36 BPS	62	Hip (OA)	73	45	NR
Parker 2012 <sup>41</sup>	ODI, VAS	47	LBP	55	64	NR
Pinsker 2015 <sup>50</sup>	NRS, WOMAC	142	Ankle	61	54	NR
Scott 2015 <sup>65</sup>	PGIC	476	Not specified	46	67	White: 72 African American/Black: 17 Other: 11
Sindhu 2011 <sup>42</sup>	NRS, VAS	33	Elbow, forearm, hand	39	48	NR
Stewart 2007 <sup>56</sup>	NRS, SF-36 BPS	132	Whiplash	43	67	NR
Stroud 2004 <sup>43</sup>	RMDQ	998	Not specified	44	57	White: 84 African American/Black: 3 Native American: 4 Other: 9
Tan 2004 <sup>a44</sup>	BPI, RMDQ	440	Not specified	55	8	White: 72 African American/Black: 21 Other: 7
Tong 2006 <sup>45</sup>	VAS	52	LBP	41	62	White: 88 African American/Black: 3 Other: 9
Trudeau 2015 <sup>46</sup>	WOMAC, NRS	47	Knee (OA)	NR	NR	NR
van der Roer 2006 <sup>66</sup>	NRS	138	LBP	44	59	NR
van Grootel 2007 <sup>67</sup>	VAS	118	TMD	32	93	NR
Wittink 2004 <sup>47</sup>	MPI, ODI, SF-36 BPS	87	Chronic pain	47	67	White: 79 Other: 21

<sup>a</sup>Enrolled exclusively US Veterans

<sup>b</sup>Enrolled US Veterans and non-Veterans; results not stratified by Veteran status

**Abbreviations:** CDMD=chronic disabling musculoskeletal disorders; LBP=low back pain; MSP=musculoskeletal pain; NR=not reported; OA=osteoarthritis; TMD=temporomandibular disorder

## Characteristics of Included Studies for Each Pain Measure

### *Brief Pain Inventory (BPI)*

The BPI is a Likert-type scale (range 0-10) originally designed to measure cancer pain intensity and pain interference.<sup>11</sup> Pain intensity is measured by 4 items: current pain, and pain at its least, worst, and average over a time of interest (often the past 24 hours or week). Pain interference is measured for 7 domains: physical functioning, work, mood, walking, social activity, relations with others, and sleep. Scores for each BPI measure range from 0 “no pain/interference” to 10 “pain as bad as you can imagine/complete interference.”

We included 6 studies that evaluated the BPI’s psychometric properties (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>20,35-37,44,53</sup> One study<sup>53</sup> assessed only the BPI’s pain severity subscale.

The BPI was administered by interview in 3 studies<sup>20,35,37</sup> and by self-report in another 3 studies.<sup>36,44,53</sup>

### *Defense and Veterans Pain Rating Scale (DVPRS)*

The DVPRS was developed to provide a standardized pain screening and assessment tool for the Department of Defense and VHA health systems.<sup>12</sup> It includes numeric rating scales for one question about pain intensity and 4 questions about pain interference. The numeric scale for pain intensity ranges from 0 to 10 and is enhanced with descriptors for each of the 11 levels, color-coded bars using traffic light colors where green indicates mild pain and red indicates severe pain, and facial expressions. The pain interference questions address activity, sleep, mood, and stress. We found no studies meeting eligibility criteria for the DVPRS.

### *Graded Chronic Pain Scale (GCPS)*

The Graded Chronic Pain Scale (GCPS), also known as Chronic Pain Grade Questionnaire (CPG) is an interview or self-administered measure used to assess pain intensity and interference related to disability.<sup>13</sup> It was designed in 1992 for use with chronic pain conditions including musculoskeletal and low back pain. Pain intensity is measured on an 11-point Likert scale from 0-10 anchored by “no pain” (0) and “pain as bad as can be” (10). The disability score is based on the number of days of disability and a numeric rating of pain disability.

We included 3 studies that evaluated psychometric properties of the GCPS (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>20,36,37</sup> Each of the studies assessed both the severity and disability components.

### *Hip Osteoarthritis Outcomes Scale (HOOS)*

The Hip Osteoarthritis Outcomes Scale (HOOS) was developed in 2002 as an extension of the WOMAC scale for hip disability among people with or without osteoarthritis.<sup>14</sup> The self-administered HOOS evaluates pain intensity and interference related to physical functioning. The HOOS consist of 5 subscales: pain, symptoms, daily living limitations, sport and recreation limitations, and hip related quality of life. The HOOS uses a 5-point Likert type scale with anchors of “no problems” (0) to “extreme problems” (4).

One study of the HOOS was included in our review (details of the study and participant characteristics are provided Supplemental Content, Table 4).<sup>64</sup> We report outcomes from the pain and activities of daily living limitations subscales.

### *Knee Osteoarthritis Outcomes Scale (KOOS)*

The KOOS is an extension of the WOMAC scale. It was designed to assess patient-relevant outcomes following a knee injury or post-traumatic osteoarthritis.<sup>15</sup> Responses to the 42 items are on a 5 point scale ranging from “none” or “never” to “extreme” or “always.” The 42 items are grouped into pain intensity, symptoms, activities of daily living, sport and recreation, and quality of life subcategories. We found no studies meeting eligibility criteria for the KOOS.

### *McGill Pain Questionnaire (MPQ)*

The MPQ measures general chronic pain using 78 items in 20 subscales. It is used to evaluate pain intensity.<sup>16</sup> Respondents are asked to respond to sensory, affective, and evaluative word descriptors of their pain. Responses are used to create a Pain Rating Index (PRI) and/or a Total Number of Words Chosen score. There is also a single item, the Present Pain Intensity (PPI), with pain rated from 0 to 5. Two revised forms of the MPQ exist: the short-form MPQ (SF-MPQ) and a revised and extended short-form MPQ (SF-MPQ-2).

We included 4 studies that assessed the psychometric properties of the MPQ (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>17,39,48,59</sup> Each of the studies assessed pain intensity using the Present Pain Intensity<sup>17,59</sup>; Total Number of Words Chosen<sup>59</sup>; Total Pain Rating Index<sup>17</sup>; total score over the continuous, intermittent, neuropathic, and affective domains<sup>39</sup>; or Adjective Checklist.<sup>48</sup> The MPQ was self-administered in all of the studies. One study administered a short-form version<sup>39</sup>; the others used the original version.

### *Multidimensional Pain Inventory (MPI/WHYMPI)*

The 52-item MPI, also known as the West Haven-Yale Multidimensional Pain Inventory (WHYMPI) was designed to measure chronic pain, including lower back pain and temporomandibular disorders.<sup>17</sup> It uses a Likert-type scale of 0-6 to measure pain intensity and pain interference. Pain interference is measured for daily activities including vocational, social, and familial functioning.

We included 4 studies that evaluated properties of the MPI (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>17,39,47,49</sup> Each of the studies assess both pain intensity and pain interference.

### *Numeric Rating Scale (NRS) for Pain*

Numerical Rating Scales (NRS) were developed to measure pain intensity for general chronic pain conditions.<sup>16</sup> The NRS studied for this report was typically an 11-point Likert type scale ranging from 0 (no pain) to 10 (severe pain), with subcategories of mild (1-3) and moderate (4-6). This self-administered NRS can be written or verbal. Of the 11 included studies, 2 administered the NRS by mail or by phone.<sup>32,50</sup>

We included 11 studies for psychometric properties of the NRS (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>32,38,42,46,50,51,54,56,58,63,66</sup> All of the

studies used the NRS to assess pain intensity. One study also assessed pain “bothersomeness” – a measure of interference or functional impairment.<sup>56</sup>

The timeframe over which patients were asked to rate pain intensity differed across the studies, such that some asked patients to rate their “current pain,”<sup>38,50</sup> their average pain over the last 24 hours,<sup>54,56</sup> their pain on the day prior to their study visit,<sup>51</sup> their pain in the past week,<sup>46,50</sup> or their pain in the past month.<sup>50</sup> Several did not specify or report a timeframe,<sup>32,58,66</sup> and one study asked patients to rate their pain intensity before and after a hand grip test.<sup>42</sup>

### *Oswestry Disability Index (ODI)*

The ODI was developed to assess disability from acute and chronic lower back pain.<sup>18</sup> It measures a combination of pain intensity and interference, referred to as disability, using a Likert-type scale. Scores range from 0 “no pain/interference/disability” to 5 “worst scenario of pain/interference/disability.” The ODI includes 10 items, one for pain or need for pain medications and 9 for interference in daily activities.

Ten studies that evaluated the psychometric properties of the ODI met our criteria for inclusion in this review (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>27,33,41,47-49,57,59,61,63</sup> All studies reported using self-administered questionnaires with one administered through the mail.<sup>33</sup>

### *Patient Global Impression of Change (PGIC)*

The Patient Global Impression of Change (PGIC) is a Likert-type scale used to assess the respondent’s overall impression of change in pain, often following an intervention.<sup>19</sup> Two studies that reported on the PGIC were included in our review (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>20,65</sup> In one study, pain and function were assessed.<sup>65</sup> The other study used the scores on the PGIC to categorize whether pain intensity was improved, unchanged, or worse over a 6 month period.<sup>20</sup>

### *PEG*

The PEG is a 3-item pain questionnaire designed to quickly assess chronic pain in primary care settings. Respondents are asked about pain intensity (P), interference with enjoyment of life (E), and interference with general activity (G) in the past week. Each item is assessed on a Likert-type scale 0-10, and individual item scores are averaged. Questions on the PEG are derived from the longer, more comprehensive BPI.

Three studies included in our review evaluated the psychometric properties of the PEG (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>20,35,37</sup> In all 3 studies, the PEG was administered by an interviewer.

### *Patient-Reported Outcomes Measurement Information System - Pain Interference (PROMIS-PI)*

The PROMIS-PI was developed in 2004 and is used for general chronic pain conditions to examine interference related to physical functioning.<sup>21</sup> PROMIS-PI consists of a 5-point Likert type scale corresponding to 1 (not at all) and 5 (very much). This PROMIS-PI can be self-administered, interview-administered, or administered through a proxy.

Five studies were included that examined the psychometric properties of the PROMIS-PI (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>28,30,31,35,40</sup>

### *Roland Morris Disability Questionnaire (RMDQ)*

The Roland Morris Disability Questionnaire (RMDQ) was developed in 1983 to evaluate disability and physical functioning interference from low back pain.<sup>22</sup> The RMDQ is self-administered with 24 items scored from 0 (no disability) to 24 (severe disability). Since its origination, 11-item, 12-item, and 18-item versions have been developed.

Psychometric properties were assessed in 9 studies (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>20,29,36,37,43,44,52,55,63</sup> All but 2 studies<sup>29,52</sup> administered the 24-item version of the RMDQ. Three studies assessed multiple versions.<sup>29,43,55</sup>

### *SF-36 Bodily Pain Scale (SF-36 BPS)*

The SF-36 Bodily Pain Scale (SF-36 BPS) uses 2 items to assess pain intensity and interference in daily activities over the past 4 weeks.<sup>23</sup> The Bodily Pain Scale is one of 8 scaled scores in the SF-36, a measure of overall health status.

We included 10 studies that evaluated the psychometric properties of the SF-36 BPS in our review (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>20,31,33,35-37,47,54,56,64</sup> Four studies asked participants to complete the SF-36 in its entirety and reported results specific to the bodily pain scale.<sup>31,33,47,64</sup> One study used only the pain intensity question from the bodily pain scale.<sup>54</sup>

One study used an interviewer-administered SF-36.<sup>35</sup> The remaining studies used a self-administered questionnaire (SAQ). One of these studies specified that the questionnaire was mailed.<sup>33</sup>

### *Visual Analogue Scale (VAS) for Pain*

Development of the Visual Analogue Scale (VAS) dates to 1952. It is used to measure pain intensity and interference related to disability.<sup>24</sup> The VAS is composed of an incrementally measured vertical line anchored with 2 opposing descriptors, such as “no pain” and “pain as bad as can be” when measuring pain intensity. The participant then places a perpendicular line at the point that best describes their pain. A ruler is then used to indicate the score.

Ten studies were included that assessed the psychometric properties of the Visual Analogue Scale (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>34,41,42,45,51,57,60-62,67</sup> One study did not specify whether the VAS was used to assess pain severity or interference.<sup>57</sup> In the other 9 studies, pain intensity was assessed.

Patients were asked to rate pain during the week after physical activity in one study.<sup>60</sup> One study asked patients to rate their pain in the last week.<sup>45</sup> A third study required patients to keep a VAS log of their pain for 14 days.<sup>67</sup> Another study asked participants to rate their pain level on the previous day.<sup>51</sup> One study asked patients to rate their change in pain from baseline (3 month

study period)<sup>34</sup> while another asked patients to rate pain prior to surgery and at 2-year follow-up.<sup>41</sup> One study asked patients to rate pain before and after performing grip exercises.<sup>42</sup> One study assessed present pain.<sup>61</sup> Two studies did not specify a timeframe.<sup>57,62</sup>

### *Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)*

The WOMAC was developed in 1982 for assessing pain severity and function in individuals with knee and hip pain associated with osteoarthritis.<sup>25</sup> Another domain, stiffness, is not addressed in this review. The index includes 24 items and can be self-administered or completed by interview. Different response formats have been used including a 5-point Likert scale, 11-point numerical rating scale, and a 100-mm visual analog scale.

We included 5 studies of the WOMAC (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>31,46,50,60,64</sup> One study assessed pain severity<sup>46</sup> while 4 studies used the WOMAC to assess both pain and function.

In all studies, the WOMAC was self-administered. One used a postal survey.<sup>50</sup> Two studies specified that participants were asked to recall pain over the past 48 hours.<sup>31,46</sup> The others did not specify a timeframe.

### *Wong Faces Scale/Wong-Baker Faces Scale*

Wong Faces Scale (also known as Wong-Baker Faces Scale) is an interview-administered, 6-point Likert-type scale ranging from 0 to 5 with corresponding faces.<sup>26</sup> Higher numbers represent greater pain. It was originally developed in 1985 to assess general pain intensity among children.<sup>26</sup>

We included one study that measured the psychometric properties of the Wong-Baker Faces Scale (details of study and participant characteristics are provided in Supplemental Content, Table 4).<sup>51</sup>

## **OUTCOMES**

Table 3 provides an overview of included pain measures and studies reporting each outcome. Of the measures that include assessment of both pain severity and pain interference, we found the greatest reporting of psychometric properties for the BPI, GCPS, MPI/WHYMPI, PEG, SF-36 BPS, and WOMAC. Of the measures that primarily assessed pain severity, we found the greatest reporting of psychometric properties for the NRS and VAS followed by the MPQ. Of the measures of pain interference, we found the greatest reporting of psychometric properties for the ODI, PROMIS-PI, and RMDQ. There was little or no reporting of psychometric properties for the DVPRS, HOOS, KOOS, PGIC, or Wong Faces Scale. Detailed psychometric data are reported in Supplemental Content, Table 5 and summarized below.

**Table 3. Summary of Results: Studies Assessing Psychometric Properties of Self-Report Measures of Pain Severity (S) and Functional Interference (I) in Chronic Musculoskeletal Pain Populations**

Pain Measure	Number of studies	Total Participants	MID	Responsiveness	Concurrent validity	Discriminant validity	Test-retest reliability
<b>Brief Pain Inventory (BPI)</b>	6	1,996	Krebs 2010 <sup>b37</sup> (S,I)	Chien 2013 <sup>53</sup> (S) Kean 2016 <sup>a35</sup> (S,I) Keller 2004 <sup>36</sup> (S,I) Krebs 2010 <sup>b37</sup> (S,I) Krebs 2009 <sup>b20</sup> (S,I) Tan 2004 <sup>a44</sup> (S,I)	Keller 2004 <sup>36</sup> (S,I) Krebs 2009 <sup>b20</sup> (S,I) Tan 2004 <sup>a44</sup> (S,I)	-	-
<b>Defense and Veterans Pain Rating Scale (DVPRS)</b>	0	-	-	-	-	-	-
<b>Graded Chronic Pain Scale (GCPS)</b>	3	1,058	Krebs 2010 <sup>b37</sup> (S,I)	Keller 2004 <sup>36</sup> (S,I) Krebs 2010 <sup>b37</sup> (S,I)	Keller 2004 <sup>36</sup> (S,I) Krebs 2009 <sup>b20</sup> (S,I)	-	-
<b>Hip Osteoarthritis Outcomes Scale (HOOS)</b>	1	62	-	Nilsson 2003 <sup>64</sup> (S,I)	Nilsson 2003 <sup>64</sup> (S,I)	-	-
<b>Knee Osteoarthritis Scale (KOOS)</b>	0	-	-	-	-	-	-
<b>McGill Pain Questionnaire (MPQ)</b>	3	366	-	Burnham 2012 <sup>48</sup> (S)	Kerns 1985 <sup>a17</sup> (S) Lovejoy 2012 <sup>a39</sup> (S)	Lovejoy 2012 <sup>a39</sup> (S)	Burnham 2012 <sup>48</sup> (S)
<b>Multidimensional Pain Inventory (MPI/WHYMPI)</b>	4	708	-	Wittink 2004 <sup>47</sup> (S,I)	Kerns 1985 <sup>a17</sup> (S,I) Lovejoy 2012 <sup>a39</sup> (S,I) Mikail 1993 <sup>49</sup> (S,I) Wittink 2004 <sup>47</sup> (S,I)	-	Kerns 1985 <sup>a17</sup> (S,I)

Pain Measure	Number of studies	Total Participants	MID	Responsiveness	Concurrent validity	Discriminant validity	Test-retest reliability
<b>Numerical Rating Scale (NRS) for Pain</b>	11	1,653	De Vet 2007 <sup>58</sup> (S) Maughan 2010 <sup>63</sup> (S) Van der Roer 2006 <sup>66</sup> (S)	Godil 2015 <sup>32</sup> (S) Maughan 2010 <sup>63</sup> (S) Sindhu 2011 <sup>42</sup> (S) Stewart 2007 <sup>56</sup> (S, I) Trudeau 2015 <sup>46</sup> (S)	De Vet 2007 <sup>58</sup> (S) Kamper 2015 <sup>54</sup> (S) Pinsker 2015 <sup>50</sup> (S) Sindhu 2011 <sup>42</sup> (S)	Krebs 2007 <sup>38</sup> (S)	Gallasch 2007 <sup>51</sup> (S) Sindhu 2011 <sup>42</sup> (S)
<b>Oswestry Disability Index (ODI)</b>	10	1,149	Hicks 2009 <sup>33</sup> (I) Maughan 2010 <sup>63</sup> (I) Parker 2012 <sup>41</sup> (I)	Anagnostis 2004 <sup>27</sup> (I) Burnham 2012 <sup>48</sup> (I) Changulani 2009 <sup>57</sup> (I) Maughan 2010 <sup>63</sup> (I) Wittink 2004 <sup>47</sup> (I)	Changulani 2009 <sup>57</sup> (I) Fisher 1997 <sup>59</sup> (I) Gronblad 1993 <sup>61</sup> (I) Hicks 2009 <sup>33</sup> (I) Mikail 1993 <sup>49</sup> (I) Wittink 2004 <sup>47</sup> (I)	Hicks 2009 <sup>33</sup> (I)	Burnham 2012 <sup>48</sup> (I) Gronblad 1993 <sup>61</sup> (I) Hicks 2009 <sup>33</sup> (I)
<b>Patient Global Impression of Change (PGIC)</b>	1	476	-	Scott 2015 <sup>65</sup> (S,I)		-	-
<b>PEG</b>	3	1,171	Krebs 2010 <sup>b37</sup> (S,I)	Kean 2016 <sup>a35</sup> (S,I) Krebs 2010 <sup>b37</sup> (S,I) Krebs 2009 <sup>b20</sup> (S,I)	Krebs 2009 <sup>b20</sup> (S,I)	-	-
<b>Patient-reported Outcomes Measurement Information System-Pain Interference (PROMIS-PI)</b>	4	864	-	Askew 2016 <sup>28</sup> (I) Deyo 2016 <sup>30</sup> (I) Kean 2016 <sup>a35</sup> (I)	Driban 2015 <sup>31</sup> (I)	Deyo 2016 <sup>30</sup> (I)	Deyo 2016 <sup>30</sup> (I)
<b>Roland-Morris Disability Questionnaire (RMDQ)</b>	9	4,023	Chansirinukor 2005 <sup>52</sup> (I) Krebs 2010 <sup>b37</sup> (I) Maughan 2010 <sup>63</sup> (I)	Chansirinukor 2005 <sup>52</sup> (I) Krebs 2010 <sup>b37</sup> (I) Macedo 2011 <sup>55</sup> (I) Maughan 2010 <sup>63</sup> (I)	Cook 2008 <sup>29</sup> (I) Keller 2004 <sup>36</sup> (I) Krebs 2009 <sup>b20</sup> (I) Stroud 2004 <sup>43</sup> (I) Tan 2004 <sup>a44</sup> (I)	-	Chansirinukor 2005 <sup>52</sup> (I)

Pain Measure	Number of studies	Total Participants	MID	Responsiveness	Concurrent validity	Discriminant validity	Test-retest reliability
<b>SF-36 Bodily Pain Scale (SF-36 BPS)</b>	10	2,174	Krebs 2010 <sup>b37</sup> (S,I)	Kean 2016 <sup>a35</sup> (S,I) Keller 2004 <sup>36</sup> (S,I) Krebs 2010 <sup>b37</sup> (S,I) Stewart 2007 <sup>56</sup> (S,I) Wittink 2004 <sup>47</sup> (S,I)	Driban 2015 <sup>31</sup> (S,I) Hicks 2009 <sup>33</sup> (S,I) Kamper 2015 <sup>54</sup> (S) Keller 2004 <sup>36</sup> (S,I) Krebs 2009 <sup>b20</sup> (S,I) Nilsson 2003 <sup>64</sup> (S,I) Wittink 2004 <sup>47</sup> (S,I)		
<b>Visual Analogue Scale (VAS) for Pain</b>	8	541	Parker 2012 <sup>41</sup> (S) Van Grootel 2007 <sup>67</sup> (S)	Gentelle-Bonnassies 2000 <sup>60</sup> (S) Sindhu 2011 <sup>42</sup> (S)	Changulani 2009 <sup>57</sup> (NR) Gronblad 1993 <sup>61</sup> (S) Sindhu 2011 <sup>42</sup> (S)	-	Gallasch 2007 <sup>51</sup> (S) Lund 2005 <sup>62</sup> (S) Sindhu 2011 <sup>42</sup> (S)
<b>Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)</b>	5	535	-	Gentelle-Bonnassies 2000 <sup>60</sup> (S,I) Nilsson 2003 <sup>64</sup> (S,I) Trudeau 2015 <sup>46</sup> (S)	Driban 2015 <sup>31</sup> (S,I) Pinsker 2015 <sup>50</sup> (S,I)	-	Pinsker 2015 <sup>50</sup> (S,I)
<b>Wong Faces Scale/ Wong-Baker Face Scale</b>	1	32	-	-	-	-	Gallasch 2007 <sup>51</sup> (S)

<sup>a</sup>Enrolled exclusively US Veterans

<sup>b</sup>Enrolled US Veterans and non-Veterans; results not stratified by Veteran status

MID=minimally important difference



## Minimally Important Difference

We identified 8 studies that estimated MIDs of 8 separate pain measures: BPI, GCPS, NRS, ODI, PEG, RMSQ, SF-36 BPS, and VAS (Table 3, Supplemental Content, Tables 5 and 6).<sup>33,37,41,52,58,63,66,67</sup> Six of the 8 measures assess pain intensity and interference/function (BPI, GCPS, PEG, SF-36 BPS, ODI, VAS), one (RMDQ) interference/function, and one (NRS) focused on intensity. Several methods for estimating MIDs were reported, including both distribution-based and anchor-based approaches. Distribution-based methods involve estimation of MID based on the distribution of the observed scores. Anchor-based methods use an external indicator (*eg*, patient rating of change) to put patients into positive change, no change, and negative change groups.<sup>70</sup> For each pain measure, MID estimates differed considerably depending on the estimation method used, the type of pain being studied, and the interval between evaluations. We broadly describe this outcome as minimally important difference, but note where studies describe the outcome differently.

Three studies calculated MID values for more than one pain measure.<sup>37,41,63</sup> One US study (n=427) estimated minimal clinically important change (MCIC) for BPI, GCPS (labeled the Chronic Pain Grade [CPG] in this study), PEG, RMDQ, and SF-36 BPS over 12 months.<sup>37</sup> A distribution-based standard error of measurement (SEM) was used to estimate MCIC. The SEM was then used to categorize patients as better, the same, or worse for each measure. “Better” indicated that the score improved at least one SEM from baseline and “worse” indicated that the score worsened at least one SEM from baseline. Kappa statistics for agreement between one-SEM and an anchor of patient’s global rating classifications were fair. The measures with the best agreement were the BPI (Kappas = 0.29 and 0.34 for trial and cohort data, respectively), the GCPS intensity (Kappas = 0.35 and 0.27), and the PEG (Kappas = 0.33 and 0.23).

Another retrospective cohort study from the US estimated minimum clinically important differences (MCID) based on 4 anchor-based approaches for 47 participants undergoing surgical treatment for pseudoarthrosis-related back pain.<sup>41</sup> MCIDs were calculated for the ODI and VAS 2 years postoperatively. The anchors were 1) patient rated global assessment with choices of “worse,” “unchanged,” “slightly better,” or “markedly better” and 2) patient rating of satisfaction with the results of their surgery (yes indicating responders, no indicating non-responders). The 4 MCID approaches included 1) average change (average change score seen in the group defined to be responders); 2) minimum detectable change (MDC) (equal to the upper value of the 95% confidence interval for average change score seen in the cohort defined to be non-responders); 3) change difference (difference of the average change score for responders and non-responders); and 4) ROC approach (the change value that provides the greatest sensitivity and/or specificity for a positive response). For the ODI, the calculated MCIDs differed by the approach used and ranged from 2.0 points for MDC up to 8.3 points for change difference. Fewer differences were seen for the VAS, where MCIDs ranged from 2.0 to 3.2 points.

One small study from the UK (n=48) estimated MCID for the ODI, NRS, and RMDQ after a 5-week class of exercise and education among patients with low back pain.<sup>63</sup> The PGIC was used to categorize patients into groups of “improved,” “unchanged,” and “deteriorated.” An anchor-based ROC approach estimated the MCID was 4 points for the NRS and RMDQ and 8 points for the ODI.

## Test-retest Reliability

Test-retest reliability, the extent to which a measure achieves the same result on 2 or more occasions when the condition is stable, was reported in 10 studies (Table 3, Supplemental Content, Table 5).<sup>17,30,33,42,48,50-52,61,62</sup> Several studies reported test-retest reliability for multiple pain measures. However, measure and timeframe comparisons differed across studies, making comparative evaluation of test-retest reliability difficult.

Test-retest reliabilities, assessed with Pearson correlations or intraclass correlations, were 0.90 or higher in many studies.<sup>33,42,50,51</sup> Pain measures evaluated in these studies included the Faces Scale, VAS, NRS, ODI, and WOMAC. There were few reports of test-retest reliabilities less than 0.80. One study evaluated test-retest reliability of the RMDQ (ICC=0.68) with approximately 3 months between assessments in patients who reported no change in work status.<sup>52</sup> Another study reported test-retest using the PROMIS-PI at baseline and 3 months apart (ICC=0.58).<sup>30</sup>

## Inter-rater Reliability

None of the included studies reported inter-rater reliability (*ie*, agreement between raters).

## Internal Consistency

The extent to which items in a measure are correlated and thus can be said to be measuring the same construct (*ie*, internal consistency) was reported in 8 studies.<sup>17,20,36,39,40,50,53,59</sup> In 7 studies, Cronbach's alpha was calculated; one calculated Spearman correlation coefficients.<sup>53</sup>

In studies reporting Cronbach's alpha, results were generally greater than 0.70, indicating good to excellent internal consistency. Pain measures evaluated include the BPI, GCPS, MPQ, ODI, PEG, PROMIS-PI, RMDQ, SF-36 BPS, WHYMPI, and WOMAC. In the study reporting Spearman correlation coefficients between elements of the BPI, values ranged from 0.38 to 0.84.<sup>53</sup>

## Concurrent and/or Criterion Validity

Concurrent validity is a measure of the extent to which scores on one measure relate to another measuring the same or a similar construct, while criterion validity measures a measure's correspondence to a gold standard or another measure. Nineteen studies reported concurrent/criterion validity.<sup>17,20,29,31,33,36,39,42-44,47,49,50,54,57-59,61,64</sup> Pain measures assessed for concurrent/criterion validity include the BPI, CGPS, HOOS, MPQ, NRS, ODI, PEG, PROMIS-PI, RMDQ, SF-36 BPS, VAS, WHYMPI/MPI, and WOMAC. Table 3 provides an overview of studies reporting this outcome; more details are presented in Supplemental Content, Table 5.

Reported correlations indicate fair to excellent concurrent and criterion validity across pain measures. Four studies provided results from multiple comparisons.<sup>20,31,36,47</sup> Krebs et al reported correlations between the PEG and other measures ranging from 0.60 (RMDQ) to 0.89 (BPI Interference component) with similar values for correlations of the BPI Severity and BPI Interference components with other measures.<sup>20</sup> Wittink et al computed R<sup>2</sup> values; values above 0.4 were considered high overlap between measures.<sup>47</sup> Observed R<sup>2</sup> values ranged from 0.37 to 0.58 among the MPI pain severity and interference components, the ODI, and the SF-36 BPS. Correlations between the PROMIS-PI and the SF-36 BPS (-0.73) and the WOMAC pain

subscale (0.47) were reported in one study.<sup>31</sup> Keller et al reported correlations between the BPI, SF-36 BPS, RMDQ, and GCPS, with values ranging from 0.47 to 0.81.<sup>36</sup>

One study reported intercorrelations (Kendall's tau) between components of the ODI and behavioral assessments of the components.<sup>59</sup> The correlation of the ODI Lifting Subscale with observed lifting was -0.38. The correlation of the ODI Walking Subscale with observed walking was -0.54. The correlation of the ODI Sitting Subscale with observed sitting was -0.40.

Two studies assessed correlations between different versions of the RMDQ.<sup>29,43</sup> In one study, Computer Adaptive Test versions with 5, 7, 9, and 11 items were evaluated with respect to a 23-item version of the RMDQ. The correlations were 0.93, 0.95, 0.97, and 0.98 for the 5-, 7-, 9-, and 11-item versions, respectively.<sup>29</sup> In the other study intercorrelations were reported for the 24-, 18-, and 11-item versions, with all values greater than 0.95.<sup>43</sup>

### Discriminant Validity

Discriminant validity is the ability of a measure to discriminate between groups. Four studies reported discriminant validity (Table 3, Supplemental Content, Table 5).<sup>30,33,38,39</sup> One study evaluated the ability of the MPQ Short Form to discriminate between number of pain diagnoses and between none/mild, moderate, and severe pain as determined with the MPI Pain Severity component.<sup>39</sup> No significant difference in total MPQ Short Form score was observed between study participants with one or with 2-3 pain diagnoses. However, scores were significantly higher in the group with 4 or more diagnoses. MPQ Short Form scores were significantly different across the 3 pain severity levels.

Krebs et al evaluated the accuracy of the NRS for predicting level of pain that interferes with function (defined in the study as BPI of 5 or higher) and level of pain that motivates a physician visit.<sup>38</sup> For both outcomes, the area under the ROC curve was 0.75-0.78 (indicating "fair" accuracy) and NRS scores of 4 and above increase the probability of interference with function or a physician visit as indicated by likelihood ratios substantially greater than 1.0.

A third study reported that ODI scores differed significantly ( $P < .001$ ) between groups with and without 1) high pain severity and high functional limitations and 2) chronic pain and high functional limitation.<sup>33</sup> Another study reported that PROMIS-PI scores differed significantly ( $P < .001$ ) between those seeking worker's compensation or not and those who had a fall in the past 3 months or not.<sup>30</sup>

### Responsiveness

We identified 22 studies that reported responsiveness, the ability of a measure to detect change in an outcome over time, in 14 of the 17 pain measures of interest. Details of study and population characteristics are provided in Table 3, Supplemental Content, Tables 5 and 7.<sup>20,27,28,30,32,35-37,42,44,46-48,52,53,55-57,60,63-65</sup>

Two common approaches to estimating responsiveness are external and internal. Internal responsiveness reflects the ability of a measure to change over a pre-specified time interval. External responsiveness relies on an anchor or external standard which is considered independent of the pain measure (eg, patient global rating of change) to assess the agreement between change in the measure and change in the external standard. Responsiveness was calculated by a variety

of metrics across studies including standardized response means (SRM) and standardized effect sizes (SES). The SRM is an effect size measure of within-group change and is calculated by taking the change of scores from time 1 to time 2 divided by the standard deviation of the change score. The studies also reported standardized effect sizes (SES), an effect size measure of between-group change which is calculated by taking the change-score means of 2 independent groups divided by the pooled the standard deviation of change. Magnitude of effect for SRM and SES are interpreted by the guidelines suggested by Cohen (0.2 is considered a small and 0.8 or greater is a large).<sup>71</sup> Area under the curve (AUC) values estimated from ROC analyses were used by several studies to also assess probability of correctly measuring discrimination between patients who improved and those who did not. A value of 0.5 can be interpreted to be the same as chance and a value of 1.0 indicates perfect discrimination. Thirteen studies estimated external (“anchored”) responsiveness.<sup>20,28,30,32,35-37,52,53,55-57,63</sup>

### *Comparative Studies*

Six studies compared *external* responsiveness across multiple pain measures (Supplemental Content, Table 7).<sup>35-38,56,63</sup> Studies that determined responsiveness based on AUC values are summarized on Table 4. The remaining 2 studies calculated SRMs for responsiveness for the BPI,<sup>20,36</sup> PEG,<sup>20</sup> GCPS,<sup>36</sup> and SF-36 BPS.<sup>36</sup>

Seven studies reported *internal* responsiveness for multiple pain measures (Supplemental Content, Tables 5 and 7).<sup>42,46-48,60,64</sup> Measures evaluated included the HOOS,<sup>64</sup> MPI,<sup>47</sup> MPQ,<sup>48</sup> NRS,<sup>46</sup> ODI,<sup>47,48</sup> SF-36 BPS,<sup>47</sup> VAS,<sup>60</sup> and WOMAC.<sup>46,60,64</sup>

### *Measure-specific*

Eleven studies reported responsiveness for one pain measure only (Supplemental Content, Tables 5 and 7).<sup>27,28,30,32,44,52,53,55,57,60,65</sup> Responsiveness varied within the individual measures, the populations, time intervals, and methods used to calculate. Pain measures included the BPI,<sup>44,53</sup> NRS,<sup>32</sup> ODI,<sup>27,57</sup> PGIC,<sup>65</sup> PROMIS-PI,<sup>28,30</sup> and RMDQ.<sup>52,55</sup>

**Table 4. Comparative External Responsiveness based on (AUC) Values for Detecting Any Improvement**

Study (sample size)	Pain Measures							
	BPI (total)	PEG	SF-36 BPS	PROMIS	RMDQ	CPG	NRS	ODI
Kean 2016 <sup>35</sup> (n=244)	0.73	0.71	0.68	Range 0.56 to 0.61 <sup>a</sup>		-		
Krebs 2010 <sup>37</sup> RCT (n=205)	0.81	0.78	0.72	-	0.81	Range 0.75 to 0.78 <sup>b</sup>		
Krebs 2010 <sup>37</sup> Cohort (n=222)	0.78	0.73	0.68	-	0.70	Range 0.65 to 0.75 <sup>b</sup>		
Maughan 2010 <sup>63</sup> (n=48)					0.64		0.50	0.67
Stewart 2007 <sup>56</sup> (n=134)			0.73				Range 0.68 to 0.70 <sup>c</sup>	

AUC=Area Under the Curve (Values range 0.5 (the same as chance) to 1.0 (perfect discrimination) and are interpreted as the probability of a measure correctly discriminating between participants who have improved and those who have not improved). BPI= Brief Pain Inventory; CPG= Chronic Pain Grade Questionnaire; ODI=Oswestry Disability Index; PROMIS=Patient-Reported Outcomes Measurement Information System; RCT=randomized controlled trial; SF-36 BPS=Short Form (36) Bodily Pain Scale

<sup>a</sup> Three versions were administered: 1) PROMIS-29 Profile Pain Interference Short Form (4-item). AUC=0.56; 2) PROMIS-57 Profile Pain Interference Short Form (8-item). AUC=0.57; 3) PROMIS Pain Interference Short Form 6b (6-item). AUC=0.61.

<sup>b</sup> Pain intensity and disability

<sup>c</sup> Two measures administered, pain intensity and pain bothersomeness

## Feasibility

### *Number of Items*

Among the 17 pain measures reviewed, the number of items used to assess pain ranged from 1 (NRS, PGIC, VAS, Wong-Baker Faces Scale) to 78 (MPQ). The 4 single-item measures assessed different dimensions of pain including pain intensity (Faces), pain intensity and/or interference (NRS and VAS), and changes in pain (PGIC). The phrasing of questions used to elicit pain scores was not consistent across the studies included in this review, and therefore in some cases it was not clear if multiple dimensions of pain were being assessed. These single-item measures also varied in how they were administered, as both the VAS and Faces involve visual cues.

Other low-item measures include the SF-36 BPS (2 items), the PEG (3 items), the DVPRS (5 items), and the GCPS (7 items). While still brief, these measures have the advantage of measuring both pain intensity and pain interference related to function. Mid-item measures include the ODI (10 items), the BPI (17 items), the WOMAC (24 items), and the RMDQ (24 items). The ODI includes one item related to pain severity (need for analgesic medications) and the RMDQ specifically measures disability related to pain.

The pain measures with the most items are the HOOS (40 items), KOOS (42 items), PROMIS-PI (41 items), MPI (52 items), and MPQ (78 items). Though lengthy, the HOOS and KOOS are the only measures that directly assess pain of the hip and knee, respectively. The PROMIS-PI is specific to pain interference in physical functioning and has 4 short form versions that are commonly used. The MPI queries pain intensity and interference in multiple domains, including social and family functioning. The highest-item measure, the MPQ, presents the patient with a list of adjectives from which to select descriptors for their subjective pain experience rather than asking them to answer questions on a Likert-type scale. In determining which measure provides sufficient items for a given research study, the intended use of the measure and the research setting will largely determine the appropriate choice.

### *Mode of Administration*

Desired mode of administration may also inform the appropriate choice of pain measure for research. Many of the pain measures can be self-administered or administered by an interviewer. Measures such as the KOOS and HOOS have been administered through the mail, while computer-based surveys have also been developed for the WOMAC, SF-36 BPS, PROMIS-PI, ODI, NRS, MPI, and HOOS. Four measures have been assessed for telephone administration, including the WOMAC, SF-36 BPS, ODI, and NRS.

### *Availability*

Pain measures readily available without restrictions on use include the DVPRS, GCPS, HOOS, KOOS, NRS, PGIC, PEG, RMDQ, and VAS. The MPI and MPQ can be obtained freely and directly from the developer. Free use of the ODI is permitted for non-funded academic research and individual clinical practice. Additionally, the PROMIS-PI is freely available after registering with an assessment center and endorsing terms and conditions. Measures that require purchase or permission to use are the BPI, SF-36 BPS, Wong-Baker Faces Scale, and WOMAC.

## SUMMARY AND DISCUSSION

### KEY MESSAGES

- Among 17 multi-item pain measures assessed, the most complete evidence on psychometric properties in chronic musculoskeletal pain populations was found for the ODI, RMDQ, and SF-36 BPS. Several key psychometric properties were available for the BPI, GCPS, MPI/WHYMPI, MPQ, PEG, PROMIS-PI, and WOMAC. Most of these measures include both pain severity/intensity and functional impairment.
- Of the measures focused primarily on pain severity, we found the greatest reporting of psychometric properties for the NRS and VAS, followed by the MPQ.
- Of the measures of pain interference, we found the greatest reporting of psychometric properties for the ODI, PROMIS-PI, and RMDQ.
- MID assessment methods differed and were often based on statistical rather than patient-noticeable differences.
- Reliability, internal consistency, concurrent or criterion validity, discriminate validity, and responsiveness differed widely but generally were in the fair to excellent range.
- Feasibility, measured by number of items, delivery mode, and public availability differed widely. The choice of measure may depend on population/condition of interest, research questions and settings, and resources available.
- Our review supplements earlier IMMPACT guidance on core outcome measures by providing recent findings on psychometric properties of measures specifically targeted for chronic musculoskeletal pain, using English language versions of measures, and including recently developed measures of pain severity and/or pain interference.
- Primary psychometric research on key measures in chronic musculoskeletal pain populations was limited overall. Future research should use consistent chronic musculoskeletal pain definitions, standardized psychometric outcomes assessment, and comprehensive descriptions of patient populations.
- Findings from this review can inform recommendations on specific core outcome measures for clinical research on chronic musculoskeletal pain interventions. Researchers' final choice of measures should consider population characteristics, pain site and type, recall period of interest and intervention length, analytic goals, and study resources.

### DISCUSSION

This rapid evidence review identified published research on psychometric properties of English-language versions of 17 key patient-reported pain outcome measures assessed in chronic musculoskeletal pain populations. The ODI, RMDQ, and SF-36 BPS were the most frequently studied multi-item pain measures and had reported data for all 4 main psychometric outcomes of interest: MID, responsiveness, validity, and test-retest reliability. Each of these measures

assesses interference; the SF-36 BPS, and ODI include a question about pain severity but no study reported separate outcomes for severity and interference. The BPI, GCPS, and PEG had data on MID, responsiveness, and validity. Each of these measures assess both pain severity and interference with all but one study reporting separate results for the 2 subscales of the BPI and GCPS; severity and interference are combined in the PEG. MPI/WHYMPI, MPQ, PROMIS-PI; and WOMAC had data on responsiveness, validity, and test-retest reliability. The MPQ is a measure of pain severity and the PROMIS-PI is a measure of pain interference. The MPI/WHYMPI and WOMAC include severity and interference subscales. All but one study reported separate results for those subscales.

Findings from our review supplement the work of IMMPACT<sup>4</sup> and IMMPACT/OMERACT.<sup>6</sup> The 2005 IMMPACT guidance on core outcome measures for chronic pain clinical trials was based on studies of any chronic pain, including cancer, dental, and neuropathic pain. The literature reviews to support the guidance included studies published through early 2003.<sup>4</sup> The 2016 IMMPACT/OMERACT guidance on assessment of physical function and participation in chronic pain clinical trials identified patient-reported outcome measures of physical functioning, including 8 addressed in our review, but did not perform detailed assessments of the measures and did not make recommendations for use of specific measures.<sup>6</sup>

While IMMPACT provides recommendations for measures that can be used to assess pain severity and/or pain interference across a broad range of pain types,<sup>4</sup> there have since been many new studies in the area of chronic musculoskeletal pain. Of 43 studies included in our review, 38 were published from 2003 to January 2017. In addition, new pain measures have been developed, notably the DVPRS, PEG, and PROMIS-PI. Therefore, our report provides updated information and a broader look at psychometric properties of measures for assessment of both pain severity and pain interference for chronic musculoskeletal pain.

Further, our findings are consistent with pain outcome measurement reviews focused on specific pain-related diagnoses or pain measures. Three reviews focused on patient-reported health outcome measures for LBP found the ODI and RMDQ to be the most comprehensively studied both for responsiveness<sup>72</sup> and for other psychometric properties.<sup>73,74</sup> There were few data on psychometric properties of pain severity measures (*ie*, NRS, VAS, BPI, MPQ) commonly used in RCTs of interventions for LBP.<sup>73</sup> Another review of back-specific functional status questionnaires for LBP found the ODI and RMDQ to have been most frequently studied, with good measurement properties in their original forms as retested in multiple settings.<sup>75</sup> A review of studies that had evaluation of psychometric properties as a main purpose found 2 of our measures of interest, the HOOS and WOMAC, to be adequately assessed for use in patients with hip and groin disability.<sup>76</sup> A review of 6 studies that used the KOOS to evaluate patients undergoing total knee arthroplasty found acceptable psychometric properties.<sup>77</sup> None of the studies included in that review were eligible for our review due to language of publication, use of a non-English language version of the KOOS, or inadequate definition of pain duration. A review of 76 studies assessing the measurement properties of the WOMAC, predominantly in patients with hip and/or knee osteoarthritis, found acceptable reliability.<sup>78</sup> Few studies assessed responsiveness and MID was not an outcome of interest for that review.

For purposes of measure selection, psychometric properties must be considered alongside conceptual and practical concerns.<sup>79</sup> The ODI and RMDQ were developed for and most often tested in low back pain, and the WOMAC was developed for and most often tested in knee and

hip pain. The BPI, GCPS, MPI/WHYMPI, MPQ, PEG, PROMIS-PI, and SF-36 BPS were designed to assess more broadly defined pain, and were tested in populations with varying chronic pain-related diagnoses. Most of these measures assess severity and functional impairment; exceptions are the MPQ (severity only) and the RMDQ and PROMIS-PI (interference only). Researchers' choice of measures should include their research goals, such as pain site, pain type, recall period of interest and length of intervention (with respect to measure responsiveness data), analytic goals (with respect to measure range and scale), and study resources (with respect to measure feasibility, including available time and mode of administration).

Versions of the NRS and the VAS were also frequently studied with respect to the 4 key psychometric outcomes of interest. However, NRS and VAS are single-item response measures, and the associated questions to which study participants responded varied with respect to phrasing, recall periods, and score ranges. For the NRS and VAS, our evidence review was thus less a review of psychometric research on 2 clearly defined pain measures and more a cataloging of multiple single-item numeric rating-based or visual analog scale-based approaches to assessing primarily pain severity.

## CHALLENGES IN ASSESSMENT OF PSYCHOMETRIC PROPERTIES

### Minimally Important Different (MID)

The range of assessment methods reflects variation in current MID-related research (Supplemental Content, Table 6). Assessments of minimum clinically important difference (MCID) for a patient-reported outcome measure should ideally involve anchoring the measure to an indicator of meaningful patient-reported change in a clinical outcome.<sup>70,80,81</sup> While some MID estimates reported here constitute MCIDs anchored to patient-reported clinical improvement via adaptations of the Patient Global Impression of Change (PGIC),<sup>37,41,58,63,66</sup> others are purely estimates of statistical minimum detectable change (MDC) based on study population distribution characteristics<sup>33,52,67</sup> without reference to clinical import of that change. Comparing anchor-based MCID findings with distribution-based MDC findings can be useful in MID estimation, as this allows researchers to consider both an external benchmark of clinical change and a measure of change detectable despite variation.<sup>58,70,80</sup> Reviewed studies, however, contained relatively few estimates via any method, precluding comparison and generalization of measure-specific MIDs. MIDs for patient-reported measures are likely to vary based on the constructs assessed by each measure, as well as by patient population, study design, and baseline measure value. It is possible that widespread application of a 30% change from baseline as an MID, originally assessed using an NRS for pain severity<sup>19</sup> and ultimately recommended for a range of patient-reported pain outcome measures,<sup>82</sup> has discouraged measure-specific MID development. Further research could explore whether the broadly adopted figure of a 30% change from baseline is empirically generalizable across patient-reported outcome measures in chronic musculoskeletal pain studies and populations. Consensus is needed on optimal approaches to developing and reporting MID for patient-reported measures in chronic musculoskeletal pain.

### Validity

There is no gold standard comparator for assessment of pain measure validity in the domains assessed. Most included studies' methods of assessing concurrent/criterion validity involved

finding correlations between a measure of interest and another measure or subscale of interest. Perhaps unsurprisingly, therefore, our review identified a self-referential network of patient-reported outcome measures validated against one another. Other assessments arguably relevant to construct validity, such as relationships of self-reported pain-related functioning measures to objective physical performance measures, were less commonly identified, consistent with the state of current physical function research in pain.<sup>6</sup> Estimates of measure validity are difficult to compare within or across measures in this review. Future research could further investigate the network of validity comparisons between measures of interest, to clarify underlying assumptions that support the validity of these measures and to identify gaps requiring conceptual research.

## Responsiveness

Responsiveness findings in reviewed studies are challenging to compare both within and across measures (Supplemental Content, Table 7). Some methods of comparing pain measure changes within clinical trials of pain interventions cannot separate the effectiveness of a pain intervention from the responsiveness of the pain measure used to assess it. Few methods recognize the inherent challenge that short-term fluctuations in pain, which commonly occur in chronic musculoskeletal pain conditions, pose to the capacity of pre-post assessments to track pain trajectory over time. Further, included pain measures have a wide range of recall periods (from 24 hours for the RMDQ to 4 weeks for the SF-36 BPS), and reviewed studies have a range of time periods between assessment points. Clinical researchers interested in comparing measures' responsiveness should consider available psychometric evidence in the context of their own work, including the recall period of interest, the expected amount and timeframe of change in the pain domains they plan to assess, and their desired study design (*eg*, pre-post assessment vs longitudinal repeated-measures assessment).

## Test-retest Reliability

Interpreting test-retest reliability estimates has conceptual challenges similar to those of responsiveness: it can be difficult to separate undesirable variability in a measure from variability that reflects actual fluctuations in subjective pain constructs, and can thus be difficult to determine the optimal test-retest reliability interval for a given measure. A short-term fluctuation in a measure may not indicate a lack of test/retest reliability, and may in fact be evidence of responsiveness to true changes in pain course. As with responsiveness, we recommend that researchers interested in specific measures' reliability consider reliability-related timeframes and design features in the context of their own work.

## LIMITATIONS AND FUTURE RESEARCH

### Limitations of a Rapid Evidence Review

Rapid evidence review development requires streamlining the scope of literature search and eligibility criteria, and language and date restrictions are among current best practice recommendations.<sup>83-85</sup> Our review was limited to studies that assessed measures or published results in English. However, this decision was also influenced and supported by evidence on the limited generalizability of self-report measures' psychometric properties derived in languages other than that of the intended population,<sup>86,87</sup> and highlights the need for linguistic and cultural validation of pain measures. With respect to search strategy, our primary abstract search was limited to dates from 2000 onward. We complemented this, however, by hand-searching

reference lists of included studies and relevant reviews, searching websites of each specific pain measure, and by querying experts for supplementary suggestions. We included identified eligible articles regardless of date, though we acknowledge that we may have missed a relevant publication. Our criteria may have excluded some studies of psychometric properties of measures developed and validated prior to the popularization of specifying chronicity and duration of pain. Researchers considering such pain measures will need to consider the relevance of past psychometric work in the context of current conceptual pain research, and of their planned studies' objectives and target populations. We excluded studies that enrolled patients with chronic musculoskeletal conditions commonly associated with pain but did not specify that enrolled patients had chronic pain (*eg*, radiologically defined osteoarthritis). In addition to a decision based on scope we believe this is justifiable scientifically as it is not clear if individuals in these studies had chronic pain, and some of these studies specifically noted that patients either did not have pain or had acute or subacute pain. We also excluded trials of interventions for pain that did not note assessment of psychometric properties in the abstract. Our focus was on primary psychometric research on the pain measures of interest, and accordingly our search required psychometric properties to be mentioned in the abstract. It is possible that this search approach did not identify some psychometric assessments embedded in studies that used the measures of interest as primary clinical outcomes. However, we believe it is unlikely that this decision excluded a large body of relevant information and took steps to address this concern within the scope of a rapid review. For example, our search of included studies from other similar evidence reviews and query of specific measures websites failed to identify trials that did not describe psychometric properties in the abstract.

### **Chronic Musculoskeletal Pain Definition and Reporting**

Chronic musculoskeletal pain definition and reporting differed widely across reviewed studies. The required duration for pain to be considered “chronic” was inconsistent, and was not always reported. Pain type (*eg*, musculoskeletal), primary diagnostic cause (*eg*, osteoarthritis), and primary bodily site(s) (*eg*, low back) were inconsistently reported. In some studies, pain-related diagnoses or bodily pain sites were reported without reference to the existence of pain duration or chronicity (*eg*, radiologically defined osteoarthritis); these studies did not meet inclusion criteria for this review. We also found inconsistent reporting of pain-relevant participant characteristics such as pain duration at baseline, baseline level of relevant pain domains, current use of pharmacological and/or non-pharmacological treatments, and co-existing physical or mental health conditions. Such differences in chronic musculoskeletal pain definition and reporting reflect active discussion in current pain research: when and how duration affects key pain qualities, when and how causal diagnoses and bodily site affect key pain qualities, and when and how intermittent pain differs meaningfully from chronic continuous pain.<sup>10,88</sup> These conceptual uncertainties underlie the wide range of approaches to defining target populations for pain studies. Research is needed to define target populations for psychometric research on measures for use in chronic musculoskeletal pain, as well as standards for reporting of pain duration, relevant diagnoses, and bodily sites. Additional work is needed to define target populations for psychometric research on measures for use in chronic musculoskeletal pain, as well as standards for reporting of chronic musculoskeletal pain duration, relevant diagnoses, and pain sites.

## Study Populations

Most studies were conducted in populations with over 50% women and mean ages 40-59. Most studies did not report race or ethnicity; of those that did, all included more than 50% white participants, and most included more than 75% white participants. No studies reported outcomes stratified by sex or gender, age range, or race/ethnicity. Generalizability of psychometric findings is thus limited by both demographic underreporting and population homogeneity. Given substantial evidence of the influence of age and psychosocial factors on individuals' experiences and reporting of both pain-related functional impairment and pain severity,<sup>87,89-91</sup> there is a need for consensus on key study population demographic and clinical characteristics, more consistent reporting of these population characteristics within studies, and further research on how measures' psychometric properties generalize or change across age ranges and psychosocial categories.

## APPLICABILITY TO VHA RESEARCH

Our findings are highly applicable to research on chronic musculoskeletal pain in the VA population. Four studies enrolled only Veterans<sup>17,35,39,44</sup> and 2 included Veterans.<sup>20,37</sup> These studies evaluated psychometric properties of several of the pain measures that overall had substantial evidence, including the BPI, MPI/WHYMPI, MPQ, PEG, PROMIS-PI, RMDQ, and SF-36 BPS.

The chronic musculoskeletal pain conditions are representative of conditions seen in a Veteran population, with measurement of back, knee, and hip pain most common. Mean ages of study participants ranged from 32 to 80 years. However, studies, other than those of Veterans, included a large percentage of women and studies reporting race/ethnicity, most from the US, enrolled a high percentage of white individuals. Additional methods work is needed in broader populations and for more consistent and complete demographic reporting.

## CONCLUSIONS

Among multi-item pain measures assessed, the most complete evidence on psychometric properties of interest within chronic musculoskeletal pain populations was found for the ODI, RMDQ and SF-36 BPS, while several additional measures (BPI, GCPS, MPI/WHYMPI, MPQ, PEG, PROMIS-PI, and WOMAC) also had evidence for several of the key psychometric properties. Most of these measures include both pain severity/intensity and functional impairment. In addition to evidence on psychometric properties, choice of pain outcome measures for a specific research study must consider both conceptual elements (*eg*, pain domains of interest, pain sites and diagnoses, time course, and population characteristics) and practical concerns (*eg*, burden to complete, mode of assessment, cost). Limitations of current chronic musculoskeletal pain measurement research relate to variations in (1) definition and reporting of chronic musculoskeletal pain and pain-related diagnoses, (2) methods of assessing psychometric outcomes, and (3) reporting on demographics of patient populations. Findings from this review can inform recommendations on specific core outcome measures for clinical research on chronic musculoskeletal pain interventions. Further methods research is needed to validate patient-reported pain outcome measures in populations with chronic musculoskeletal pain and develop a framework for determining outcome measurement selection that incorporates feasibility and applicability.

## REFERENCES

1. Butchart A, Kerr EA, Heisler M, Piette JD, Krein SL. Experience and management of chronic pain among patients with other complex chronic conditions. *Clin J Pain*. 2009;25(4):293-298.
2. Gereau RWt, Sluka KA, Maixner W, et al. A pain research agenda for the 21st century. *J Pain*. 2014;15(12):1203-1214.
3. Department of Health and Human Services. National pain strategy: a comprehensive population health-level strategy for pain. 2015. Available at: [https://iprcc.nih.gov/docs/HHSNational\\_Pain\\_Strategy.pdf](https://iprcc.nih.gov/docs/HHSNational_Pain_Strategy.pdf). Accessed 1 August 2017.
4. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2005;113(1-2):9-19.
5. Dworkin RH, Turk DC, McDermott MP, et al. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2009;146(3):238-244.
6. Taylor AM, Phillips K, Patel KV, et al. Assessment of physical function and participation in chronic pain clinical trials: IMMPACT/OMERACT recommendations. *Pain*. 2016;157(9):1836-1850.
7. Turk DC, Dworkin RH, Burke LB, et al. Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. *Pain*. 2006;125(3):208-215.
8. Turk DC, Dworkin RH, McDermott MP, et al. Analyzing multiple endpoints in clinical trials of pain treatments: IMMPACT recommendations. Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials. *Pain*. 2008;139(3):485-493.
9. Turk DC, Dworkin RH, Revicki D, et al. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. *Pain*. 2008;137(2):276-285.
10. Younger J, McCue R, Mackey S. Pain outcomes: a brief review of instruments and techniques. *Curr Pain Headache Rep*. 2009;13(1):39-43.
11. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. *Ann Acad Med Singapore*. 1994;23(2):129-138.
12. Buckenmaier CC, 3rd, Galloway KT, Polomano RC, McDuffie M, Kwon N, Gallagher RM. Preliminary validation of the Defense and Veterans Pain Rating Scale (DVPRS) in a military population. *Pain Med*. 2013;14(1):110-123.
13. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain*. 1992;50(2):133-149.
14. Klassbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score: An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. *Scand J Rheumatol*. 2003;32:46-51.

15. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes*. 2003;1:64.
16. McCaffery M, Beebe A. *Pain: Clinical Manual for Nursing Practice*. St. Louis, MO: Mosby, 1989.
17. Kerns RD, Turk DC, Rudy TE. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain*. 1985;23(4):345-356.
18. Smeets R, Koke A, Lin CW, Ferreira M, Demoulin C. Measures of function in low back pain/disorders: Low Back Pain Rating Scale (LBPRS), Oswestry Disability Index (ODI), Progressive Isoinertial Lifting Evaluation (PILE), Quebec Back Pain Disability Scale (QBPDS), and Roland-Morris Disability Questionnaire (RDQ). *Arthritis Care Res*. 2011;63 Suppl 11:S158-173.
19. Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*. 2001;94(2):149-158.
20. Krebs EE, Lorenz KA, Bair MJ, et al. Development and initial validation of the PEG, a three-item scale assessing pain intensity and interference. *J Gen Intern Med*. 2009;24(6):733-738.
21. Pain Interference: A brief guide to the PROMIS Pain Interference instruments. 2015. Available at: <https://www.assessmentcenter.net/documents/PROMIS%20Pain%20Interference%20Scoring%20Manual.pdf>. Accessed 1 August 2017.
22. Roland MO, Morris RW. A study of the natural history of back pain. Part 1: Development of a reliable and sensitive measure of disability in low back pain. *Spine*. 1983;8:141-144.
23. Ware JE, Jr., Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *J Clin Epidemiol*. 1998;51(11):903-912.
24. Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. *Res Nurs Health*. 1990;13(4):227-236.
25. American College of Rheumatology. Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Available at: <http://www.rheumatology.org/I-Am-A/Rheumatologist/Research/Clinician-Researchers/Western-Ontario-McMaster-Universities-Osteoarthritis-Index-WOMAC>. Accessed 1 August 2017.
26. Wong-Baker FACES<sup>®</sup> History. Available at: <http://wongbakerfaces.org/us/wong-baker-faces-history/>. Accessed 1 August 2017.
27. Anagnostis C, Gatchel RJ, Mayer TG. The pain disability questionnaire: a new psychometrically sound measure for chronic musculoskeletal disorders. *Spine*. 2004;29(20):2290-2302.
28. Askew RL, Cook KF, Revicki DA, Cella D, Amtmann D. Evidence from diverse clinical populations supported clinical validity of PROMIS pain interference and pain behavior. *J Clin Epidemiol*. 2016;73:103-111.

29. Cook KF, Choi SW, Crane PK, Deyo RA, Johnson KL, Amtmann D. Letting the CAT out of the bag: comparing computer adaptive tests and an 11-item short form of the Roland-Morris Disability Questionnaire. *Spine*. 2008;33(12):1378-1383.
30. Deyo RA, Katrina R, Buckley DI, et al. Performance of a Patient Reported Outcomes Measurement Information System (PROMIS) Short Form in older adults with chronic musculoskeletal pain. *Pain Med*. 2016;17(2):314-324.
31. Driban JB, Morgan N, Price LL, Cook KF, Wang C. Patient-Reported Outcomes Measurement Information System (PROMIS) instruments among individuals with symptomatic knee osteoarthritis: a cross-sectional study of floor/ceiling effects and construct validity. *BMC Musculoskelet Disord*. 2015;16:253.
32. Godil SS, Parker SL, Zuckerman SL, Mendenhall SK, McGirt MJ. Accurately measuring the quality and effectiveness of cervical spine surgery in registry efforts: determining the most valid and responsive instruments. *Spine J*. 2015;15(6):1203-1209.
33. Hicks GE, Manal TJ. Psychometric properties of commonly used low back disability questionnaires: are they useful for older adults with low back pain? *Pain Med*. 2009;10(1):85-94.
34. Jensen MP, Schnitzer TJ, Wang H, Smugar SS, Peloso PM, Gammaitoni A. Sensitivity of single-domain versus multiple-domain outcome measures to identify responders in chronic low-back pain: pooled analysis of 2 placebo-controlled trials of etoricoxib. *Clin J Pain*. 2012;28(1):1-7.
35. Kean J, Monahan PO, Kroenke K, et al. Comparative responsiveness of the PROMIS Pain Interference Short Forms, Brief Pain Inventory, PEG, and SF-36 Bodily Pain Subscale. *MedCare*. 2016;54(4):414-421.
36. Keller S, Bann CM, Dodd SL, Schein J, Mendoza TR, Cleeland CS. Validity of the Brief Pain Inventory for use in documenting the outcomes of patients with noncancer pain. *Clin J Pain*. 2004;20:309-318.
37. Krebs EE, Bair MJ, Damush TM, Tu W, Wu J, Kroenke K. Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. *Med Care*. 2010;48(11):1007-1014.
38. Krebs EE, Carey TS, Weinberger M. Accuracy of the pain numeric rating scale as a screening test in primary care. *J Gen Intern Med*. 2007;22(10):1453-1458.
39. Lovejoy TI, Turk DC, Morasco BJ. Evaluation of the psychometric properties of the revised short-form McGill Pain Questionnaire. *J Pain*. 2012;13(12):1250-1257.
40. Merriwether EN, Rakel BA, Zimmerman MB, et al. Reliability and construct validity of the Patient-Reported Outcomes Measurement Information System (PROMIS) Instruments in women with fibromyalgia. *Pain Med*. 2016.
41. Parker SL, Adogwa O, Mendenhall SK, et al. Determination of minimum clinically important difference (MCID) in pain, disability, and quality of life after revision fusion for symptomatic pseudoarthrosis. *Spine J*. 2012;12(12):1122-1128.
42. Sindhu BS, Shechtman O, Tuckey L. Validity, reliability, and responsiveness of a digital version of the visual analog scale. *J Hand Ther*. 2011;24(4):356-363; quiz 364.

43. Stroud MW, McKnight PE, Jensen MP. Assessment of self-reported physical activity in patients with chronic pain: development of an abbreviated Roland-Morris disability scale. *J Pain*. 2004;5(5):257-263.
44. Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the Brief Pain Inventory for chronic nonmalignant pain. *J Pain*. 2004;5(2):133-137.
45. Tong HC, Geisser ME, Ignaczak AP. Ability of early response to predict discharge outcomes with physical therapy for chronic low back pain. *Pain Pract*. 2006;6(3):166-170.
46. Trudeau J, Van Inwegen R, Eaton T, et al. Assessment of pain and activity using an electronic pain diary and actigraphy device in a randomized, placebo-controlled crossover trial of celecoxib in osteoarthritis of the knee. *Pain Pract*. 2015;15(3):247-255.
47. 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(3):133-142.
48. Burnham R, Stanford G, Gray L. An assessment of a short composite questionnaire designed for use in an interventional spine pain management setting. *PM R*. 2012;4(6):413-418; quiz 418.
49. Mikail SF, DuBreuil S, D'eon JL. A Comparative Analysis of Measures Used in the Assessment of Chronic Pain Patients. *Psychol Assess*. 1993;5(1):117-120.
50. Pinsker E, Inrig T, Daniels TR, Warmington K, Beaton DE. Reliability and validity of 6 measures of pain, function, and disability for ankle arthroplasty and arthrodesis. *Foot Ankle Int*. 2015;36(6):617-625.
51. Gallasch CH, Alexandre NM. The measurement of musculoskeletal pain intensity: a comparison of four methods. *Rev Gaucha Enfer*. 2007;28(2):260-265.
52. Chansirinukor W, Maher CG, Latimer J, Hush J. Comparison of the functional rating index and the 18-item Roland-Morris Disability Questionnaire: responsiveness and reliability. *Spine*. 2005;30(1):141-145.
53. Chien CW, Bagraith KS, Khan A, Deen M, Strong J. Comparative responsiveness of verbal and numerical rating scales to measure pain intensity in patients with chronic pain. *J Pain*. 2013;14(12):1653-1662.
54. Kamper SJ, Grootjans SJ, Michaleff ZA, Maher CG, McAuley JH, Sterling M. Measuring pain intensity in patients with neck pain: does it matter how you do it? *Pain Pract*. 2015;15(2):159-167.
55. Macedo LG, Maher CG, Latimer J, Hancock MJ, Machado LA, McAuley JH. Responsiveness of the 24-, 18- and 11-item versions of the Roland Morris Disability Questionnaire. *Eur Spine J*. 2011;20(3):458-463.
56. Stewart M, Maher CG, Refshauge KM, Bogduk N, Nicholas M. Responsiveness of pain and disability measures for chronic whiplash. *Spine*. 2007;32(5):580-585.
57. Changulani M, Shaju A. Evaluation of responsiveness of Oswestry low back pain disability index. *Arch Orthop Trauma Surg*. 2009;129(5):691-694.

58. de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Qual Life Res.* 2007;16(1):131-142.
59. Fisher K, Johnston M. Validation of the Oswestry Low Back Pain Disability Questionnaire, its sensitivity as a measure of change following treatment and its relationship with other aspects of the chronic pain experience. *Physiother Theory Pract.* 1997;13:67-80.
60. Gentelle-Bonnassies S, Le Claire P, Mezieres M, Ayral X, Dougados M. Comparison of the responsiveness of symptomatic outcome measures in knee osteoarthritis. *Arthritis Care Res.* 2000;13(5):280-285.
61. Gronblad M, Hupli M, Wennerstrand P, et al. Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. *Clin J Pain.* 1993;9:189-195.
62. Lund I, Lundeberg T, Sandberg L, Budh CN, Kowalski J, Svensson E. Lack of interchangeability between visual analogue and verbal rating pain scales: a cross sectional description of pain etiology groups. *BMC Med Res Methodol.* 2005;5:31.
63. Maughan EF, Lewis JS. Outcome measures in chronic low back pain. *Eur Spine J.* 2010;19:1484-1494.
64. Nilsson AK, Lohmander LS, Klassbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)--validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord.* 2003;4:10.
65. Scott W, McCracken LM. Patients' impression of change following treatment for chronic pain: global, specific, a single dimension, or many? *J Pain.* 2015;16(6):518-526.
66. van der Roer N, Ostelo RW, Bekkering GE, van Tulder MW, de Vet HC. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine.* 2006;31(5):578-582.
67. van Grootel RJ, van der Bilt A, van der Glas HW. Long-term reliable change of pain scores in individual myogenous TMD patients. *Eur J Pain.* 2007;11(6):635-643.
68. Polomano RC, Galloway KT, Kent ML, et al. Psychometric testing of the defense and veterans pain rating scale (DVPRS): a new pain scale for military population. *Pain Med.* 2016;17:1505-1519.
69. Ornetti P, Parratte S, Gossec L, et al. Cross-cultural adaptation and validation of the French version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in knee osteoarthritis patients. *Osteoarthritis Cartilage.* 2008;16:423-428.
70. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61(2):102-109.
71. Cohen J. *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

72. Cleland J, Gillani R, Bienen EJ, Sadosky A. Assessing dimensionality and responsiveness of outcomes measures for patients with low back pain. *Pain Pract.* 2011;11(1):57-69.
73. Chapman JR, Norvell DC, Hermsmeyer JT, et al. Evaluating common outcomes for measuring treatment success for chronic low back pain. *Spine.* 2011;36(21 Suppl):S54-68.
74. Rocchi MB, Sisti D, Benedetti P, Valentini M, Bellagamba S, Federici A. Critical comparison of nine different self-administered questionnaires for the evaluation of disability caused by low back pain. *Eura Medicophys.* 2005;41(4):275-281.
75. Grotle M, Brox J, Vollestad N. Functional status and disability questionnaires: what do they assess?: a systematic review of back-specific outcome questionnaires. *Spine.* 2005;30(1):130-140.
76. Thorborg K, Roos EM, Bartels EM, Petersen J, Holmich P. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. *Br J Sports Med.* 2010;44:1186-1196.
77. Peer MA, Lane J. The Knee Injury and Osteoarthritis Outcome Score (KOOS): a review of its psychometric properties in people undergoing total knee arthroplasty. *J Orthop Sports Phys Ther.* 2013;43(1):20-28.
78. Gandek B. Measurement properties of the Western Ontario and McMaster Universities Osteoarthritis Index: a systematic review. *Arthritis Care Res.* 2015;67(2):216-229.
79. Mokkink LB, Prinsen CA, Bouter LM, Vet HC, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther.* 2016;20(2):105-113.
80. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003;56(5):395-407.
81. Turner D, Schunemann HJ, Griffith LE, et al. The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol.* 2010;63(1):28-36.
82. Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain.* 2008;9(2):105-121.
83. Haby MM, Chapman E, Clark R, Barreto J, Reveiz L, Lavis JN. What are the best methodologies for rapid reviews of the research evidence for evidence-informed decision making in health policy and practice: a rapid review. *Health Res Policy Syst.* 2016;14(1):83.
84. Tricco AC, Antony J, Zarin W, et al. A scoping review of rapid review methods. *BMC Medicine.* 2015;13:224.
85. Tricco AC, Zarin W, Antony J, et al. An international survey and modified Delphi approach revealed numerous rapid review methods. *J Clin Epidemiol.* 2016;70:61-67.
86. Beaton D, BscOt M, Bombardier C, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine.* 2000;25(24):3186-3191.

87. Booker SS, Herr K. The state-of-“cultural validity” of self-report pain assessment tools in diverse older adults. *Pain Med.* 2014;16(2):232-239.
88. Von Korff M. *Assessment of chronic pain in epidemiological and health services research.* New York: Guilford Publications; 2011.
89. Fillingim RB, King CD, Ribeiro-Dasilva MC, Rahim-Williams B, Riley JL, 3rd. Sex, gender, and pain: a review of recent clinical and experimental findings. *J Pain.* 2009;10(5):447-485.
90. Kroenke K, Spitzer RL. Gender differences in the reporting of physical and somatoform symptoms. *Psychosom Med.* 1998;60(2):150-155.
91. Tait RC, Chibnall JT. Racial/ethnic disparities in the assessment and treatment of pain: psychosocial perspectives. *Am Psychol.* 2014;69(2):131-141.