

## PRIMARY CARE & HEALTH SERVICES SECTION

# Core Outcome Measures for Chronic Musculoskeletal Pain Research: Recommendations from a Veterans Health Administration Work Group

Kurt Kroenke, MD,\* Erin E. Krebs, MD,<sup>†</sup> Dennis Turk, PhD,<sup>‡</sup> Michael Von Korff, ScD,<sup>§</sup> Matthew J. Bair, MD,\* Kelli D. Allen, PhD,<sup>¶</sup> Friedhelm Sandbrink, MD,<sup>||</sup> Andrea L. Cheville, MD,<sup>|||</sup> Lynn DeBar, PhD,<sup>§</sup> Karl A. Lorenz, MD,\*\* and Robert D. Kerns, PhD<sup>††</sup>

\*Indiana University and Roudebush VAMC, Indianapolis, Indiana; <sup>†</sup>University of Minnesota and Minneapolis VAMC, Minneapolis, Minnesota; <sup>‡</sup>University of Washington, Seattle, Washington; <sup>§</sup>Kaiser Permanente Washington Health Research Institute, Seattle, Washington; <sup>¶</sup>University of North Carolina and Durham VAMC, Chapel Hill, North Carolina; <sup>||</sup>George Washington University and Washington DC VAMC, Washington, District of Columbia; <sup>|||</sup>Mayo Clinic, Rochester, Minnesota; \*\*Stanford University and Palo Alto VAMC, Palo Alto, California; <sup>††</sup>Yale University and West Haven VAMC, West Haven, Connecticut, USA

*Correspondence to:* Kurt Kroenke, MD, Regenstrief Institute, Room 221, 1101 W. 10th Street, Indianapolis, IN 46202, USA. Tel: 317-274-9046; Fax: 317-274-9304; E-mail: kkroenke@regenstrief.org.

Funding sources: This report was sponsored by the Department of Veterans Affairs (VA) Health Services Research and Development Service.

Disclaimer: The sponsor had no role in work group deliberations; data collection, analysis, or interpretation; the writing of the report; or the decision to submit the article for publication. The views expressed in this article are those of the authors and do not represent the views of the VA or the US Government.

Conflicts of interest: The authors have no conflicts of interest to disclose.

### Abstract

**Objective.** Chronic musculoskeletal pain (CMSP) disorders are among the most prevalent and disabling conditions worldwide. It would be advantageous to have common outcome measures when comparing results across different CMSP research studies. **Methods.** The Veterans Health Administration appointed a work group to recommend core outcome measures for assessing pain intensity and interference as well as important secondary domains in clinical research. The work group used three streams of data to inform their recommendations: 1) literature synthesis augmented by three recently completed trials; 2) review and comparison of measures recommended by other expert groups; 3) two Delphi surveys of work group members. **Results.** The single-item numerical rating scale and seven-item Brief Pain Inventory interference scale emerged as the recommended measures for assessing pain intensity and interference, respectively. The secondary domains ranked most important included physical functioning and depression, followed by sleep, anxiety, and patient-reported global impression of change (PGIC). For these domains, the work group recommended the Patient-Reported Outcome Information System four-item physical function and sleep scales, the Patient Health Questionnaire two-item depression scale, the Generalized Anxiety Disorder two-item anxiety scale, and the single-item PGIC. Finally, a single-item National Health Interview Survey item was favored for defining chronic pain. **Conclusions.** Two scales comprising eight items are recommended as core outcome measures for pain intensity and interference in all studies of chronic musculoskeletal pain, and brief scales comprising 13 additional items can be added when possible to assess important secondary domains.

**Key Words:** Pain; Low Back Pain; Psychometrics; Measures; Minimal Data Set

Musculoskeletal pain conditions account for 70–80% of all chronic pain [1–3]. Low back pain (LBP), neck pain, osteoarthritis (OA), and other musculoskeletal disorders represent four of the top nine causes of disability in the

United States and result in more years lived with disability than the 12 leading medical causes of disability combined [4]. Chronic pain costs the United States an estimated \$560–635 billion annually, and although

analgesic medications are one of the most commonly prescribed classes of drugs [5], they have, on average, relatively modest effects on reducing pain and improving physical functioning. Moreover, increasing awareness of the risks of opioid analgesics has made this particular class of analgesics even less desirable as a treatment modality for the treatment of individuals with chronic pain [6]. Thus, the Veterans Health Administration (VHA) Health Services Research and Development Service convened a State-Of-The-Art (SOTA) research conference in November 2016 focused on nonpharmacological treatments for chronic musculoskeletal pain (CMSP).

One request that emerged from the SOTA was to develop a consensus about core measures for clinical research investigating CMSP in order to facilitate cross-study comparability including the degree of effectiveness of different treatments. Therefore, the VHA convened a work group for which the principal charge was “to recommend core *outcome* measures for *pain intensity and interference* to be used in all VHA prospective clinical research studies of *chronic musculoskeletal* pain (both interventional and observational).” Secondly, outcome measures for other important domains were evaluated. Although several other expert groups have recommended core measures for pain research, priorities for the present group included not only strong psychometrics of measures but also pragmatic issues such as brevity, applicability to any type of CMSP (rather than a specific condition only such as low back pain or osteoarthritis), and utility in studies that may incorporate patient-reported outcomes in practice-based research and electronic health records.

## Methods

### Pain Measures Work Group Process

An 11-member Pain Measures Work Group (PMWG) composed of VHA and non-VHA individuals with expertise in pain research, measurement development, and psychometrics was appointed and completed its task over 12 months (January through December 2017). The process involved 1) eight conference calls (90 minutes each); 2) a one-day in-person meeting; 3) evaluating the results of a systematic review on pain intensity and interference measures (see below) [7,8]; 4) reviewing and comparing the recommendations of several previous consensus reports [9–11]; 5) conducting two Delphi surveys of PMWG members to identify core domains and desirable characteristics of pain measures and to achieve consensus on the final outcome measures to recommend. This process was adapted from guidelines for developing core outcome sets for research [12].

### Systematic Review of Pain Intensity and Interference Measures

To inform the PMWG, the VHA sponsored a systematic review of pain intensity and interference measures

through its Evidence-based Synthesis Program (ESP). Intensity refers to the severity of the pain (the two terms are often used interchangeably), and interference refers to the effect of pain on specific areas of functioning. The PMWG members, as well as key participants with expertise in pain measurement who had attended the SOTA, identified a list of 17 potential measures for the ESP to review. Published in detail elsewhere [7,8], the ESP report identified 43 articles that met the inclusion criteria of 1) evaluating at least one of the 17 pain measures; 2) including adults with chronic musculoskeletal pain of at least three months' duration or adults with musculoskeletal pain described as “chronic” by the study authors; and 3) reporting on at least one of the four psychometric outcomes—minimally important difference, reliability, validity, and responsiveness to change. The ESP systematic review 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.

### Recommendations by Other Expert Groups

Three other reports relevant to the PMWG's charge were identified and reviewed, including the 1) Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) [9,13]; 2) National Institute of Health Task Force on Research Standards for Chronic Low Back Pain [10]; 3) Core Outcome Set steering group recommendations for nonspecific LBP research [11,14]. Other literature syntheses were reviewed but in general did not achieve consensus on specific measures [15–17].

### Delphi Surveys

Two surveys of PMWG members using a Delphi process were conducted. The first survey identified and prioritized core outcome domains for CMSP clinical research, and the second survey achieved consensus on a recommended measure for each domain.

## Results

### Defining Chronic Pain

Although not an outcome measure, the definition of chronic pain is also considered important to CMSP research. The National Health Interview Survey (NHIS) defines chronic pain as pain that has been present on most days for the past three months [18], whereas the National Institutes of Health (NIH) Task Force defines it as pain for at least three months that has been present on more than half of the days in the past six months [10].

The PMWG favors the NHIS definition due to its simplicity and its extensive use in epidemiological surveys. Also, the two definitions are conceptually equivalent in that both approximate an accumulation of three months' worth of pain days.

### Systematic Review of Pain Intensity and Interference Measures

**Table 1** summarizes psychometric data on the 17 pain intensity and interference measures included in the ESP systematic review [7,8]. Citations for these measures are available in the ESP report. The table was adapted from original tables in the ESP review with two important modifications. First, data from three recent Department of Veterans Affairs (VA) trials comprising 759 patients [19] have been added. Second, the type of response options (numeric vs verbal) for each scale has been noted. The pain intensity and interference measures with the strongest psychometric data include three numeric (numerical rating scale [NRS], Brief Pain Inventory [BPI], and Pain intensity and pain interference with Enjoyment of life and General activities [PEG] and two verbal (Patient-Reported Outcome Information System [PROMIS] and Short-Form 36 [SF-36]) rating scales. Some scales measure intensity/severity or interference only, others have subscales for both domains, and two (PEG, SF-36) provide a single composite score that includes both domains. Other scales with substantial psychometric evidence, such as the Roland Morris Disability Questionnaire (RMDQ) and Oswestry Disability Index (ODI) for low back pain and the WOMAC for osteoarthritis, focus on condition-specific types of musculoskeletal pain. As the PMWG charge was to recommend general measures (i.e., ones that could be used across different types of CMSP conditions), condition-specific measures were not further considered. One other general scale with moderate evidence is the visual analog scale (VAS), but because it is similar in purpose to the single-item NRS and the latter has much more psychometric evidence [8] as well as several practical advantages (patient preferences, fewer missing data) [9], the VAS was not further considered.

### Responsiveness of Pain Intensity and Interference Measures

**Table 2** summarizes the area under the curve (AUC) for detecting pain improvement using data from seven studies for eight different scales [19–23]. The original table in the ESP review [7] has been augmented by data from the three recent VHA trials [19]. In estimating the AUC, the data from these three trials involved 594 instead of 759 participants as longitudinal (two time points) rather than cross-sectional data are required to calculate AUCs for improvement. AUCs for pain improvement ranged from 0.61 to 0.77. Whereas good AUCs for diagnostic tests compared with a criterion standard typically exceed

0.80, a reasonable AUC for the ability of a measure to detect improvement is often somewhat lower (0.65 or higher) [20,21,24,25]. Moreover, the key purpose of our report is to assess the comparative rather than absolute responsiveness of different pain intensity and interference measures.

**Table 3** summarizes data from five trials on another measure of responsiveness—standardized response mean (SRM)—which in this case was the standardized changes in the scale score between baseline and follow-up assessments among patients who rate their pain at follow-up as “better,” “unchanged (same),” or “worse.” The SRM data from two trials (SCAMP, SCOPE) have been published [20,21], whereas some psychometric data (but not the SRM findings) from the other three trials have been reported [19]. The measures with the most evidence to estimate SRMs across trials were the BPI interference (five trials), PEG (five trials), and PROMIS (four trials) scales, followed by the SF-36 (three trials). The average SRM across trials for these four scales in patients who reported pain improvement ranged from 0.56 to 0.84; for those who reported their pain being the same, 0.31 to 0.38; and for those who reported their pain as worse,  $-0.09$  to  $-0.24$ . Interpretation of SRMs is similar to that of effect sizes, where 0.2, 0.5, and 0.8 represent small, moderate, and large changes, respectively [26]. Three findings should be noted. First, the SRMs for improvement and worsening were not symmetric; the SRMs for improvement were substantially larger in absolute magnitude than the SRMs for worsening. Second, although SRMs for the unchanged group typically did not center on 0, they were not as large as the SRMs for improvement. Third, the SRMs for the improved, unchanged, and worse groups differed significantly from one another, meaning the scales differentiated among the three groups. These findings are consistent with previous reports [20,21,24].

### Recommended Measures by Other Expert Groups

**Table 4** compares the core domains and specific measures recommended by various expert groups [7,9,10,14]. The NRS and BPI interference scales are the most commonly recommended pain intensity and interference measures, respectively.

### PMWG Prioritized Secondary Domains and Recommended Measures

**Table 5** shows the PMWG's prioritized domains and recommended measure for each domain. Panelists rated each domain as 1 (important), 2 (very important), or 3 (essential). After pain intensity and inference, the next most highly ranked domains were physical functioning and depression (2.4 each), followed by anxiety (1.7), sleep (1.4), and patient-rated change (1.3). The NRS (one item) and BPI (seven items) are the measures endorsed for assessment of pain intensity and interference,

**Table 1.** Overview of pain severity and interference measures\*

Pain Measure	Response Options	No. of Studies	No. of Participants	Length No. of Items	No. of Studies Reporting					Pain Domain	Public Domain
					MID	Responsive	Concurrent Validity	Test-Retest Reliability			
General pain											
Numerical rating scale (NRS)	Numeric	11	1,653	1	3	5	4	2	S	Yes	
Brief Pain Inventory (BPI)	Numeric	9	2,590	11	4	9	6		S+I	No	
PEG	Numeric	6	1,765	3	4	6	4		SI comp	Yes	
Patient-Reported Outcome Information System (PROMIS)	Verbal	7	1,458	4-8	3	6	4	1	I	Yes	
SF-36 Bodily Pain Scale	Verbal	11	2,308	2	2	6	8		SI comp	No	
Graded Chronic Pain Scale (GCPS)	Numeric	3	1,058	12	1	2	2		S+I	Yes	
McGill Pain Questionnaire (MPQ)	Verbal	3	366	78 <sup>†</sup>		1	2	1	S	Yes	
Multidimensional Pain Inventory (MPI)	Numeric	4	708	9		1	4	1	S+I	Yes	
Visual analog scale (VAS)	Numeric	8	541	1	2	2	3	3	S	Yes	
Patient global impression of change (PGIC)	Verbal	1	476	1		1			n/a	Yes	
Disorder-specific											
Oswestry Disability Index (ODI)	Verbal	10	1,149	10	3	5	6	3	S+I	No	
Roland Morris Disability Questionnaire (RMDQ)	Verbal	10	4,157	24	4	5	6	1	I	Yes	
Western Ontario McMasters Osteoarthritis Index (WOMAC)	Verbal	5	535	24		3	2	1	S+I	No	
Minimal evidence											
Wong Faces Scale	Numeric	1	32	1				1	S	No	
Hip disability and Osteoarthritis Outcome Score (HOOS)	Verbal	1	62	40		1	1		S+I	Yes	
Knee injury and Osteoarthritis Outcome Score (KOOS)	Verbal	0		42					S+I	Yes	
Defense & Veterans Pain Rating Scale (DVPRS)	Numeric	0		5					S+I	Yes	

I = interference score only; MID = minimally important difference; S = severity (intensity) score only; S+I = separate severity and interference scores; SI comp = composite severity-interference score.

\*Table was adapted from Goldsmith et al. [7,8], including the addition of data from Chen et al. [19] from three other trials.

<sup>†</sup>Pain Rating Index based on 78 pain descriptors; Present Pain Intensity based on six additional items.

respectively. For secondary domains, the recommended measures are the Patient Health Questionnaire two-item depression scale (PHQ-2) for depression, Generalized Anxiety Disorder two-item anxiety scale (GAD-2) for anxiety, and PROMIS four-item scales for physical function and sleep. Fatigue and pain catastrophizing were discussed as potential domains but were not recommended as core measures due to greater uncertainty about treatability, lack of endorsement by other expert groups, and the desire to limit the number of domains to those considered most important. The final set of measures recommended by the PWMQ are included in the [Supplementary Data](#).

PWMG members were asked to rank the importance of secondary (nonpsychometric) characteristics of pain measures on a 1 (low importance) to 5 (high importance) scale. The secondary scale characteristics most highly rated included the scale being public domain (i.e., no fee to use), brief, and feasible for clinical use (all with a median rating of 5), followed by feasibility of incorporating it into electronic health records (4.0), emerging/increasing use (3.0), number of studies that have used the measure (3.0), and number of translations into languages

other than English (2.0). Although not ranked, other desirable characteristics suggested by a few PMWG members included simplicity of scoring, clinically interpretable cut-points, and data on cross-walking scores with other measures.

A single measure rather than several options was recommended for each domain to encourage common metrics that would facilitate direct comparison of results across different studies. The closest competitors to secondary domain scales that were ultimately chosen included the SF-36 (10 items) for physical function; the PHQ-9 (nine items) and PROMIS (four items) scales for depression; the brief Pittsburgh Insomnia Scale (two items) for sleep; and the GAD-7 (seven items) and PROMIS (four items) for anxiety.

## Discussion

The PMWG's principal charge was to recommend core pain intensity and interference measures to be included in prospective studies focused on chronic musculoskeletal pain. Findings from the ESP systematic review augmented by three recent studies did not identify a clear

**Table 2. Comparative responsiveness of measures based on AUC values for detecting any improvement\***

Study (Sample Size)	BPI (Severity) (N = 3)	BPI (Interference) (N = 6)	BPI (Total) (N = 3)	PEG (N = 6)	SF-36 BPS (N = 5)	PROMIS (Mean) <sup>†</sup> (N = 4)	RMDQ (N = 4)	GCPS (Severity) (N = 2)	GCPS (Disability) (N = 2)	NRS (N = 2)	ODI (N = 1)
Average	0.77	0.71	0.77	0.71	0.68	0.64	0.71	0.77	0.70	0.61	0.67
Krebs et al. [20] 2010 RCT (N = 205)	0.81	0.78	0.81	0.78	0.72	-	0.81	0.78	0.75		
Krebs et al. [20] 2010 cohort (N = 222)	0.83	0.70	0.78	0.73	0.68	-	0.70	0.75	0.65		
Kean et al. [21] 2016 (N = 244)	0.68	0.73	0.73	0.71	0.68	0.58	-				
Maughan and Lewis [22] 2010 (N = 48)							0.64			0.50	0.67
Stewart et al. [23] 2007 (N = 134)					0.73					0.69 <sup>‡</sup>	
Bair [19] 2017 (N = 134)		0.73		0.72	0.58	0.71	0.67				
Krebs [19] 2017 (N = 222)		0.77		0.79		0.69					
Damush [19] 2017 (N = 238)		0.54		0.54		0.56					

AUC=area under the curve (values range from 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; BPS = Bodily Pain Scale; GCPS = Graded Chronic Pain Scale; NRS = numerical rating scale; ODI = Oswestry Disability Index; PROMIS = Patient-Reported Outcome Information System; RCT = randomized controlled trial; RMDQ = Roland Morris Disability Questionnaire.

\* Adapted from Goldsmith et al. [7,8] including the addition of data from three other trials (Chen et al. [19]).

<sup>†</sup>Three versions were administered in Kean (PROMIS four- and eight-item profiles, and short-form), and four versions were administered in Bair, Krebs, and Damush (four-, six-, and eight-item profiles, and short-form). The range across the PROMIS scales was very tight, so the mean is reported.

<sup>‡</sup>Two measures administered, pain intensity and pain bothersomeness (mean is reported).

**Table 3. Standardized response means for patient global impression of change categories by pain scales in five clinical trials\***

Scale (No. of Trials)	Average Across Trials			CAMEO (N = 134)			SPACE (N = 222)			SSSM (N = 238)			SCOPE (N = 244)			SCAMP (N = 427)			
	B	S	W	B	S	W	B	S	W	B	S	W	B	S	W	B	S	W	
<b>Most evidence</b>																			
BPI interference (5)	0.79	0.31	-0.09	0.75	0.34	-0.24	1.19	0.42	0.22	0.17	0.19	0.19	0.38	0.03	0.24	0.86	0.24	-0.24	-0.15
PEG (5)	0.84	0.31	-0.13	0.77	0.30	-0.22	1.42	0.58	0.08	0.18	0.18	-0.24	0.86	0.25	-0.14	0.99	0.25	-0.15	-0.15
PROMIS, range (4)	0.56	0.31	-0.23	0.55	0.18	-0.19	1.05	0.52	-0.02	0.14	0.16	-0.62	0.33	0.27	-0.02	0.67	0.11	-0.45	-0.57
SF bodily pain (3)	0.63	0.38	-0.24	0.65	0.23	-0.27	1.08	0.62	0.14	0.20	0.19	-0.69	0.51	0.30	-0.16	0.71	0.38	-0.17	-0.36
<b>Less evidence</b>																			
BPI severity (2)	0.86	0.05	-0.48	0.47	0.53	-0.20							0.71	0.13	-0.47	1.00	-0.04	-0.49	-0.37
BPI total (2)	0.98	0.25	-0.30										0.93	0.31	-0.22	1.02	0.18	-0.37	-0.45
Roland (2)	0.73	0.19	-0.34	0.56	0.27	-0.23							0.90	0.11	-0.45	0.67	-0.02	-0.57	-0.27
GCPS severity (1)	0.67	-0.02	-0.57										0.77	0.12	-0.27				
GCPS disability (1)	0.77	0.12	-0.27																

Standardized response mean = [baseline mean score - follow-up mean score]/SD of change score.

B= better; S= same; W= worse; CAMEO = Care Management for the Effective use of Opioids; SPACE = Strategies for Prescribing Analgesics Comparative Effectiveness; SSSM = Stroke Survivor Self-Management; SCOPE = Stepped Care Optimizing Pain Care Effectiveness; SCAMP = Stepped Care for Affective disorders and Musculoskeletal Pain (as classified by single-item patient-rated global impression of change).

\* Four trials were interventions for chronic musculoskeletal pain. One trial (SSSM) focused on risk factor self-management in stroke survivors.

**Table 4.** Core domains/measures recommended by other groups for pain research

Domain (Pain Site)	Recommended Scale (No. of Items)					
	NIH Research Task Force [10]		IMMPACT [9,13] (Any)	Core Outcome Set Nonspecific LBP [11,14] (Low Back)	VA ESP Report [7,8] (Any)	VA Work Group (Any)
	Recommended (Low Back)	Alternative				
Pain intensity	NRS (1)	BPI (4)	NRS (1)	NRS (1)	NRS (1), VAS (1), MPQ (78)	NRS (1)
Pain interference	PROMIS (4)	BPI (7)	BPI (7) or MPI (9)	BPI (7) or MPI (9)	BPI (7), PEG (3), SF-36 (2), MPI (9), GCPS (7)	BPI (7)
Physical function	PROMIS (4)	RMDQ (24) or ODI (10)	SF-36 (10)	RMDQ (24) or ODI (10)	RMDS (24), ODI (10), WOMAC (24)	PROMIS (4)
Depression	PROMIS (4)	PHQ-9 or BDI	BDI and xc POMS	HADS (7)		PHQ-2 (2)
Sleep	PROMIS (4)					PROMIS (4)
Patient-rated change			PGIC (1)			PGIC (1)
Anxiety				HADS (7)		GAD-2 (2)
Work disability	2 items			1–2 items		
Other	*			†		

BDI=Beck Depression Inventory; GAD-2=Generalized Anxiety Disorder two-item anxiety scale; HADS=Hospital Anxiety and Depression Scale; MPI=Multidimensional Pain Inventory; MPQ = McGill Pain Questionnaire; ODI=Oswestry Disability Index; NRS=numerical rating scale; PGIC=patient-rated global impression of change; PHQ-2=Patient Health Questionnaire two-item depression scale; POMS=Profile of Mood States; PROMIS=Patient-Reported Outcomes Measurement Information System; RMDQ=Roland Morris Disability Questionnaire; VAS=visual analog scale; WOMAC=Western Ontario McMasters Osteoarthritis Index.

\*Other domains (No. of items) include substance use (tobacco, alcohol, drugs; 3), other pain sites (5), prior surgery (3), pain treatments (4), catastrophizing (1).

†Other domains (No. of items) include health-related quality of life (5–12).

**Table 5.** Recommended measures for prospective research of musculoskeletal pain

Domain	Mean Rating*	Recommended Measure	Items
Pain intensity	3.0	Numerical rating scale (NRS)	1
Pain interference	3.0	Brief Pain Inventory (BPI) <sup>†</sup>	7
Physical function	2.4	PROMIS Profile–29	4
Depression	2.4	Patient Health Questionnaire–2 (PHQ-2)	2
Anxiety	1.7	Generalized Anxiety Disorder–2 (GAD-2)	2
Sleep	1.4	PROMIS Profile–29	4
Global change	1.3	Patient global impression of change (PGIC)	1

\*Mean importance rating of domain by 10 expert panelists, assigning three points for essential, two points for very important, and one point for important.

<sup>†</sup>For permission to use the Brief Pain Inventory, go to its developers' website: <https://www.mdanderson.org/research/departments-labs-institutes/departments-divisions/symptom-research/symptom-assessment-tools/brief-pain-inventory.html>.

“psychometric winner” but instead several psychometrically valid scales from which to choose. Although psychometric soundness was considered a necessary criterion, it could not be the sole deciding factor for several reasons: 1) the ESP review found only a small number of studies for any given measure; 2) there were substantial methodological and population differences across these studies; 3) studies comparing the same measures across different populations can show substantial psychometric differences; 4) studies comparing different

measures within the same population tend to show relatively comparable responsiveness.

It is important to acknowledge that the recommendations of the PMWG are directly related to its specific charge, namely, to focus on pain intensity and interference. The recommendations of the PMWG should not be taken as precluding the inclusion of other important domains both in clinical practice and research. The inclusion of other domains and additional measures should be directly related to the questions of particular interest addressed by a clinician or a clinical researcher. In addition, we are not suggesting which measures should be used as the primary outcome for studies. For example, a study focused exclusively on low back pain could include the PMWG measures but still select a validated LBP scale as the primary outcome (plus items that capture LBP-unique features such as the radicular pain of sciatica).

For several reasons, the PMWG recommends the NRS (one item) and BPI interference scale (seven items) as minimum core measures to assess pain intensity and interference, respectively. First, rating scales with numeric rather than verbal response options have a somewhat longer track record of use in clinical trials and other research studies. Second, the cut-points for differing levels of pain on 0–10 NRS have been extensively studied. For example, one common cut-point range (1–3 representing mild pain, 4–6 moderate pain, and 7–10 severe pain) has been used in major VHA quality improvement initiatives [27]. A recent literature synthesis suggests 1–4, 5–6, and 7–10 as evidence-based ranges [28]. Third, the

responsiveness of the BPI and NRS equals or exceeds that of the most-studied verbal response scales (PROMIS and SF-36), as shown in [Tables 2](#) and [3](#). Fourth, the BPI and NRS are the most commonly recommended measures by other expert groups ([Table 4](#)). Although the evidence is inconclusive regarding how well verbal ratings of pain correspond to numerical ratings [[21,29](#)], methods providing a cross-walk between the two types of ratings have been developed [[19,30,31](#)].

Use of the BPI does require permission from its developers, and in some circumstances a modest fee. If this is a barrier to use for some researchers or clinicians, two alternative options using public domain measures are 1) the NRS for pain severity and the PROMIS (four items) for pain interference or 2) the three-item PEG, which includes the NRS plus two items from the BPI and provides a composite severity-interference score. Nonetheless, the PMWG recommends that VHA researchers include the NRS and BPI as a minimal data set to facilitate cross-study comparisons using common metrics.

There is some debate as to whether pain intensity and pain interference need to be assessed as separate domains vs a single composite score. Although the separate domain approach has been traditionally favored, some recent evidence shows that composite scales such as the PEG, BPI total score, and SF-36 bodily pain scale perform similarly to separate scores [[19–21,24](#)]. Composite scales have the additional advantage of providing a single measure as a primary outcome in research studies as well as a single score with which clinicians can monitor outcomes and adjust treatment. The brevity of a composite measure like the PEG may also be suited to studies where pain is a secondary outcome and a smaller number of items is desirable. Nonetheless, the core outcomes recommended by the PMWG include separate measures of pain intensity and interference.

Three measures for which the ESP review provided substantial evidence include two for LBP (RMDQ and ODI) and one for OA (WOMAC). Because the PMWG sought measures that could be used across all studies regardless of CMSP condition, these three condition-specific measures were not deemed eligible for the core measure set. However, they are certainly strong measures for LBP and OA research, although we also recommend inclusion of the core measures in [Table 5](#). Of note, the RMDQ has been adapted for research that includes patients with CMSP other than LBP [[32–35](#)]. It is also available in a shorter 11-item version [[36](#)].

IMMPACT is the other expert group to have provided recommendations for pain measures that were not condition-specific. Our PMWG report differs from IMMPACT in several ways. First, there have been new pain measures and some advances in the psychometric research since the IMMPACT report was published 15 years ago. Second, IMMPACT provided more than one option for several domains, whereas a principal intent of PMWG deliberations was to recommend only one

measure for each domain so that a common set of measures might be used across different pain studies. Third, brevity of measures was an important aim of the PMWG in order to maximize the use of core measures across many types of research settings, including practice-based studies and those using the electronic health record.

A key issue was whether to preferentially recommend a PROMIS measure when it was available for a specific domain. Strengths of PROMIS measures include heavy investment by the NIH in rigorous psychometric development, their grounding in item-response theory, the ability to obtain highly reliable scores by completion of fewer than 10 items when administered by computer adaptive testing (CAT), and conversion of raw to T-scores, for which 50 represents the population norm regardless of domain. However, CAT administration is not yet feasible across all research and clinical settings. Also, some scale users find responses to individual items useful (complementing the overall scale score), preferring fixed-item scales where the same items are administered at each assessment. PROMIS does provide the option of fixed-item scales (which are included for some domains by the NIH Task Force and a few domains by our VHA PMWG). However, the advantages of PROMIS measures are somewhat attenuated when PROMIS fixed-length scales are compared with non-PROMIS legacy scales, such as the BPI for pain interference, the PHQ-9/PHQ-2 for depression, and the GAD-7/GAD-2 for anxiety. In these domains, the PROMIS scales do not yet have as strong a track record in clinical trials and other research. However, with increasing use of PROMIS measures, it is good that cross-walks between PROMIS and non-PROMIS legacy scales are available [[19,31,37](#)].

At a minimum, the PMWG recommends that common measures of pain intensity and interference be used across all prospective clinical studies of CMSP. The other domains in [Table 5](#) would be desirable to include whenever possible. For example, a general measure of physical functioning does not require individuals with chronic pain and comorbid medical and mental conditions to decide how much of their functional impairment is attributable to chronic pain vs other concurrent disorders (an often difficult task). The focus on brief measures for both primary and secondary domains is driven by the desire to minimize respondent burden and to encourage use in pragmatic trials, which may depend upon measures routinely gathered in clinical care and incorporated into electronic medical records. Individuals can typically complete around four items per minute; thus, the BPI plus NRS should only take a few minutes to complete and, even with the addition of the secondary measures in [Table 5](#), no longer than five minutes. As pain is ubiquitous across most medical and mental disorders, the availability of brief public domain measures may encourage assessment of pain as a secondary outcome in many other areas of clinical research.

Although the anticipated initial audience for this report is VHA researchers, four of the PMWG members

were clinical researchers from outside the VHA. Also, much of the evidence supporting the recommended measures was generated in nonveteran samples. Thus, the recommendations are likely relevant to clinical research outside the VHA. Finally, although the targeted condition was chronic musculoskeletal pain, the recommended domains and measures may also be applicable across non-CMSP pain disorders. This is especially salient given the fact that many patients with chronic pain report pain across multiple musculoskeletal and nonmusculoskeletal locations [38,39]. Thus, brief pain measures that are not restricted to a specific bodily site may serve as valuable common metrics across pain studies, even when longer or more disease-specific measures are used as primary outcomes.

## Supplementary Data

Supplementary data are available at *Pain Medicine* online.

## References

- Institute of Medicine. *Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research*. Washington, DC: National Academies Press; 2011.
- Schappert SM. National Ambulatory Medical Care Survey: 1989 summary. *National Center for Health Statistics. Vital Health Stat* 1992;13(110):1–80.
- Lorenz KA, Krebs EE, Bentley TG, et al. Exploring alternative approaches to routine outpatient pain screening. *Pain Med* 2009;10(7):1291–9.
- U'S Burden of Disease Collaborators. The state of US health, 1990-2010: Burden of diseases, injuries, and risk factors. *JAMA* 2013;310(6):591–608.
- Turk DC. Clinical effectiveness and cost-effectiveness of treatments for patients with chronic pain. *Clin J Pain* 2002;18(6):355–65.
- Kroenke K, Cheville A. Management of chronic pain in the aftermath of the opioid backlash. *JAMA* 2017; 317(23):2365–6.
- Goldsmith ES, Murdoch M, Taylor B, et al. Rapid Evidence Review: Measures for Patients with Chronic Musculoskeletal Pain. VA ESP Project#09-009; 2017.
- Goldsmith ES, Taylor BC, Greer N, et al. Focused evidence review: Psychometric properties of patient-reported outcome measures for chronic musculoskeletal pain. *J Gen Intern Med* 2018;33(suppl 1):61–70.
- Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005;113(1):9–19.
- Deyo RA, Dworkin SF, Amtmann D, et al. Report of the NIH Task Force on Research Standards for Chronic Low Back Pain. *J Pain* 2014;15(6):569–85.
- Chiarotto A, Boers M, Deyo RA, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain* 2018;159(3):481–95.
- Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a “Core Outcome Set”—a practical guideline. *Trials* 2016;17(1):449.
- Turk DC, Dworkin RH, Allen RR, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2003;106(3):337–45.
- Chiarotto A, Terwee CB, Ostelo RW. Choosing the right outcome measurement instruments for patients with low back pain. *Best Pract Res Clin Rheumatol* 2016;30(6):1003–20.
- Kaiser U, Neustadt K, Kopkow C, Schmitt J, Sabatowski R. Core outcome sets and multidimensional assessment tools for harmonizing outcome measure in chronic pain and back pain. *Healthcare (Basel)* 2016;4(3):63.
- 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):1836–50.
- Turk DC, Fillingim RB, Ohrbach R, Patel KV. Assessment of psychosocial and functional impact of chronic pain. *J Pain* 2016;17(suppl 9):T21–49.
- Kennedy J, Roll JM, Schraudner T, Murphy S, McPherson S. Prevalence of persistent pain in the U.S. adult population: New data from the 2010 National Health Interview Survey. *J Pain* 2014;15(10):979–84.
- Chen CX, Kroenke K, Stump T, et al. Estimating minimally important differences for the PROMIS Pain Interference scales: Results from 3 randomized clinical trials. *Pain* 2018;154(4):775–82.
- Krebs EE, Bair MJ, Wu J, et al. Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. *Med Care* 2010;48(11):1007–14.
- 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. *Med Care* 2016;54(4):414–21.
- Maughan EF, Lewis JS. Outcome measures in chronic low back pain. *Eur Spine J* 2010;19(9):1484–94.
- Stewart M, Maher CG, Refshauge KM, Bogduk N, Nicholas M. Responsiveness of pain and disability measures for chronic whiplash. *Spine (Phila Pa 1976)* 2007;32(5):580–5.
- Kroenke K, Theobald D, Wu J, Tu W, Krebs EE. Comparative responsiveness of pain measures in cancer patients. *J Pain* 2012;13(8):764–72.
- Johns SA, Kroenke K, Krebs EE, et al. Longitudinal comparison of three depression measures in adult cancer patients. *J Pain Symptom Manage* 2013;45(1):71–82.



26. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989; 27(suppl 3):S178–89.
27. Cleeland CS, Reyes-Gibby CC, Schall M, et al. Rapid improvement in pain management: The Veterans Health Administration and the Institute For Healthcare Improvement Collaborative. *Clin J Pain* 2003;19(5):298–305.
28. Woo A, Lechner B, Fu T, et al. Cut points for mild, moderate, and severe pain among cancer and non-cancer patients: A literature review. *Ann Palliat Med* 2015;4(4):176–83.
29. Dijkers M. Comparing quantification of pain severity by verbal rating and numeric rating scales. *J Spinal Cord Med* 2010;33(3):232–42.
30. Edelen MO, Saliba D. Correspondence of verbal descriptor and numeric rating scales for pain intensity: An item response theory calibration. *J Gerontol A Biol Sci Med Sci* 2010;65(7):778–85.
31. Cook KF, Schalet BD, Kallen MA, Rutsohn JP, Cella D. Establishing a common metric for self-reported pain: Linking BPI Pain Interference and SF-36 Bodily Pain Subscale scores to the PROMIS Pain Interference metric. *Qual Life Res* 2015;24(10):2305–18.
32. Kroenke K, Bair MJ, Damush TM, et al. Optimized antidepressant therapy and pain self-management in primary care patients with depression and musculoskeletal pain: A randomized controlled trial. *JAMA* 2009;301(20):2099–110.
33. Dobscha SK, Corson K, Perrin NA, et al. Collaborative care for chronic pain in primary care: A clustered randomized trial. *JAMA* 2009;301(12):1242–52.
34. Bair MJ, Ang D, Wu J, et al. Evaluation of Stepped Care for Chronic Pain (ESCAPE) in veterans of the Iraq and Afghanistan conflicts: A randomized clinical trial. *JAMA Intern Med* 2015;175(5):682–9.
35. Krebs EE, Jensen AC, Nugent S, et al. Design, recruitment outcomes, and sample characteristics of the Strategies for Prescribing Analgesics Comparative Effectiveness (SPACE) trial. *Contemp Clin Trials* 2017;62:130–9.
36. 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):257–63.
37. Choi SW, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: Linking the BDI-II, CES-D, and PHQ-9 to PROMIS Depression. *Psychol Assess* 2014;26(2):513–27.
38. Kroenke K, Krebs E, Wu J, et al. Stepped Care to Optimize Pain Care Effectiveness (SCOPE) Trial: Study design and sample characteristics. *Contemp Clin Trials* 2013;34(2):270–81.
39. Kroenke K, Evans E, Weitlauf S, et al. Comprehensive vs. Assisted Management of Mood and Pain Symptoms (CAMMPS) trial: Study design and sample characteristics. *Contemp Clin Trials* 2018;64:179–87.