QUERI

Rapid Evidence Review: Measures for Patients with Chronic Musculoskeletal Pain

Supplemental Content

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.

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 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

SUPPLEMENTAL TABLE 1. SEARCH STRATEGY

1	exp Low Back Pain/ or exp Shoulder Pain/ or exp Back Pain/ or exp Musculoskeletal Pain/ or exp Chronic Pain/ or exp Neck Pain/
2	(pain and (musculoskeletal or (low adj back) or neck or shoulder or hip or knee or joint)).mp.
3	osteoarthritis.mp. or exp Osteoarthritis/
4	1 or 2 or 3
5	exp Pain Measurement/mt
6	(pain adj5 (questionnaire$ or assess$ or measur$ or scale$ or inventor$ or rating$ or tool$)).mp.
7	(BPI or PEG or SF-36 or PROMIS or McGill or DVPRS or Roland-Morris or WOMAC or Oswestry or KOOS or HOOS or (Faces adj Scale)).mp.
8	5 or 6 or 7
9	(pain adj (severity or intensity or function$ or limit$ or activit$ or impact$ or interfer$ or disabilit$)).mp.
10	(valid$ or reliab$ or feasib$ or generalizab$ or respons$ or implement$).mp.
11	4 and 8 and 9 and 10
12	limit 12 to (english language and humans and yr="2000 -Current")

SUPPLEMENTAL TABLE 2. PEER REVIEW COMMENTS/AUTHOR RESPONSES

Question Text	Comment	Author Response
Are the objectives, scope, and methods for this review clearly described?	Yes	Thank you
	Yes	Thank you
	Yes	Thank you
	Yes	Thank you
	Yes	Thank you
Is there any indication of bias in our synthesis of the evidence?	No	Thank you
	No	Thank you
	No	Thank you
	No	Thank you
	No	Thank you
Are there any published or unpublished studies that we may have overlooked?	Yes - Please see my major comment below.	Please see our response to these major comments below.
	No	Thank you
	Yes - I have some concerns about the time period examined, as detailed below.	Please see our response below.
	Yes - I am concerned that some studies may have been missed. For example, re: the PROMIS-PI scale please verify that the following studies were screened and excluded. Amtmann, D. A., Cook, K. F., Jensen, M. P., Chen, W-H., Choi, S. W., Revicki, D., Cella, D., Rothrock, N., Keefe, F., Callahan, L., Lai, J-S. (2010). Development of a PROMIS item bank to measure pain interference. Pain, 150(1), 173-82. Amtmann, D., Kim, J., Chung, H., Askew, R. L., Park, R., & Cook, K. F. (2016). Minimally important differences for Patient Reported Outcomes Measurement Information System pain interference for individuals with back pain. Journal of Pain Research, 9, 251-255. Askew, R. L., Kim, J., Chung, H., Cook, K. F., Johnson, K. L., & Amtmann, D. (2013). Development of a crosswalk for pain interference measured by the BPI and PROMIS pain interference short form. Quality of Life Research, 10.1007/s11136-013-0398-5. Broderick, J. E., Schneider, S., Junghaenel, D. U., Schwartz, J. E., & Stone, A. A. (2013). Validity and reliability of Patient-Reported Outcomes Measurement Information System instruments in osteoarthritis. Arthritis Care and Research, 5(10), 1625-1633.	The suggested references were reviewed for eligibility and did not meet inclusion criteria. Amtmann (2010): the study population did not meet the requirement that >75% of participants have chronic musculoskeletal pain Amtmann (2016): the study population did not meet the requirement that >75% of participants have chronic musculoskeletal pain Askew: the study population was comprised of multiple sclerosis patients (not musculoskeletal pain) Broderick: the study population was comprised of patients who self-reported a physician diagnosis of osteoarthritis. It is unclear whether such diagnoses were radiologically or clinically defined. Details were not provided on presence or duration of pain. Merriwether: we agree with a reviewer's suggestion to include fibromyalgia; this study is included in the final report. Papuga: the duration of pain associated with conditions of the spine in this population was not reported.

Question Text	Comment	Author Response
	Merriwether, E. N., Rakel, B. A., Zimmerman, M. B., Dailey, D. L., Vance, C. G., Darghosian, L., . . . Sluka, K. A. (2016). Reliability and construct validity of the Patient-Reported Outcomes Measurement Information System (PROMIS) instruments in women with fibromyalgia. Pain Medicine. doi:10.1093/pm/pnw187 Papuga, M. O., Mesfin, A., Molinari, R., & Rubery, P. T. (2016). Correlation of PROMIS physical function and pain CAT instruments with Oswestry Disability Index and Neck Disability Index in spine patients. Spine. doi:10.1097/BRS.0000000000001518 Also, I am concerned that the exclusion criteria may have resulted in exclusion of relevant studies (see comments below). For example, many studies investigating the psychometric properties of the pain scales have been published using non-English language versions of the scales of interest. It is not clear to me the rationale for excluding such studies, as this excludes a huge chunk of the literature on this topic. If the authors were concerned that findings could be affected by use of translated versions of a scale, it would be easy to assess whether that is the case.	We excluded results from non-English language versions of scales. We added information to support this decision in the Limitations section (page 31 We disagree that "it would be easy to assess" whether findings could be affected by use of translated versions of a scale, as psychometric properties are affected by a number of factors other than linguistic and cultural variation, and isolating the influence of language variation would not be a straightforward process.
	No	Thank you.
Additional suggestions or comments can be provided below. If applicable, please indicate the page and line numbers from the draft report.	I have one major concern and a number of smaller comments. Major comment: I am confused about the inclusion/exclusion criteria related to chronic pain conditions. Exclusion criteria include "studies of patients with chronic conditions typically associated with pain unless the study specified that the patients had CMP (eg, osteoarthritis)." Does this mean that a study conducted in an osteoarthritis population would be excluded unless the authors specified that patients had "CMP"? If so, this seems to contradict the key question, which indicates that the population of interest has "chronic (≥ 3 months) musculoskeletal pain (eg, low back pain, osteoarthritis, and non-traumatic joint pain)." The pain field suffers from a lack of consensus on terminology (e.g., "chronic pain" vs "persistent pain") and pain diagnosis categories, so substantial heterogeneity in descriptions of clinical populations is to be expected. "Chronic musculoskeletal pain" is not a specific entity, just an umbrella term used to capture a group of patients with chronic painful conditions, such as those with low back pain and osteoarthritis. Excluding studies of chronic pain measures conducted in patients with chronic back pain and osteoarthritis that do not describe patients as specifically having "chronic musculoskeletal pain" does not seem to make sense. It's possible that I'm just misunderstanding the exclusion criterion. It would	We clarified the study inclusion criteria. The requested scope of this rapid review was to assess the psychometric properties of specified scale scores in individuals with chronic musculoskeletal pain (defined as at least 3 months duration). We were generous in our inclusion criteria. We did not specifically require the phrase "chronic musculoskeletal pain" be used. We included studies if the authors reported that participants had pain of at last 3 months duration or the authors described the participants as having chronic pain associated with a musculoskeletal condition (eg, osteoarthritis) even if duration was not described. We disagree that articles should have been included if they evaluated pain measures in patients with chronic conditions often associated with pain. Such individuals do not necessarily have chronic pain. From a clinical perspective, many patients with radiologically defined osteoarthritis do not have pain or only have acute or subacute pain and thus would fall outside the scope of this review. Nonetheless, we recognize that some may wish to extrapolate less reliable findings from individuals with acute or subacute pain or those with osteoarthritis without pain. We provide a discussion of findings from results of systematic reviews that assessed pain scale scores from studies that included these populations.

Question Text	Comment	Author Response
	be helpful to have a table of excluded studies along with the reasons for exclusion so readers of the report can better understand how criteria were applied. Without this information, I am wondering why the following studies were not included: • 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 Sep-Oct;20(5):309-18. • Elliott AM, Smith BH, Smith WC, Chambers WA. Changes in chronic pain severity over time: the Chronic Pain Grade as a valid measure. Pain. 2000 Dec 1;88(3):303-8. • Holm I, Friis A, Storheim K, Brox JI. Measuring self-reported functional status and pain in patients with chronic low back pain by postal questionnaires: a reliability study. Spine (Phila Pa 1976). 2003 Apr 15;28(8):828-33.	In the process of full text review, we exclude an article if it meets any one of our exclusion criteria – we do not document all the reasons it was not eligible. Therefore a table of excluded studies would not provide the level of requested detail. The suggested references were reviewed for eligibility. Keller is now included in the final report. Elliott did not meet all inclusion criteria. The study was designed to assess the Chronic Pain Grade as a measure useful in prospective studies of the general population and included patients with any chronic pain, including pain due to angina, arthritis, back pain, injury, women's problems, and unknown sources. Results were not stratified by pain type. Holm did not meet all inclusion criteria (not English version of scale of interest). The study assessed the Norwegian version of the ODI.
	Other comments: • I would not use "CMP" in the text because I think it's generally preferable to avoid unnecessary use of idiosyncratic abbreviations. (I have no objection to use in the tables, where space is limited).	We removed the CMP abbreviation. Thank you for the suggestion.
	• It would be helpful to provide a bit more descriptive information about each of the included measures. Many of these measures are currently described incorrectly as using "Likert" type items. Numeric rating items, such as those in the BPI and PEG, are not the same as Likert-type items.	We reviewed Table 1 (table of measures) and Supplemental Content Table 3 and updated them to reflect correct descriptions of the scales. Scales, including the BPI and PEG, have been corrected to indicate they are numeric rating scales. Of note, there are discrepancies among various sources of information we reviewed about the scales.
	• Reporting of pain medication use is commented on for most of the measures and in the table, but it is not clear to me how this information is relevant to this report. Given the purpose, it would have made more sense to describe patients' use of non-pharmacological therapies. (Please note: I am not suggesting adding info about reporting of non-pharm therapies. Rather, I think the text about reporting pain medications, such as "studies failed to report if patients were using pain medication," could be eliminated because it's not a relevant limitation for these types of studies.)	We thought that "use of pain medication" might provide an important descriptor of the study population but there are limitations, as the reviewer noted. We have deleted this information from the text and tables.
	• For measures with no included studies (e.g., DVPRS, KOOS), it would be helpful to provide information about why studies were excluded.	We appreciate the reviewer's point about the value of information on why no studies of the DVPRS and KOOS met inclusion criteria. We have added this information to the report (page 10).
	• Table 1: What is meant by "yes" and "request" in the public domain column? Some measures are copyrighted but available without charge. Some measures require payment. The relevant information here is the requirement for payment. If "public domain" is meant to mean the scale is available without charge, then the information is incorrect for several of the scales. I am confident that the GCPS and PROMIS measures are both available and free. BPI is not free. The SF-36 is copyrighted and some	As noted above, we updated Table 1. We replaced the "domain" column with "Restrictions on Use." The BPI and SF-36 have restrictions and may require payment. Some scales may also be obtained directly from the original author.

Question Text	Comment	Author Response
	versions are available free while others are only available with payment. I don't know about many of the other measures. • Page 5: Authors may wish to add publication prior to 2000 to the list of exclusion criteria, provide a brief justification (e.g., rapid review, focus on measures in current wide use) and briefly describe how they handled situations where the original measure paper was published prior to 2000. • Page 9: The BPI items are 0-10 numeric rating scales, not Likert scales. • Page 12: Numeric rating scale is a generic term not specific to pain numeric rating scales. • Page 17: Shouldn't the van Grootel study been excluded because it is a study of patients with orofacial pain? • Page 23: It would be helpful to comment on the fact that pain normally varies in intensity and would not be expected to remain static over days, weeks, or months. For most of the measures assessed, test-retest reliability doesn't make conceptual sense. • Page 29: The BPI scales have a total of 11 items (not 17). • Page 29: Shouldn't the Brazilian study (Gallasch 2007) be excluded because measures were not administered in English? • Page 30: I'm confused overall by the wording in this section. What does "all other scales we reviewed can be self-administered" mean? Does this mean they can't be administered any other way or that the preceding scales can't be self-administered? If so, why? Also, there is at least one error: the PEG is not designed to be administered by an interviewer; like the BPI from which it originated, it can be self-administered or administered by an interviewer. I would guess that almost all of the scales are commonly administered both by self-complete questionnaire and by interview. I suspect very few have been subject to rigorous evaluation of whether they perform similarly when self-administered or telephone/in-person interviewer-administered. Either way, I think it is outside of the scope of the review to determine all of the validated modes of administration for each scale. It would be helpful to know which scales require specific tools or modes for administration (e.g., computerized adaptive testing, visual aids). • Page 30: There are several errors in the availability section. GCPS and PROMIS are freely available without charge. I think the MPI is too. Also, availability of SF-36 and its pain subscale is complex. The original version is available from RAND for free, but there is a revised copyrighted commercial version that requires payment to use.	Although our literature search was limited to 2000 to the present, we included studies prior to 2000 that were identified in hand-searching of reference lists of eligible studies and systematic reviews as well as websites of individual pain scales. This is noted in the Methods section (page 6). We updated Supplemental Content Table 3 to show the BPI as a set of numeric rating scales. We reviewed Table 1 and Supplemental Content Table 3 and added "for pain' to the titles for the NRS and VAS We included TMJ references, as this type of orofacial pain is musculoskeletal in nature. We clarified this in the report (page 7). We agree that pain varies in intensity and is not expected to remain static over specific periods of time. We comment further on the limited conceptual value and applicability of test-retest reliability in the discussion (page 31). We updated Table 1 and Supplemental Content Table 3 showing the BPI as an 11 item scale. We reviewed Gallasch 2007. For non-US studies, we included the study if the authors did not specify the language used for the measures and Gallasch doesn't provide this information. We modified our exclusion criteria to reflect this. The section on "Mode of Administration" has been revised. We appreciate the reviewer's attention to the information about availability of measures. We updated Table 1 and Supplemental Content Table 3 as well as the text on Availability. We chose to indicate that there are restrictions on the SF-36.
	This report will serve an important role in informing a Pain Measurement WG deliberations regarding optimal measures for use in clinical and research settings. The process for establishing the parameters of the review, the enactment of a high fidelity review process, and an exceptionally clear report are important strengths. At the same time, the narrow scope of the review and the narrow	Thank you. We disagree that the scope and parameters are too narrow. Our report was based on decisions made jointly with our partners and within the parameters of a Rapid Review and the Topic Nomination. We previously described our concerns about including findings from studies of non-English language versions of scales and

Question Text	Comment	Author Response
	parameters for identifying published articles on this topic will likely limit the value of the report. In particular, the decision to "exclude trials of interventions for pain, unless assessment of psychometric properties of interest was noted in the abstract" seems to be a "fatal flaw." It seems intuitive that most published clinical trials would not report psychometric properties of the key measures in the Abstract, since this would not likely be the major focus of the report. It would also seem intuitive that published reports of the psychometric properties of the measures should have been considered, as well as systematic reviews of the psychometric properties of these measures. These decisions likely greatly limited the data pool upon which measures' quality could be evaluated. Other comments: Why only Medline? Results seem to be based on a binary determination about whether specific psychometric properties were reported in published studies, rather than the strength of the psychometric properties. Results are reported for CMP generally, without regard to specific CMP conditions. The failure to examine non-English language versions of the measures will greatly limit the value of the report beyond its use in informing VHA policy. Use of quality assessment (COSMIN) is a strength. Some of the data reported in Table 1 could be considered misleading. The pain severity and pain interference scales of the WHYMPI total only 12 items. The Pain Rating Index of the MPQ that assesses pain severity has only five response options. There are questions about the accuracy of Table 2. Kerns (1985) reported on the concurrent and criterion-related validity of the measure. Page 24 reports on internal consistency, but it reports on a range of alphas for the several scales of the WHYMPI, not just the pain severity and interference	recommendations for providing some information related to these studies. We also disagree that our decision to "exclude trials of interventions for pain, unless assessment of psychometric properties of interest was noted in the abstract" is a "fatal flaw." We reviewed search and triage strategies of relevant systematic reviews. It is extremely likely that these authors used similar strategies. For example, the systematic review by Gandek (2015) evaluating the WOMAC excluded nearly 2000 articles at the abstract level because they did not meet inclusion criteria (without further elaboration). Included articles described psychometric properties in the abstract and none of the included articles identified in prior systematic reviews or scale score websites reported information in the body of the manuscript without also describing in the abstract. A review of a subset of excluded articles confirmed these findings and support our rationale. We considered MEDLINE to be the most pertinent database. Rapid Reviews typically utilize a single electronic search engine that is likely to capture the most relevant information in an expedient fashion. As noted, in the Methods, we searched beyond MEDLINE to identify relevant evidence. Evaluating the quality of the statistical and other methodological approaches to psychometric assessments, and therefore the quality of their findings, was not set out as one of our goals for this systematic review. We report the CMP conditions present within the population for each study, and thus for the psychometric properties of interest assessed in that study. We attempted to synthesize results across CMP conditions as per the understood goals of this report. We have commented further on patterns in the CMP conditions of populations within which the most frequently studied measures were assessed. We have previously described our rationale for excluding non-English language versions of the measures. We modified the Methods section. Although the COSMIN checklist is an appropriate tool for the quality of studies of measurement properties, on further examination, the checklist (beyond identifying the appropriate measurement properties to evaluate) is extensive and not feasible to use in a Rapid Review. We reviewed Table 1 (table of measures) and Supplementary Content Table 3 to include the number of pain severity and number of pain interference items in each scale (as appropriate based on purpose of

Question Text	Comment	Author Response
	scales. Also on page 24, the section on concurrent and criterion-related validity does include a reference to Kerns et al (1985) but this publication is not listed among those that were reviewed earlier in the same section and the fact that these psychometric properties were evaluated is not acknowledged in Table 2. I'm not clear how MID or responsiveness is operationalized. It seems intuitive that if an RCT included one of these measures as an outcome, and if there is evidence of significant change over time, it should be considered as evidence or responsiveness. Similarly, if the study included a prespecified MID and included a "responder analysis" then this should be included.	scale) Thank you for noting this. We cross-checked all tables and text for accuracy. We appreciate the reviewer's point about the role of change over time in a measure used in an RCT. We also recognize that in such a situation, some forms of responsiveness assessment cannot separate the effect of an intervention from the ability of a measure to assess that effect. We considered the variety of approaches to responsiveness in assessed studies, and comment on this in the discussion (page 31). We have also attempted to clarify our approach to MID assessment, which focuses on whether studies developed an estimate of a minimum clinically important difference and/or minimum detectable change specific to a given measure. This question of primary MID development is distinct from questions about whether studies used, for example, prespecified MIDs and responder analyses as part of their approach to an RCT.
	This is overall a very well done review. The methods are clearly described and the conclusions are appropriate for the findings. I have one main concern, which is the time period examined, starting in 2000. I think this would be an ample look-back period. However, some of these measures are a bit older, and the early psychometric data may have been published before 2000. Even though the included articles from 2000 forward were scanned for references other relevant articles, this strategy could still miss relevant articles for measures that did not have any newer articles meeting inclusion criteria. If it is decided that the review will not search for articles prior to 2000, I think this limitation should be very clear and prominent, and the report should make it clear that the lack of finding psychometric data does not necessarily mean it isn't there in earlier years, or that this finding means the measure is not recommended. I have some familiarity with the KOOS and was very surprised to see that there were no relevant articles, since there certainly are papers on its psychometric properties in the literature. I am not sure if the lack of finding is because these manuscripts were published before 2000. Some additional minor comments: - Page 2, Lines 18-19: This seems to be somewhat in conflict with the first sentence of the paragraph - Page 26, lines 32-34: This does not seem to be a complete sentence.	Thank you. As noted in responses above, we have taken multiple other steps to enhance the literature retrieval process. We appreciate the point about the inherent limitation in any date restrictions, and have commented on this in the Limitations section (page 31). The reviewer does not provide article references that we may have missed. We have reviewed and included (or excluded) all suggested references provided by peer reviewers. Several articles reporting on the KOOS were included in our full text review (*eg,* Roos 2003, Ornetti 2008) but were not eligible because they used non-English versions of the KOOS. We revised the Conclusions paragraph in the Abstract. We corrected this sentence.
	This manuscript gives an overview of the properties of a number of pain-related measures in persons with chronic pain. Though it does provide some summarization of the literature, I have some concerns about the methods as well as the conclusions. My main concern is that relevant	Thank you for the feedback. We have modified the conclusion statements and address specific comments below.

Question Text	Comment	Author Response
	studies were likely to have been excluded due to the way the inclusion/exclusion criteria were specified, failure to assess the quality of included studies, and unclear synthesis methods. I also feel that the conclusions--which are basically "we can't make any conclusions" are rather superficial when there do appear to be some measures that are supported by more evidence/testing/validation than others. 1. It is not clear to me why studies that used non-English language versions of the scales were excluded. This is a big chunk of the literature. As the main conclusion is that there isn't enough evidence to know the properties of the scales, it is problematic to focus exclusively on English language versions of the scales when there is a lot of other data available. 2. I don't understand why studies of patients with chronic conditions typically associated with pain were excluded unless the study specified that the patient had CMP. Why else would these studies be using/assessing pain-related outcome measures? 3. It isn't clear to me what conditions were included. The "Key Question" section says "e.g. LBP, OA, and non-traumatic joint pain." What about things like chronic neck pain, fibromyalgia, tension HA's, shoulder pain, etc. I guess there may be some debate about whether FM and tension HA's are "musculoskeletal" but I would generally consider them in that category. Also RA and the inflammatory arthritis conditions seem to have been excluded. 4. If the focus is specifically on musculoskeletal pain that should be specified in the title--right now it talks about measures for pain in general. 5. The methods indicate that studies were excluded unless they specifically note that duration of pain was >3 months. But OA is almost by definition a chronic condition so I don't think that studies should have been excluded if they didn't specify duration of symptoms. 6. It is not clear why quality of studies was not assessed. This is not the same as the checklist on measurement properties that is mentioned in the Methods, which mainly seem to be about what kinds of properties should be evaluated to determine whether a measure if valid or not, not about internal validity of the studies themselves. The methods say that quality was not assessed because they did a "qualitative synthesis of findings" but I don't see that as a valid reason--we assess quality all the time when we do quantitative syntheses. I don't see how we are to assess the validity of the studies without some quality assessment. 7. The Data Synthesis/Rating the Body of Evidence sections really don't describe any methods. Saying that the methods were "qualitative" is not	1. As noted above, the review team decided that non-English language versions of the scales of interest could potentially produce different results due to variations in interpretation of descriptors of subjective ratings for pain intensity and interference. We provide references and descriptions that highlight the limitations in extrapolating findings from non-English language versions though we recognize that our stakeholders and other researchers/clinicians/policy makers may wish to make decisions based on studies of more broadly defined populations and study settings, in particular information from studies: 1) using non-English language versions of scales; 2) evaluating patients with musculoskeletal conditions often associated with chronic pain but not specifying the presence or duration of pain; 3) pain related to conditions outside of chronic musculoskeletal disorders (*eg,* headache, cancer). 2. As noted above, we clarified the study inclusion criteria. It was not sufficient for studies to include participants with a condition that is potentially *associated* with pain. This decision was based on the understanding that not all patients with a diagnosed condition such as radiologically defined osteoarthritis experience chronic pain. 3. We agree that fibromyalgia should have been an included condition and thank the reviewer for noting this. We reviewed our excluded studies and, as described in the Methods, did a separate search for studies of patients with fibromyalgia. We excluded rheumatoid arthritis (an inflammatory condition), headache, and orofacial pain (with the exception of temporomandibular pain, a musculoskeletal condition). 4. We modified the title to include musculoskeletal pain. 5. We included studies with duration of pain ≥3 months or pain described as “chronic” by study authors. We did not automatically include studies on osteoarthritis as it is possible to have the condition without chronic pain. For example, some excluded studies included participants with radiologically defined osteoarthritis per a series of imaging reviews, which does not necessarily address the presence or chronicity of pain. 6. We modified the Quality Assessment section of the report. We established inclusion criteria that would focus our review on studies that appropriately evaluated the psychometric properties of the pain measures. We did not go further into evaluating the quality of the articles since it is very difficult to evaluate the wide range of statistical approaches to assessing multiple psychometric attributes. A study

Question Text	Comment	Author Response
	sufficient--we do qualitative syntheses all the time and account for the same kinds of things (quality, inconsistency, directness, precision, etc) as we do for quantitative syntheses. I get very little sense of whether the findings are reliable--the results mostly read like a laundry list of results. Also, there are no pre-specified criteria for interpreting the findings--e.g. what would be considered decent test-retest reliability, responsiveness, etc? What would be necessary to establish a MID? 8. I think the conclusions are too quick to basically say that they can't support any of the measures. There are clearly some measures that have been validated/tested more than others. The conclusions should do a more nuanced job of highlighting those measures that are supported by better evidence. 9. It should be noted that the PEG scale is derived from the BPI (takes three items from the BPI). 10. There is also a lot of other overlap between scales--e.g. the NRS or a VAS is included in a number of outcome measures and similar items regarding function have been incorporated into a number of scales. I think that the overlap between scales warrants some discussion. If several items or scales has been validated how much additional validation is required when they are incorporated into another scale?	that is good for some aspects could be poor for others. The extensive list of criteria set forth in the COSMIN checklist speaks to this difficulty in evaluating a large list of measures for multiple quality criteria and clinical contexts. 7. We modified these sections of the report. Regarding criteria for interpreting findings, we describe some of the difficulties with establishing such across-the-board criteria in the Discussion. We provide particular attention to methods of developing and interpreting MID (the primary outcome, as approved by Topic Nominators). As previously noted, evaluating the quality of the statistical and other methodological approaches to psychometric assessments, and therefore the quality of their findings, was not set out as one of our goals in this systematic review. 8. We appreciate the reviewer's point, and have attempted to highlight measures that have been more frequently assessed with respect to psychometric properties of interest. 9. Thank you for the clarification. We changed the wording on Supplemental Content Table 3 from "based on" to "derived from." 10. We appreciate the reviewer's point, and have commented further on the conceptual overlap between some scales. We also comment on the wide variation in content among NRS-based and VAS-based approaches. The question of how much additional validation is required when items of an existing scale are adapted and incorporated into a new scale could inspire interesting debate within the field of psychometric methodology. To our knowledge, there are no concrete criteria addressing this question that we could operationalize in this review.
	Boonstra, Anne M. et al. "Cut-Off Points for Mild, Moderate, and Severe Pain on the Numeric Rating Scale for Pain in Patients with Chronic Musculoskeletal Pain: Variability and Influence of Sex and Catastrophizing." Frontiers in Psychology 7 (2016): 1466. PMC. Web. 8 May 2017.	The suggested reference was reviewed for eligibility and did not meet all inclusion criteria (English language versions of scales). The scales were administered in Dutch.

SUPPLEMENTAL TABLE 3. CHARACTERISTICS OF INCLUDED PAIN MEASUREMENT SCALES

Scale Reference	Measure Properties				Feasibility			
	Year Developed	Developed for Specific Conditions (write in)	Pain Severity/ Intensity *or* Functioning/ Interference	Scoring (write in)	Number of Items	Scale Description	Restrictions on use: Yes, No	Reading Level (write in)
Brief Pain Inventory (BPI)[1]	1983	Cancer pain	Pain intensity Interference (physical functioning, work mood, walking, social activity, relations with others, and sleep)	11-point numeric rating scale of 0-10, corresponding to: 0=no pain, no interference, to 10=pain as bad as you can imagine, complete interference Diagram also provided for respondents to indicate where pain is felt Mean of pain intensity and interference scores indexed separately	11 total (4 severity, 7 interference)	Range: none=0, mild= 1-3, moderate=4-6, severe=7-10 Direction: Higher indicates worse	Scale available for purchase with price dependent on use	NR
Defense & Veterans Pain Rating Scale[2]	2010	Pain among military and Veterans Designed to enhance NRS with visual cues and word descriptors to anchor pain	Pain intensity Interference (general activity, sleep, mood, stress)	11-point numeric rating scale of 0-10, corresponding to: 0=no pain, no interference, no affect, to 10=pain as bad as can be, completely interferes, completely affects	5 total (1 severity, 4 interference)	Range: Green (0-4)=mild pain or interference Yellow (5-6)= moderate pain or interference Red (7-10)= severe pain or interference Direction: Higher indicates worse	Free use of the scale is permitted without revisions or alterations	9th grade reading level

Scale Reference	Measure Properties				Feasibility			
	Year Developed	Developed for Specific Conditions (write in)	Pain Severity/ Intensity *or* Functioning/ Interference	Scoring (write in)	Number of Items	Scale Description	Restrictions on use: Yes, No	Reading Level (write in)
Graded Chronic Pain Scale[3,4]	1992	Chronic pain conditions including musculoskeletal pain and LBP	Pain Intensity Interference (disability)	11-point Likert-type scale of 0-10, corresponding to: 0=no pain to 10=pain as bad as can be. Mean intensity ratings multiplied by 10 calculated for 2 subscales ranging from 0-100 and 1 subscale ranging from 0-3 points Subscale scores for pain intensity and disability are combined to calculate chronic pain grade. Patients are then divided into 5 hierarchical categories: grade 0 (no pain) and 5 (high disability and severely limiting)	7 total	Range: Pain intensity= 0-100 Disability score= 0-100 (0-3 points) Disability pts (points from disability days + disability score) =0-6 *Disability days 0-6 days=0 points 7-14 days=1 points 15-30 days=2 points 31+ days=3 points *Disability score 0-29=0 points 30-49=1 points 50-69=2 points 70+=3 points Direction: Higher indicates worse	Free version of scale is available from original reference or directly from author	Basic
Hip Osteoarthritis Outcomes Scale (HOOS)[5]	2002	Hip disability with or without OA Extension of WOMAC scale	Pain intensity Interference (physical functioning)	5-pt Likert-type scale of 0-4, corresponding to: 0=no problems to 4=extreme problems All subscale scores transformed to a 0-100 scale with zero representing no hip problems and 100 representing extreme hip problems	40 total with 5 subscales pain, symptoms, daily living limitations, sport and recreation limitations, hip-related quality of life	Range: 0-100 Direction: Higher indicates worse	Free version of scale available online	NR

Scale Reference	Measure Properties				Feasibility			
	Year Developed	Developed for Specific Conditions (write in)	Pain Severity/ Intensity *or* Functioning/ Interference	Scoring (write in)	Number of Items	Scale Description	Restrictions on use: Yes, No	Reading Level (write in)
Knee Osteoarthritis Outcomes Scale (KOOS)[6]	1998	Knee injury or OA Extension of the WOMAC pain scale	Pain intensity Interference (physical functioning)	5-point Likert-type scale of 0-4, corresponding to: 0=no problems to 4=extreme problems All subscale scored as sum of items answered Scores are then transformed to a 0-100 scale with zero representing extreme knee problems and 100 representing no knee problems	42 total with 5 subscales: pain (9 items), symptoms (7 items), daily living limitations (17 items), sport and recreation limitations (5 items), knee-related quality of life (4 items)	Range: 0-100 Direction: Higher indicates better	Free version of scale available online	NR
McGill Pain Questionnaire[7,8]	1970	General chronic pain	Pain Intensity and quality in multiple domains (*eg,* sensory, affective, evaluative)	Three classes of rank order-type words and a 5-point numeric rating scale MPQ scored by counting number of words selected to obtain a Number of Words Chosen score (0-20). PRI scores (0-78) based on rank order of words in each subclass Rank scores are summed in each subclass as well as overall Normative scored range from 24-50% of maximum score	78 total with 20 subscales (PRI) sensory=42, affective=14, evaluative=5, miscellaneous=17; 6 additional items (5 point score range) for the present pain intensity scale (PPI)	Range: Number of Words Chosen= 0-20 PRI= 0-78 PPI= 0-6 Direction: Higher indicates worse *No established cut points	Free version of scale available from author	Words may be defined by adminis-trator

Scale Reference	Measure Properties				Feasibility			
	Year Developed	Developed for Specific Conditions (write in)	Pain Severity/ Intensity *or* Functioning/ Interference	Scoring (write in)	Number of Items	Scale Description	Restrictions on use: Yes, No	Reading Level (write in)
Multidimension-al Pain Inventory[7,9] (also known as the West Haven-Yale Multidimensional Pain Inventory [WHYMPI])	1985	Chronic pain including LBP and temporo-mandibular disorders	Pain intensity Interference (daily activities including vocational, social, and familial functioning)	7-point numeric rating scale of 0-6, corresponding to: 0=none, not at all, extremely low, never, to 6=extreme, very intense, very often Subscale scores are derived by from sum of individual terms in subscale divided buy number of items in subscale To calculate total score divide by the number of items	52 total with 3 parts interference=9, support=3, pain severity=3, life-control=2, affective distress=3, negative responses=4, solicitous responses=6, distracting responses=4, household chores=5, outdoor work=5, activities away from home=4, social activities=4	Range: Pain experience= (0-120) Significant others' responses to communication of pain=(0-84) Participation in common daily activities= (0-108) Direction: Higher indicates worse	Free version of scale available from author	Words may be defined by adminis-trator
Numeric Rating Scale for Pain (NRS)[3,8]	NR	General chronic pain	Pain intensity Pain interference	11-point numeric rating scale of 0-10, corresponding to: 0=none to 10=severe Horizontal line commonly used	1 total	Range: none=0, mild= 1-3, moderate=4-6, severe=7-10 Direction: Higher indicates worse	Free version of scale available online	Basic
Oswestry Disability Index/Oswestry Low Back Pain Disability Questionnaire (ODI/ODQ)[10]	1980	Disability from acute and chronic LBP	Pain intensity (need for pain medications) Interference (physical functioning, disability)	6-point ordinal scale of 0-5, corresponding to: 0=no pain, no interference/disability, to 5=worst scenario of pain, interference/disability Scoring for each item increases from 0-5 Missing values omitted. Sum of scores divided by total possible scores to obtain percentage	10 total with 2 possible subscales; pain or need for pain medication=1 item, interference on daily activities=9 items	Range: Minimal disability=0-20 Moderate disability=20-40 Severe disability=40-60 Housebound-60-80 Bedbound=80-100 Direction: Higher indicates worse (disability)	Free use of scale permitted for non-funded academic research and individual clinical practice	NR

Scale Reference	Measure Properties				Feasibility			
	Year Developed	Developed for Specific Conditions (write in)	Pain Severity/ Intensity *or* Functioning/ Interference	Scoring (write in)	Number of Items	Scale Description	Restrictions on use: Yes, No	Reading Level (write in)
Patient Global Impression of Change[11]	NR	NR	NR	7-point categorical scale of 1-7, which corresponds to: 1=no change in condition to 7= a great deal better	1 total	Range: 1-7 Direction: Higher indicates better	Free version of scale available online	N/A
PEG[12]	2008	Chronic pain in primary care Derived from the BPI	Pain intensity Interference (physical functioning)	11-point numeric rating scale of 0-10, corresponding to: 0=no pain, no interference, to 10=pain as bad as you can imagine, completely interferes The PEG is scored by averaging the 3 individual item scores	3 total	Range: 0-10 Direction: Higher indicates worse	Free use of the scale is permitted	NR
PROMIS Pain Interference (PROMIS-PI)[13,14]	2004	General chronic pain conditions	Interference (physical functioning)	5-point numeric rating scale of 1-5, corresponding to: 1=not at all, never, to 5=very much, always, every few hours Sum response scores for questions that were answered. Multiply sum by total number of items in form then divide by number of items answered	41 total 4a, 6a, 6b, and 8a item short version SFs often used	Range:4a=4-20 6a=6-30 6b=6-30 8a=8-40 Direction: Higher indicates worse	Free use of the scale is permitted after registration with assessment center and endorsing terms and conditions of use	NR
Roland Morris Disability Questionnaire (RMDQ)[15]	1983	Disability from LBP	Interference (physical functioning, disability)	1 point for each item completed	24 total	Range: 0=no disability to 24=severe disability Direction: Higher indicates worse (disability)	Free version of scale available and in the public domain	Basic

Scale Reference	Measure Properties				Feasibility			
	Year Developed	Developed for Specific Conditions (write in)	Pain Severity/ Intensity *or* Functioning/ Interference	Scoring (write in)	Number of Items	Scale Description	Restrictions on use: Yes, No	Reading Level (write in)
SF-36 Bodily Pain Scale (SF-36 BPS)[3,16]	1996	Overall health status in ages ≥14	Pain intensity Interference (daily activities)	6-point pain severity rating where 1=none and 6=very severe; 5-point pain interference rating where 1=not at all and 5=extremely Responses transformed to a 0-100 point scale.	2 total	Range: 0-100 Direction: Higher indicates more favorable health state.	Scale available for purchase with price dependent on use	NR
Visual Analogue Scale for Pain (VAS)[7,17]	1952	Rheumatic diseases	Pain intensity Interference (disability)	One vertical line (usually 10cm or 100 mm) in length anchored with verbal descriptors of "no pain" to "pain as bad as it could be". Perpendicular lines placed at point that best indicates pain. Metric ruler placed along line to indicate score in mm or cm	1 total	Range: 0-10 cm or 0-100 mm Direction: Higher indicates worse Scores below 4 cm or 20 mm considered desirable for chronic pain management	Free version of scale available and in the public domain	NR
Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[18]	1982	OA (knee and hip)	Pain intensity Interference (physical functioning)	5-point Likert-type scale of 0-4, corresponding to: 0=none to 4=extreme 100mm Visual Analog version uses anchors of no pain/ stiffness/ difficulty and extreme pain/ stiffness/ difficulty	24 total with 3 subscales pain=5, interference (functioning)=17, stiffness=2	Range: Pain 0-20 Function 0-68 Stiffness 0-8 Direction: Higher indicates worse	Scale available for purchase with price dependent on use	NR
Wong Faces Scale[19]	1985	Pain among children	Pain intensity	6-point numeric rating scale of 0-10 (increasing by 2), which corresponding to: 0=no pain to 10=hurts worst Person chooses the face that best describes their pain	1 total	Range: No pain=0 hurts little bit=2 hurts little more=4 hurts even more=6 hurts whole lot=8 hurts worst=10 Direction: Higher indicates worse pain	Scale available for purchase with price dependent on use	NR

OA=Osteoarthritis; NR=not reported, PPI=present pain intensity, PRI=Pain rating index, LBP= low back pain, CAT= computer adaptive test, N/A= not applicable

SUPPLEMENTAL TABLE 4. STUDY CHARACTERISTICS

Study (name and year)/ Location/ Funding	Scale of Interest/ Others	Mode of Administration	Setting[a]	Condition/ Study Inclusion/ Exclusion Criteria	Baseline Pain Characteristics	Demographics
Anagnostis 2004[20] Location: United States Funding: Government	**ODI**	SAQ, written	Community treatment clinic	Condition: CDMD (current) Inclusion: Enrolled in chronic pain management course; ≥ 4 months partial or total disability since work related injury; ≥ 1 injury related to spine or extremities, failed response to primary or secondary non-operative care or surgery; severe functional limitations; English or Spanish speaking Exclusion: NR	Baseline pain score(s): NR Average intensity: NR	N=230 Age (mean, SD): 43.3, δ 9.4 Women (%): 53 Race/Ethnicity (%): White: 59.7 African-Amer./Black: 29.2 Hispanic: 11 Other: 0.1
Askew 2016[21] Location: United States Funding: Government	**PROMIS-PI**	SAQ, written	Spine center, local clinics	Condition: LBP Inclusion: Receiving, or about to receive; a spinal injection Exclusion: NR	Baseline pain score(s): NRS= 78% scored ≥8, range 0-10 Average intensity: NR	N=218 Age: 62% ≥ 50 years Women (%): 56 Race/Ethnicity (%): White: 84 African-Amer./Black: 4.1 Hispanic: 1.3 Other: 10.6
Burnham 2012[22] Location: Canada Funding: NS	**MPQ** **ODI**	SAQ, written	Chronic pain management clinic	Condition: Spine pain Inclusion: Attending a chronic pain management clinic; received a lumbopelvic spinal intervention Exclusion: NR	Baseline pain score(s): NR Average intensity: NR	N=60 Age (mean, SD): 60, δ 12.4 Women (%): 67 Race/Ethnicity (%):NR
Changulani 2009[23] Location:	**ODI** **VAS**	SAQ, written	Outpatient clinic	Condition: LBP Inclusion: Undergoing caudal epidural steroid injections for	Baseline pain score(s): ODI Spinal stenosis= 48 (δ15);	N=107 Age (mean): 58 Women (%): 58 Race/Ethnicity (%): NR

Study (name and year)/ Location/ Funding	**Scale of Interest/** Others	**Mode of Administration**	**Setting[a]**	**Condition/ Study Inclusion/ Exclusion Criteria**	**Baseline Pain Characteristics**	**Demographics**
United Kingdom Funding: NS				lumbosacral radicular pain with symptoms persisting for more than 4 weeks; unrelieved by analgesia and physiotherapy Exclusion: NR	Disc prolapse=50 (δ16); Spondylolisthesis=41 (δ15) Average intensity: NR Type of pain (%): Spinal stenosis=59 Disc prolapse=36 Spondylolisthesis=5	
Chansirinukor 2005[24] Location: Australia Funding: None	**RMDQ**	SAQ, written	Physical therapy clinic	Condition: LBP Inclusion: Work-related pain, at least 2 complete Functional Rating Indexes and RMDQs Exclusion: NR	Baseline pain score(s): RMDQ= 57.2 (δ23.7) Average intensity: NR Type of pain (%): LBP=78.3	N=143 Age (mean, SD): 37.9, δ 9.8 Women (%): 26.4 Race/Ethnicity (%):NR
Chien 2013[25] Location: Australia Funding: Academic	**BPI**	SAQ, written	Pain clinic	Condition: General musculoskeletal pain Inclusion: Age ≥18 years; nonmalignant pain Exclusion: Cancer-related pain	Baseline pain score(s): BPI (S) =6.0 (δ 1.6); BPI (I) =5.9 (δ 1.9) Average intensity: Moderate	N=254 Age (mean): 51 Women (%): 50 Race/Ethnicity (%): NR
Cook 2008[26] Location: United States Funding: Government	**RMDQ** (24-, modified, 18-, 12-, and 11- item)	SAQ, written and CATs	NR	Condition: LBP Inclusion: Study 1 (Discogenic study) participants had 1- or 2-level disc degeneration Study 2 (Seattle Lumbar Imaging Project) participants randomly assigned to rapid	Baseline pain score(s): NR Average intensity: NR	*Data combined from 2 studies N=875 Age (mean, range): 47, 18-93 Women (%):NR Race/Ethnicity (%): White: 85

Study (name and year)/ Location/ Funding	**Scale of Interest/** Others	**Mode of Administration**	**Setting[a]**	**Condition/ Study Inclusion/ Exclusion Criteria**	**Baseline Pain Characteristics**	**Demographics**
				magnetic resonance imaging or standard radiographs Exclusion: NR		African-Amer./Black: 9 Hispanic: 3 Asian: 2 Other: 1
de Vet 2007[27] Location: Netherlands Funding: NS	**NRS**	SAQ, written	Physio-therapy clinics	Condition: LBP Inclusion: Referred for physiotherapy Exclusion: NR	Baseline pain score(s): NR Average intensity: NR	N=438 Age (mean, range): NR Women (%): NR Race/Ethnicity (%): NR
Deyo 2016[28] Location: United States Funding: Government	**PROMIS-PI SF**	SAQ, Telephone interview	Primary care clinics	Condition: General musculoskeletal pain Inclusion: Age ≥55 years; ≥ 2 visits for musculoskeletal pain; moderate pain (≥ 5 points on 10-point pain scale); no opioid use for ≥ 1 month; telephone access; no cognitive impairments Exclusion: Adverse reaction to opioids; life expectancy <2 years	Baseline pain score(s): NR Average intensity: ≥5, 10-point scale Type of pain (%): Back=30.8 Neck=7.5 Joint=14.1 Arthritis=15.6 Other=31.8	N=198 Age (mean, SD): 66.5, δ 8.2 Women (%): 62.1 Race/Ethnicity (%): White: 92.3 Hispanic: 3.6 Other:4.1
Driban 2015[29] Location: United States Funding: Government	**PROMIS-PI** **SF-36 BPS** **WOMAC**	SAQ, written	University hospital	Condition: Pain from OA of the knee Inclusion: Participation in RCT (comparison of Tai Chi and physical therapy); age ≥40 years, WOMAC pain subscale score (100 mm visual analog scales) >40 on at least 1 out of 5 questions; fulfillment of the American College of Rheumatology criteria for knee	Baseline pain score(s): PROMIS-PI= 58 (δ 7.0); SF-36 BPS= 47.5 (δ 18.6); WOMAC= 254 (δ 98.6) Average intensity: NR	N=204 Age (mean): 60.2 Women (%): 70 Race/Ethnicity (%): White: 52.7 African-Amer./Black: 35.5 Other: 11.8

Study (name and year)/ Location/ Funding	Scale of Interest/ Others	Mode of Administration	Setting[a]	Condition/ Study Inclusion/ Exclusion Criteria	Baseline Pain Characteristics	Demographics
				osteoarthritis; radiographic evidence of knee osteoarthritis; confirmation of knee pain, discomfort; or disability by clinical examination Exclusion: Experience with physical therapy in past year, Tai Chi training/ use of alternative medicine; serious medical conditions limiting ability to participate, intraarticular steroid injections or replacement surgery on the affected knee in the last 3 months; or a Mini-Mental examination score <24		
Fisher 1997[30] Location: United Kingdom Funding: NS	**MPQ** **ODI**	SAQ, written	Clinical Psychology Department, outpatient clinic	Condition: LBP Inclusion: Undergoing, or about to undergo, a back pain rehabilitation program Exclusion: NR	Baseline pain score(s): ODI= 54.5 (δ12.3); MPQ =2.8 (δ1.1) Average intensity: NR Type of pain (%): Back=87 Leg or neck=13	N=54 Age (mean, range): 41, 20-62 Women (%): 63 Race/Ethnicity (%): NR
Gallasch 2007[31] Location: Brazil Funding: Government	**Faces** **NRS** **VAS**	SAQ, written	University health center	Condition: General musculoskeletal pain Inclusion: Physiotherapy treatment due to musculoskeletal symptoms, age 18 to 70 years; education no more than middle school level Exclusion: Illiterate	Baseline pain score(s): NR Average intensity: NR Type of pain (%): OA=19 Tendonitis=16 Back=13	N=32 Age (mean, range): 51, 33-69 Women (%): NR Race/Ethnicity (%): NR

Study (name and year)/ Location/ Funding	**Scale of Interest/** Others	**Mode of Administration**	**Setting[a]**	**Condition/ Study Inclusion/ Exclusion Criteria**	**Baseline Pain Characteristics**	**Demographics**
Gentelle-Bonnassies 2000[32] Location: France Funding: NS	**VAS** **WOMAC**	SAQ, written and mail	Hospital	Condition: Pain from OA of the knee Inclusion: OA fulfilling the criteria of the American College of Rheumatology; primary or secondary OA (osteonecrosis; chondro-calcinosis); involvement of the medial tibiofemoral; the lateral tibiofemoral, or the patellofemoral compartment of the knee joint; active disease (pain and disability) justifying joint lavage Exclusion: Serious chronic disease; intra-articular procedures (arthroscopy or surgery) performed ≤ 2 years or osteotomy performed ≤ 3 years; prescription of intra-articular injections ≤ 1 month before entry	Baseline pain score(s): VAS=57 (δ22) (pain after activity) Average intensity: NR	N=80 Age (mean, SD): 62, δ 12 Women (%): 70 Race/Ethnicity (%): NR
Godil 2015[33] Location: United States Funding: Industry	**NRS - neck pain** **NRS - arm pain**	Telephone interview	Medical center	Condition: Neck and radicular arm pain Inclusion: Age 18-70 years; undergoing anterior cervical discectomy and fusion for neck and radicular arm pain; radiological evidence of cervical nerve root impingement from herniated disc or osteophyte Exclusion: Myelopathic symptoms; previous cervical spine surgery	Baseline pain score(s): NRS-neck pain=6.3 (δ2.6); NRS-arm pain= 5.5 (δ3) Average intensity: NR	N=88 Age (mean, SD): 52.3, δ 10.7 Women (%): 44 Race/Ethnicity (%): NR

Study (name and year)/ Location/ Funding	**Scale of Interest/** Others	**Mode of Administration**	**Setting[a]**	**Condition/ Study Inclusion/ Exclusion Criteria**	**Baseline Pain Characteristics**	**Demographics**
Gronblad 1993[34] Location: Finland Funding: Foundation	**ODQ** **VAS**	SAQ	Tertiary care center	Condition: LBP Inclusion: With or without radiation to legs Exclusion: Pain due to underlying disease, psychiatric disease requiring continuous medication	Baseline pain score(s): NR Average intensity: Among subset VAS=54.1 (δ19.48)	N=94 Age (mean): 42.7 Women (%): 51 Race/Ethnicity (%): NR N=20 (re-test) Age (mean, SD): 42.3 Women (%): 55 Race/Ethnicity (%): NR
Hicks 2009[35] Location: USA Funding: Government	**ODI** **SF-36 BPS**	SAQ, mail	Retirement communities	Condition: LBP requiring activity modification Inclusion: Age ≥ 62 years, living independently Exclusion: NR	Baseline pain score(s): ODI=29.4 (δ16.6) Average intensity: NR	N=107 (validity) Age (mean): 80 Women (%): 72 Race/Ethnicity (%): White: 100 N=56 (re-test) Age (mean, range): 79 Women (%): 71 Race/Ethnicity (%): White: 100
Jensen 2012[36] Location: United States Funding: Industry	**VAS**	SAQ, written	Clinic	Condition: LBP Inclusion: Participation in an RCT (comparison of Etoricoxib and placebo); age 18-75 years; pain for majority previous month; taking NSAID or acetaminophen for 24 of previous 30 days; pain met Quebec Task Force criteria for spinal disorders (class 1 or 2); no surgery for LBP in past 6 months; no symptomatic depression or drug/alcohol abuse in past 5 years; no opioids > 4 days/month; no corticosteroid injections within 3	Baseline pain score(s): VAS= 76.7; RMDQ= 14.7 Average intensity: NR	N=639 Age (mean): 52.4 Women (%): 61.5 Race/Ethnicity (%): White: 90.1 African-Amer./Black: 5.1 Asian: 0.6 Other: 4.2

Study (name and year)/ Location/ Funding	**Scale of Interest/** Others	**Mode of Administration**	**Setting[a]**	**Condition/ Study Inclusion/ Exclusion Criteria**	**Baseline Pain Characteristics**	**Demographics**
				months Exclusion: NR		
Kamper 2015[37] Location: Australia Funding: Government, Industry	**NRS-24 hours** **NRS-week** **SF-36 BPS**	SAQ (See Stewart 2007)	Clinic (See Stewart 2007)	Condition: Whiplash associated disorders (neck pain) Inclusion: Participation in RCT (exercise therapy for chronic whiplash); pain from car accident; age 18-65 years; English speaking Exclusion: Cervical fractures or dislocations; serious spinal pathology; serious psychiatric illness	Baseline pain score(s): NR Average intensity: NR	N=280 Age (mean): 43.5 Women (%): 65 Race/Ethnicity (%): NR
Kean 2016[38] Location: United States **(Enrolled Veterans)** Funding: Government	**BPI** **PEG** **PROMIS-PI** **SF-36 BPS**	IAQ	Primary care clinic	Condition: Musculoskeletal pain Inclusion: Participation in RCT (effectiveness of collaborative telecare management for moderate to severe and persistent musculoskeletal pain); Veteran; age 18-65 years; receiving care at a VAMC; persistent pain despite trying ≥ 1 analgesic medication; other non-musculoskeletal pain; English speaker; pain of moderate severity Exclusion: Inflammatory arthritis; pending pain-related disability claim; cognitive impairment; psychoses; actively suicidal; current illicit drug use; life expectancy < 12 months	Baseline pain score(s): BPI= 5.3 (δ 1.8); SF-36 BPS= 34.8 (δ 16.8); PROMIS-PI= 22.1 (δ 8.8) Average intensity: Moderate	N= 244 Age (mean, range): 55, 28-65 Women (%): 17 Race/Ethnicity (%): White: 77 African-Amer./Black: 19 Other: 4
Keller 2004[39]	**BPI-SF**	SAQ	Primary care	Condition: LBP	Baseline pain	N=131

Study (name and year)/ Location/ Funding	**Scale of Interest/** Others	**Mode of Administration**	**Setting[a]**	**Condition/ Study Inclusion/ Exclusion Criteria**	**Baseline Pain Characteristics**	**Demographics**
Location: United States Funding: Government	**GCPS** **SF-36 BP** **RMDQ**		clinic	Inclusion: Age 18-80 years, not permanently disabled, at least 8th grade reading level, prescribed change of therapy requiring follow-up visit Exclusion: NR	score(s): NR Average intensity: NR	Age (mean): 46.5 Women (%): NR Race/Ethnicity (%): NR
Kerns 1985[9] Location: United States **(Enrolled Veterans)** Funding: Government	**WHYMPI (MPI)** **MPQ**	SAQ	Pain clinic	Condition: Chronic pain Inclusion: consecutive referrals to pain management program at 2 VAMCs Exclusion: NR	Baseline pain score(s): NR Average intensity: NR Type of pain (%): Back=36	N=120 (test-retest reliability for n=60 from one site) Age (mean, SD): 51, δ 14.5 Women (%): 18.5 Race/Ethnicity (%): NR
Krebs 2010[40] Location: United States **(Enrolled Veterans)** Funding: Government	**BPI** **GCPS** **PEG** **RMDQ** **SF-36 BPS**	SAQ, written See Scamp papers[b]	See Scamp papers	Condition: Back, hip, or knee pain Inclusion: Participation in SCAMP study[b]; Veteran; primary care patients, receiving care at a VAMC; persistent pain of at least moderate severity [BPI≥5]) (See SCAMP Study papers) Exclusion: NR	Baseline pain score(s): BPI (S)= 5.7; BPI (I)= 5.8; PEG= 6.0; GCPS= 68.3; RMDQ=14.8; SF-36 BPS=35.3 Average intensity: NR Type of pain (%): Back=55 Hip or knee=45	N=427 Age (mean): 59 Women (%): 53.4 Race/Ethnicity (%): White: 58 African-Amer./Black: 38 Other: 4
Krebs 2009[12] Location: United States **(Enrolled Veterans)** Funding: Government	**BPI** **GCPS** **PEG** **PGIC** **RMDQ**	SAQ, written See Scamp papers[b]	University and VA affiliated clinics	Condition: Back, hip, or knee pain Inclusion: Participation in SCAMP study[b]; Veteran; primary care patients, receiving care at a VAMC; persistent pain of at least moderate severity [BPI≥5])	Baseline pain score(s): (Mean) NRS = 6.1 (δ1.9), 0-10 scale Average intensity: NR	N=500 Age (mean): 59 Women (%): 52 Race/Ethnicity (%): White: 58 African-Amer./Black: 38 Other: 4

Study (name and year)/ Location/ Funding	**Scale of Interest/** Others	**Mode of Administration**	**Setting[a]**	**Condition/ Study Inclusion/ Exclusion Criteria**	**Baseline Pain Characteristics**	**Demographics**
	SF-36 BPS			(See SCAMP Study papers) Exclusion: NR		
Krebs 2007[41] Location: United States Funding: Foundation, Government	**NRS**	IAQ	Hospital	Condition: General musculoskeletal pain (including extremities, back, and neck) Inclusion: Adults presenting to general medicine clinic for a return visit Excluded: Non-English speaking; patients chosen by physicians	Baseline pain score(s): NRS=6.0 (among those with NRS ≥1), 0-10 scale Average intensity: NR Type of pain (%): Lower extremity= 21 Back or neck= 18 Upper extremity= 8 No pain= 28	N=275 Age (mean): 59 Women (%): 59 Race/Ethnicity (%): White: 70 African-Amer./Black: 24 Other: 6
Lovejoy 2012[42] Location: United States **(Enrolled Veterans)** Funding: Government, Industry	**MPQ-2 SF** **MPQ** **MPI**	SAQ, written	Unclear	Condition: LBP, neck or joint pain Inclusion: Veterans; age ≥18 years; English speaking; ≥1 pain diagnosis in medical record; reported current symptoms of (or receiving treatment for) chronic pain, previous tests for hepatitis C Exclusion: Age >70 years; current unstable psychiatric disorder; pending litigation or disability compensation for pain; advanced liver disease	Baseline pain score: MPQ-2 SF= 3.22 (δ2.36), range 0-9.82 Average intensity: NR Type of pain (%): Neck or joint=76 Back=59	N=186 Age (mean, SD): 54.4, δ 7.7 Women (%): 7.5 Race/Ethnicity (%): White: 75.3 Other: 14.7
Lund 2005[43] Location: Sweden Funding:	**VAS**	SAQ, written	Unclear	Condition: Idiopathic musculoskeletal pain Inclusion: Recruited from rehabilitation medicine clinic; previously classified as	Baseline pain score(s): (Median) VAS=59, range12-96 Average intensity: NR	N=30 Age (mean, SD): 42.8, δ 10.6 Women (%): 43 Race/Ethnicity (%): NR

Study (name and year)/ Location/ Funding	**Scale of Interest/** Others	**Mode of Administration**	**Setting[a]**	**Condition/ Study Inclusion/ Exclusion Criteria**	**Baseline Pain Characteristics**	**Demographics**
NS				chronic/idiopathic pain by physician Exclusion: NR		
Macedo 2011[44] Location: Australia Funding: Government	**RMDQ** (24-, 18[c,d]-, and 11- item)	SAQ, written	Unclear	Condition: LBP Inclusion: Patients with or without leg pain, age 18-80 years Exclusion: Previous spinal surgery; specific pathology; contraindication to exercise; insufficient English ability to complete questionnaires	Baseline pain score(s): (Mean) RMDQ 24=12.8 (δ 5.1);18[c] = 10.6 (δ 4.6); 18[d] = 10.8 (δ 4.4); 11= 7.3 (δ 2.9) Average intensity: NR	N=461 Age (mean, SD): 52.5, δ 14 Women (%): 61 Race/Ethnicity (%): NR
Maughan 2010[45] Location: United Kingdom Funding: NS	**NRS** **ODI-2** **RMDQ**	SAQ, written	Pain manage-ment back class	Condition: LBP Inclusion: Age ≥ 18 years, not undergoing treatment for pain, sufficient level of spoken and written English Exclusion: Spinal surgery in past 12 months, unstable neurological symptoms, pregnancy	Baseline pain score(s): NRS= 5 (δ2.6), RMDQ= 11(δ6.1), ODI= 29 (δ20) Average intensity: NR	N=48 Age (mean): 52 Women (%): 67 Race/Ethnicity (%): NR
Merriwether 2016[46] Location: United States Funding: Government	**PROMIS- PI-SF**	SAQ	University outpatient clinics	Condition: Fibromyalgia Inclusion: Women, ages 20-67 years, English speaking, stable medical treatment regime Exclusion: Prior transcutaneous electrical nerve stimulation use in last 5 years, pain intensity less than 4 out of 10 on the NRS	Baseline pain score: NR Average intensity: NR	N=106 Age (mean): 49.1 Women (%): 100 Race/Ethnicity (%): White: 96 Other: 4

Study (name and year)/ Location/ Funding	**Scale of Interest/** Others	**Mode of Administration**	**Setting[a]**	**Condition/ Study Inclusion/ Exclusion Criteria**	**Baseline Pain Characteristics**	**Demographics**
Mikail 1993[47] Location: Canada Funding: Foundation	**MPI** **ODI**	SAQ	Pain clinic	Condition: Chronic pain Inclusion: Patients seen at Chronic Pain clinic; diagnosis of chronic pain by physiatrist, psychologist, and physiotherapist Exclusion: NR	Baseline pain score: NR Average intensity: NR Type of pain (%): Neck=6, Back=43, Extremities=18, Multiple=25, Other=8	N=315 Age (mean): 43.5 Women (%): 53 Race/Ethnicity (%): NR
Nilsdotter 2003[48] Location: Sweden Funding: NS	**HOOS** **SF-36 BPS** **WOMAC**	SAQ, written	Clinic	Condition: OA of the hip Inclusion: Assigned total hip replacement; completed follow-up Exclusion: NR	Baseline pain score(s): NR Average intensity: NR	N=62 Age (mean, range): 72.8, 53-85 Women (%):45 Race/Ethnicity (%): NR
Parker 2012[49] Location: United States Funding: NS	**ODI** **VAS-back pain**	SAQ, written	University medical center	Condition: Symptomatic pseudoarthrosis, mechanical LBP Inclusion: Patients undergoing revision-instrumented fusion; age 18-70 years; prior lumbar instrumented fusion; failed to complete at least 3 months of non-operative care Exclusion: Extra-spinal cause of back pain; trauma, infection, or neoplasm; previous lumbar revision surgery for other causes	Baseline pain score(s): (Mean) VAS-back pain=7.3mm (δ 0.8mm); ODI= 59.4% (δ 10.8%) Average intensity: NR	N= 47 Age (mean, SD): 54.5, δ 10.5 Women (%): 64 Race/Ethnicity (%): NR
Pinsker 2015[50] Location:	**NRS** **WOMAC**	SAQ, mail	Patients home	Condition: Ankle arthroplasty or arthrodesis	Baseline pain score(s): NRS pre-op= 6.6, range 2-10;	N=142 N=124 (test/retest) Age (mean, range): 61.2,

Study (name and year)/ Location/ Funding	**Scale of Interest/** Others	**Mode of Administration**	**Setting[a]**	**Condition/ Study Inclusion/ Exclusion Criteria**	**Baseline Pain Characteristics**	**Demographics**
Canada Funding: Foundation				Inclusion: Age ≥ 18 years; able to complete survey in English; end-stage ankle arthritis (pre- or post-operative); surgical patients ≥ 6 months post-surgery Exclusion: NR	NRS post-op= 4.0, range 0-10; WOMAC (overall)=51.4, range 0-95.2 Average intensity: NR Type of pain (%): Arthroplasty=60 Arthrodesis=10 Pre-operative=30	22-92 Women (%): 54 Race/Ethnicity (%): NR
Scott 2015[51] Location: United Kingdom Funding: International Association	**PGIC**	SAQ, written	Pain treatment center	Condition: Chronic pain Inclusion: Significant levels of distress and disability Exclusion: Incomplete data	Baseline pain score(s): NR Average intensity: NR Type of Pain (%): Low back/spine=43.8 Upper shoulder=7.80 Lower limbs=13.30 Other=35.1	N=476 Age (mean, SD): 46.2, δ 11.2 Women (%): 66.8 Race/Ethnicity (%): White: 71.9 African-Amer./Black: 16.6 Asian: 7.1 Other: 4.4
Sindhu 2011[52] Location: United States Funding: NS	**NRS** **VAS-digital** **VAS**	VAS, written and digital NRS, verbal	Hand therapy clinics	Condition: Unilateral musculoskeletal disorder or injury to elbow, forearm, or hand Inclusion: Age 18-65 years; recruited from hand therapy clinics Exclusion: Verbally reported pain intensity > 7 (1-10); unable to perform grip test	Baseline pain score(s): NR Average intensity: (Mean) NRS=<2; VAS=<2mm	N= 33 Age (Mean, SD): 39, δ 12.3 Women (%): 48% Race/Ethnicity (%): NR
Stewart 2007[53]	**NRS**	SAQ, written	New-South Wales	Condition: Whiplash associated disorders (neck pain)	Baseline pain score(s): NR	N= 132 Age (mean, SD): 43, δ

Study (name and year)/ Location/ Funding	**Scale of Interest/** Others	**Mode of Administration**	**Setting[a]**	**Condition/ Study Inclusion/ Exclusion Criteria**	**Baseline Pain Characteristics**	**Demographics**
Location: Australia Funding: Government	**SF-36 BPS**	Baseline; follow up after completion of 6-week treatment period	community, physio-therapy clinics	Inclusion: Patients enrolled in RCT (effects of exercise and advice to exercise alone); Motor Accident Authority claimants seeking medical care for whiplash associated disorder (Grades I to III) within 1 month of accident; reported at least "mild" disability compared to pre-injury; significant pain or disability indicated by score of at least 20% on NRS scales or Patient-Specific Functional Scale Exclusion: Previous neck surgery, nerve root compromise, current physical therapy neck treatment	Average intensity: NR	14.7 Women (%): 67 Race/Ethnicity (%): NR
Stroud 2004[54] Location: United States Funding: NS	**RMDQ (**24-, 18-, and 11- item**)**	SAQ, written	University pain treatment center	Condition: Chronic pain Inclusion: Available RMDQ scale data Exclusion: NR	Baseline pain score(s): NR Average intensity: NR Type of pain (%): Lower back=36.2 Lower extremities=14.1 Head=12.5 Shoulder and arms=9.8 Upper back=4.8 Other=22.6	N=993 Age (mean, SD): 43.5, δ 12.6 Women (%): 57 Race/Ethnicity (%): White: 84.4 African-Amer./Black: 3 Asian: 2.3 Hispanic: 3.9 Native American: 3.7 Other: 2.7
Tan 2004[55] Location:	**BPI** **RMDQ**	SAQ, written Baseline; follow	VA chronic pain center	Condition: Chronic pain Inclusion: Completed BPI	Baseline pain score(s): NR	N=440 Age (mean, SD): 54.9, δ 21-85

Study (name and year)/ Location/ Funding	**Scale of Interest/** Others	**Mode of Administration**	**Setting[a]**	**Condition/ Study Inclusion/ Exclusion Criteria**	**Baseline Pain Characteristics**	**Demographics**
United States **(Enrolled Veterans)** Funding: NS		up assessments on subsequent visits		before initial visit; referred to chronic pain center Exclusion: NR	Average intensity: NR Type of pain (%): Multiple sites (including back)=50 Back only=28	Women (%): 8.2 Race/Ethnicity (%): White: 72.3 African-Amer./Black: 21.2 Other: 6.5
Tong 2006[56] Location: United States Funding: Academic institution	**VAS**	SAQ, written Baseline; follow up 2nd, 3rd, and 4th visits	University spine care center	Condition: LBP Inclusion: Referred for physical therapy at University spine care facility Exclusion: NR	Baseline pain score(s): (Mean) VAS=5.2mm (δ 2.1mm) Average intensity: NR	N=52 Age (mean, SD):41.1, δ 12.6 Women (%): 61.5 Race/Ethnicity (%): White: 88 African-Amer./Black: 3 Asian: 3 Other: 6
Trudeau 2015[57] Location: United States Funding: Industry	**NRS-Now** **NRS- 24 hours** **NRS-1 week** **WOMAC-48 hours**	SAQ, written and digital NRS-Now reported 4 times daily, NRS-24 hours reported daily, NRS- 1 week reported weekly, WOMAC-48 hours reported every 48 hours	Unclear	Condition: Pain from OA of the knee Inclusion: Age ≥ 21 years; diagnoses of functional classes 1-3 of knee OA; pain intensity on NRS ≤ 6; able to withdraw from OA medications; ≤ 10 on hospital anxiety and depression scale, Exclusion: History of major depressive disorders not controlled with medication	Baseline pain score(s): Treatment-placebo: NRS-24 =3.7 (δ1.22); WOMAC-48=8.6 (2.69); NRS-1 wk =4.2 (δ1.61) Placebo-treatment: NRS-24 =3.9 (δ1.32) WOMAC-48=8.7 (δ2.97); NRS-1 wk =4.4 (δ1.33) Average intensity: NR	N=47 Age (mean, SD): NR Women (%): NR Race/ethnicity (%): NR
Van der Roer, 2006[58] Location: Netherlands	**NRS**	SAQ, written Baseline; follow up 6, 12, 26, and 52 weeks	Physio-therapy clinics	Condition: LBP Inclusion: Participants in an RCT (comparison of physiotherapy strategies); referred to physiotherapy	Baseline pain score(s): NRS (by GPE category) Improved= 6.0 (δ 2.1); Unchanged= 6.4 (δ1.8)	N=138 Age (mean, SD): 44.0, δ 13.4 Women (%): 58.7 Race/ethnicity (%): NR

Study (name and year)/ Location/ Funding	Scale of Interest/ Others	Mode of Administration	Setting[a]	Condition/ Study Inclusion/ Exclusion Criteria	Baseline Pain Characteristics	Demographics
Funding: NS				treatment by physician for new episode of pain. Exclusion: Pregnant; unable to give consent	Average intensity: NR	
van Grootel 2007[59] Location: Netherlands Funding: International academic research institute	**VAS**	SAQ, written Baseline (pre-treatment); follow up post-treatment	University medical center	Condition: Myogenous temporomandibular disorders (TMD) Inclusion: Pain and tenderness of the mastication muscles; restricted mandibular opening of 3 months duration or longer; age 18-65 years Exclusion: Clinical and/or radiographic evidence of organic TMJ changes; recent TMD treatment (<1 year); other pain treatment; evidence of serious psychopathology (psychotherapy and/or psychomedication, recent dramatic life events)	Baseline pain score(s): (Mean) VAS=40mm (δ 22.3mm) Average intensity: NR	N=118 Age (mean, range): 31.6, 18-65 Women (%): 93 Race/Ethnicity (%): NR
Wittink 2004[60] Location: United States Funding: Government, Industry	**MPI** **ODI** **SF-36 BPS**	SAQ, written Baseline; follow up after 3 visits	Medical center pain program	Condition: Chronic pain Inclusion: More than 3 visits to medical center; referred to pain program	Baseline pain score(s): NR Average intensity: NR Type of Pain (%): Back=52.9 Neck=21.8 Myofascial=19.5	N=87 Age (mean): 46.9 Women (%): 66.5 Race/Ethnicity (%): White: 79.3 Other: 20.7

δ=standard deviation
(I)=interference; (S)=severity; BPI=Brief Pain Inventory; CAT= computer adaptive testing (subsequent questions depend on previous response); CDMD= chronic disabling musculoskeletal disorders; CPG=Chronic Pain Grade Questionnaire; IAQ=interview administered questionnaire; LBP=low back pain; MPQ=McGill Pain Questionnaire; MPI=Multidimensional Pain Inventory; MS=multiple sclerosis; NRS=numeric rating scale; NR=not reported; NS=not specified; OA=osteoarthritis

arthritis; ODI=Oswestry Disability Index (also known as Oswestry Low Back Pain Disability Questionnaire); PEG=items assess average pain intensity (P), interference with enjoyment of life (E), and interference with general activity (G); PGA=patient global assessment; PGART=patient-rated global assessment of response to therapy; PRGC=patient-reported Global Change; PROMIS-PI=Patient-Reported Outcomes Measurement Information System-Pain Interference; SAQ=self-administered questionnaire; SCAMP=Stepped Care for Affective Disorders and Musculoskeletal Pain; SF-36 BPS=Medical Outcomes Study Short Form-36 Bodily Pain Scale; SF= Short form; RCT=randomized controlled trial; RMDQ=Roland-Morris Disability Questionnaire; VAMC=Veterans Affairs Medical Center; WHYMPI=West Haven-Yale Multidimensional Pain Inventory (see also MPI); WOMAC=Western Ontario and McMaster Universities Arthritis Index

[a]Primary care, pain clinic, etcetera

[b]SCAMP study included RCT of combined depression medication and pain self-management vs usual care in patients with depression of at least moderate severity and observational study in patients with absence of clinical depression; responsive results analyzed separately

[c]William and Myers Version

[d]Stratford and Binkley Version

SUPPLEMENTAL TABLE 5. OUTCOMES REPORTED

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	Test-Retest Reliability	Inter-Rater Reliability	Concurrent and/or Criterion Validity	Discriminant and/or Predictive Validity	Responsive-ness (Describe)	Other Outcomes
Anagnostis 2004[20] **Scale(s): ODI** (0-100) Time: Varied (upon completion of program)						Responsiveness assessed by effect size P<.001 ES=0.95 Mean pre-post treatment change = 14.8, δ 15.6	
Askew 2016[21] **Scale(s): PROMIS-PI** (0-66) Time: 12 weeks						Responsiveness assessed using SRMs for PROMIS-PI scores SRM scores ≥0.30 indicated responsiveness Change by “general health” anchor Better=-0.94, δ 7.96 Same=-0.58, δ 7.97 Worse=-0.47, δ 7.18 Change by “pain” anchor Better=-1.09, δ 7.43 Same=-0.26, δ 6.33 Worse=0.44, δ 4.95	

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Burnham 2012[22] **Scale(s): MPQ (short form)** (pain 0-6), **ODI** (0-100) Time: 2 weeks (corticosteroid injection); 6-8 weeks (radiofrequency neurotomy or TransDiscal Biacuplasty)		Pearson's correlation coefficient between mean change 1 month before intervention and day of intervention MPQ r=0.88 (95%CI 0.72, 0.95) ODI r=0.89 (95%CI 0.75, 0.95)				Pre-post treatment responsiveness ratios (RR) (significant RR values >1.96) MPQ RR=1.9 ODI RR=2.3	
Changulani 2009[23] **Scale(s): ODI** (0-100%), **VAS** (domain not reported) Time: 6 weeks				Pearson's correlation coefficient between mean change in ODI scores and mean change in VAS scores r=0.44 (P<.05)		Based on: ES=1.05 Measured by SRM=0.84	

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Chansirinukor 2005[24] **Scale(s): RM-18** (Shortened version of **RMDQ** () Time: 12.1 weeks (±0.9)	Minimal detectable change (MDC) for RM-18: 7.5 points	Assessed in subset of patients whose work status had not changed from baseline to follow-up visit ICC=0.68 (95% CI 0.52, 0.79)				RM-18 correlated with change in work status using Spearman's ρ (0.30; Z=123, P=.02) AUC=0.69 (95%CI=0.60, 0.78) *1=perfect discrimination ES=0.44 (0.37-0.51) SES=0.38 (0.32-0.44) SRM=0.44 (0.37-0.51); paired t= 5.25	
Chien 2013[25] **Scale(s): BPI pain intensity items*** (4), each item scored 0-10 Time: 10 days *NOTE: 4 BPI items (current pain, worst pain [past 24 hr], least pain [past 24 hr], average pain) used to compute composite average pain						SMRs all participants (improved/unimproved) Current pain: 0.36 (0.89/-0.03) Worst pain: 0.37 (0.63, 0.14) Least pain: 0.17 (0.50, -0.03) Average pain: 0.40 (0.53, 0.28) Composite average pain: 0.42 (0.81/0.10) ROC Current pain: 0.75 Worst pain: 0.66 Least pain: 0.65 Average pain: 0.61 Composite average pain: 0.71	**Internal consistency** (Spearman correlation) -moderate/high correlations between BPI composite average pain score and component items: ρ=0.71-0.84, P< .01 -small/moderate correlations between BPI pain items: ρ=0.38-0.65, P< .01

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Cook 2008[26] **Scale(s): RMDQ-23-** (0-23);**11-** (0-11); **5-** (0-5) Time: Single administration of scale *Data from 2 previously published studies				Correlations between each CAT condition and scores based on RMDQ 23 ranged from 0.93 (5-item) to 0.98 (11-item) Standard error of measurement-based CAT scores correlated 0.95 with RM-MOD$_{IRT}$ scores			
de Vet 2007[27] **Scale(s): NRS (pain intensity)** (0-10) Time: 12 weeks	**Anchor-based (with global perceived effect)** **MIC, for chronic pain subjects (n=135):** 1) 95% cut-off limit 4.7 points 2) ROC cut-off 3.5 points **Change on NRS** 1) 0.5 sensitivity 95%, specificity 37% 2) 1.5 sensitivity 89% specificity 59% 3) 2.5 sensitivity 81% specificity 78% 4) 3.5 sensitivity 69% specificity 89% 5) 4.5 sensitivity 53% specificity 94%			**Changes in PI-NRS scores and the global perceived effect categories** (Spearman correlation): ρ=0.6			

Author Year/ Scale (range)/ Mean time between surveys	**OUTCOMES REPORTED**						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Deyo 2016[28] **Scale(s): PROMIS-PI SF** (4-20) Time: 12 weeks		At 3 months: -Patients that rated pain "about the same" ICC=0.58 (0.44, 0.71) -Patients that rated pain as "changed ± 1 point" ICC=0.67 (0.56, 0.77)			PROMIS-PI scores in those a) seeking worker's compensation (65.0) or not (59.8) (P<.001) b) who had a fall in past 3 months (62.7) or not (59.7) (P<.001)	Change of pain (much less to much worse) at 3 months compared to baseline *Pain Interference:* ES range: -1.03 (much less) to 0.71 (much worse) SRM range: -1.07 (much less) to 0.74 (much worse)	
Driban 2015[29] **Scale(s): PROMIS-PI SF** (41-78.3), **SF-36 BPS** (0-100), **WOMAC** (pain 0-500), **WOMAC** (function 0-1700) Time: baseline data *Secondary analysis of previously published RCT				Spearman's correlation coefficient (95%CI) PROMIS-PI/SF-36 BPS: ρ=-0.73 (-0.79, 0.65) PROMIS-PI/WOMAC Pain: ρ=0.47 (0.35, 0.57) PROMIS-PI/WOMAC Function: ρ=0.42 (estimated from Figure 3, confidence interval not provided)			

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Fisher 1997[30] **Scale(s): ODI** (0-100), **MPQ** (pain 0-6) Time: 15 weeks				**Criterion Validity (**Kendall's tau, all P<.01) a) ODI Lifting Subscale with behavioral assessment of lifting: τ=0.38 b) ODI Walking Subscale with behavioral assessment of walking: τ=0.54 c) ODI Sitting Subscale with behavioral assessment of sitting: τ=-0.40 **Sensitivity/ Specificity of ODI Subscales** a) Lifting: 81%/52% b) Walking: 76%/96% c) Sitting: 72%/69%			**Internal Consistency** Cronbach's alpha ODI=0.76 **Effect size for post-treatment change** ODI=0.6 MPQ a) Total Number of Words Chosen=0.5 b) Present Pain Inventory=NR but reported to be not significant

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Gallasch 2007[31] **Scale(s): Wong Faces, VAS** (0-10)**, NRS** (0-10) All scales: pain on previous day Time: Same day (pre- and post-physiotherapy)		Before and after physiotherapy session ICC: Faces=0.96 VAS=0.97 NRS=0.99					Rated easiest to understand: 1) Faces scale 38.7% 2) NRS 32.3% Easiest to fill out: 1) NRS 37.5% 2) verbal rating scale 32.2% Most difficult to understand VAS 58% Most difficult to fill out: VAS 67.8%

Author Year/ Scale (range)/ Mean time between surveys	**OUTCOMES REPORTED**						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Gentelle-Bonnassies 2000[32] **Scale(s): VAS** (pain, 0-100 mm)**, WOMAC** (pain and function 0-100) Time: 6 months						Based on SRM (95%CI) VAS, Pain, ITT (n=80) Month 1: -0.40 (-0.64, -0.16) Month 3: -0.13 (-0.35, 0.10) Month 6: -0.25 (-0.48, -0.02) WOMAC, Pain, ITT (N=80) Month 1: -0.39 (-0.60, -0.18) Month 3: -0.28 (-0.53, -0.02) Month 6: -0.30 (-0.55, -0.06) WOMAC, Function, ITT (N=80) Month 1: -0.37 (-0.64, -0.10) Month 3: -0.15 (-0.39, 0.09) Month 6: -0.09 (-0.33, 0.14)	

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Godil 2015[33] **Scale(s): NRS-neck pain, NRS-arm pain** (intensity, 0-10) Time: 12 months						**Based on SRM** *Responders:* NRS-neck pain=0.95 NRS-arm pain=0.97 *Non-responders*: NRS-neck pain=0.49 NRS-arm pain=0.38 SRMs in patients reporting meaningful improvement (responders) versus non-responders (greater difference = more responsive scale) Mean change: NRS-neck pain=0.46 NRS-arm-pain=0.59 Based on ROC (AUC) curve NRS-neck pain: AUC=0.69 (poor discriminator) NRS-arm-pain: AUC=0.74 (valid discriminator)	
Gronblad 1993[34] **Scale(s): ODQ** (0-50), **VAS** (present pain intensity, 0-100) Time: Subset after 1 week		Subset chosen, n=20, 1 week interval ODQ ICC=0.83		Pearson's correlation ODQ/VAS r=0.62			

Author Year/ Scale (range)/ Mean time between surveys	**OUTCOMES REPORTED**						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Hicks 2009[35] **Scale(s): ODI** (0-100), **SF-36 BPS** (0-100) Time: Mean 11 days	Standard error of measurement = 4.57 (using data from participants classified as stable) Minimum detectable change ODI: 10.7 points 14.5% scored below 10.7 0% scored above 89.3	ODI subset of patients with stable LBP status from baseline to follow-up (mean 11 days) ICC 0.92 (95% CI 0.86, 0.95)		"Convergent" ODI/ SF-36 BPS: r=-0.69 (-0.78, -0.60) (P<.0001)	ODI scores significantly different (P<.0001) between groups with and without 1) high pain severity/high functional limitation and 2) chronic pain/high functional limitation (n=107)		

Author Year/ Scale (range)/ Mean time between surveys	**OUTCOMES REPORTED**						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Jensen 2012[36] **Scale(s): VAS** (pain intensity, 0-100mm) Time: 12 weeks post-randomization *Data obtained from 2 previously published RCTs							**Discriminating active tx from placebo** VAS ≥20mm: OR 1.94 (1.37, 2.75) VAS ≥30%: OR 1.97 (1.41, 2.77) VAS ≥50%: OR 2.46 (1.72, 3.50) **Agreement between response criteria (kappa)** a) 20 mm improvement with 30% improvement: k=.90 b) 20 mm improvement with 50% improvement: k=.51 c) 30% improve-ment with 50% improvement: k=.58

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Kamper 2015[37] **Scale(s): SF-36 BPS** (pain only, 0-100)**, NRS-24hr** and **NRS-Wk** (pain intensity, 0-10) Time: 3, 6, and 12 months *Secondary analysis of data from 3 clinical studies; studies 2 and 3 were chronic pain cohorts				Pearson's correlation coefficient *Study 2*:* NRS-24/SF-36 BPS Baseline: r=0.37 3 months: r= 0.71 12 months: r= 0.68 *Study 3*:* NRS-24/SF-36 BPS Baseline: r= 0.40 3 months: r=0.65 6 months: r=0.66 12 months: r=0.65 NRS-Wk/SF-36 BPS Baseline: r=0.46 3 months: r=0.64 6 months: r=0.72 12 months: r=0.70 NRS-24/NRS-Wk Baseline: r=0.72 3 months: r=0.87 6 months: r=0.90 12 months: r=0.93			

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Kean 2016[38] **Scale(s): PROMIS-PI SF (6b)** (6 items, total score 6-30), **BPI** (4 item severity [S], 7 item interference [I], and 11 item total, each item scored 0-10), **PEG** (severity and interference combined, each item scored 0-10), **SF-36 BPS** (0-100) Time: 3 months						Responsiveness to intervention (SCOPE trial) (Cohen's *d)* BPI-S: 0.37 BPI-I: 0.33 BPI total: 0.38 PEG: 0.35 SF-36 BPS: -0.24 PROMIS-PI-SF: 0.21 AUC (SE) for detecting any improvement BPI-S= 0.73 (0.03) BPI-I= 0.68 (0.04) BPI total= 0.73 (0.03) PEG= 0.71 (0.03) SF-36 BPS= 0.68 (0.04) PROMIS-PI-SF= 0.61 (0.04) AUC (SE) for detecting moderate improvement BPI-S= 0.74 (0.04) BPI-I= 0.69 (0.04) BPI total= 0.74 (0.04) PEG= 0.72 (0.04) SF-36 BPS= 0.64 (0.05) PROMIS-PI-SF= 0.66 (0.04) SRMs significantly different (P<.05) between those who report being better vs stayed the same and those who report being worse vs stayed the same for all BPI scales, PEG, SF-36 BPS, and PROMIS-PI-SF	

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Keller 2004[39] **Scale(s): BPI-SF (15 items, 4 severity, 7 interference, 4 other), GCPS (7 items, 3 intensity, 4 disability), SF-36 BPS, RMDQ-24** Time: First follow-up visit Note: all results for low back pain group only				**Pearson's r correlations** BPI severity and GCPS intensity=0.60 GCPS disability=0.49 RMDQ=0.57 SF-36 BPS=0.61 BPI interference and GCPS intensity=0.64 GCPS disability=0.69 RMDQ=0.64 SF-36 BP=0.64 SF-36 BPS and GCPS intensity=0.47 GCPS disability=0.45 RMDQ=0.53		Standardized Response Means among improved patients BPI severity=-1.09 BPI interference=-1.13 GCPS intensity=-0.47 GCPS disability=-0.47 SF-36 BPS=0.69	**Internal consistency (Cronbach's alpha)** BPI severity=0.82 BPI interference=0.93 GCPS intensity=0.65 GCPS disability=0.94 RMDQ=0.92 SF-36 BPS=0.84
Kerns 1985[9] **Scale: WHYMPI (pain severity and pain interference), MPQ (Present Pain Intensity, Total Pain Rating Index)** Time: 2 weeks		WHYMPI scales Pain severity: r=0.82 Pain interference: r=0.86		**Correlation with a factor related to severity and interference** WHYMPI Pain Severity: 0.81 WHYMPI Pain Interference: 0.70 MPQ Total Pain Rating Index: 0.47 MPQ Present Pain Intensity: 0.44			**Internal consistency (Cronbach's alpha)** WHYMPI Pain severity=0.72 Pain interference=0.90

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Krebs 2010[40] **Scale(s): BPI** (4 item severity [S], 7 item interference [I], and 11 item total, each item scored 0-10), **PEG** (severity and interference combined, each item scored 0-10), **GCPS** (3 item intensity [S], 3 item disability [D], 0-10), **RMDQ-24** (0-24), **SF-36 BPS** (0-100) Time: 12 months	**Kappa for agreement between one-SEM and global rating classification** *Observational cohort* BPI-S=0.31 BPI-I=0.20 BPI total=0.34 PEG=0.23 GCPS-S=0.27 CGPS-D=0.14 RMDQ-24=0.18 SF-36 BPS=0.27 *RCT group* BPI-S=0.32 BPI-I=0.24 BPI total=0.29 PEG=0.33 GCPS-S=0.35 GCPS-D=0.27 RMDQ-24=0.36 SF-36 BPS=0.19					**AUC for responsiveness – detecting *moderate* improvement** *Observational cohort* BPI-S=0.81 (0.04) BPI-I=0.67 (0.05) BPI total=0.76 (0.04) PEG=0.70 (0.05) GCPS-S=0.73 (0.06) GCPS-D=0.66 (0.06) RMDQ-24=0.70 (0.05) SF-36 BPS=0.70 (0.05) *RCT group* BPI-S=0.85 (0.04) BPI-I= 0.77 (0.05) BPI total=0.81 (0.04) PEG=0.79 (0.04) GCPS-S=0.82 (0.04) CGPS-D=0.76 (0.04) RMDQ-24=0.85 (0.04) SF-36 BPS=0.77 (0.04) *Observational cohort* Mean SRMs differed significantly between “worse” and “same” groups and between “better” and “same” groups for each measure *RCT group* Mean SRMs differed significantly between “better” and “same” groups for each measure; values did not differ significantly between “worse” and “same” groups	

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Krebs 2009[12] **Scale(s): BPI (**4 item severity [S] and 7 item interference [I]; each item scored 0-10), **PEG** (severity and interference combined, each item scored 0-10), **SF-36 BPS** (0-100), **GCPS** (3 item intensity [S], 3 item disability [D], transformed to 0-100 scores), **PGIC** (1 item, change in pain,1-7), **RMDQ-24** (0-24) Time: 6 months *Brief Pain Inventory (BPI), Graded Chronic Pain (GCP), RMDQ, and SF-36 BPS administered at baseline; BPI, GCP, and patient global rating of change administered at 6 months				Validity (Pearson's r) PEG/BPI-S: r=0.69 PEG/BPI-I: r=0.89 PEG/GCPS-S: r=0.64; PEG/GCPS-D: r=0.67 PEG/RMDQ-24: r=0.60 PEG/SF-36 BPS: r=-0.61 BPI-S/BPI-I: r=0.58 BPI-S/CPGS-S: r=0.82 BPI-S/CPGS-D: r=0.47 BPI-S/RMDQ-24: r=0.41 BPI-S/SF-36 BPS: r=-0.46 BPI-I/CPGS-S: r=0.62 BPI-I/CPGS-D: r=0.71 BPI-S/RMDQ-24: r=0.70 BPI-S/SF-36 BPS: r=-0.65		Proportion of pain improvement after 6 months according to PGIC (31.4%) and GCPS (29.5%) - "similar" Improved group (based on PGIC) had mean improvement on PEG of 3 points (δ 2.5) and GCPS of 2.6 points (δ 2.7) - "similar" SRM among improved patients according to PGIC similar for PEG (1.20, 95%CI 0.96, 1.44), BPI-S (1.04, 95%CI 0.80, 1.28), and BPI-I (1.13, 95%CI 0.89, 1.37) For all measures of improvement ES and SRM were consistent with large effects	Reliability (internal consistency) – PEG: 0.73 Construct validity "good" PEG: r=0.60-0.89

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Krebs 2007[41] **Scale(s): NRS** (current pain intensity, 0-10) Time: Single administration of scale					**Accuracy of NRS - predicting** *1) pain that interferes with function (BPI≥5):* a) AUC=0.76 b) likelihood ratios: i) NRS=0: 0.39 (0.29, 0.53) ii) NRS=1-3: 0.99 (0.38, 2.60) iii) NRS=4-6: 2.67 (1.56, 4.57) iv) NRS=7-10: 5.60 (3.06, 10.26) *2) pain that motivates a visit:* a) AUC=0.78 b) likelihood ratios: i) NRS=0: 0.35 (0.26, 0.48) ii) NRS=1-3: 2.00 (0.78, 5.13) iii) NRS=4-6: 3.06 (1.75, 5.37) iv) NRS=7-10: 6.04 (3.18, 11.48)		

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Lovejoy 2012[42] **Scale(s): MPQ-2 SF** (22 pain descriptors [6 continuous, 6 intermittent, 6 neuropathic, and 4 affective] each rated 0-10 and total pain score), **MPI (severity [S] and interference [I] scales)** Time: Single administration of scale				Bivariate correlations using Pearson's r MPQ-2 SF/ MPI-S: r=0.72 MPQ-2 SF/ MPI-I: r=0.66	MPQ-2-SF discriminant validity (mean) vs a) 1 pain diagnoses 2.44 (δ2.14)* b) 2-3 pain diagnoses 2.97 (δ2.13)* c) ≥4 pain diagnoses 3.81 (δ2.36) *P<.01 vs c) And vs MPI-S a) None/Mild (score 0-2) 1.16 (δ1.69)** b) Moderate (score 2-4) 3.08 (δ1.68)** c) Severe (score >4) 5.55 (δ2.00)** **All different (P<.01)		**Internal consistency reliability** (Cronbach's α) MPQ-2-SF Total score: α =0.96
Lund 2005[43] **Scale(s): VAS** (pain intensity, 0-100) Time: Same day		Same day agreement: 20%					

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Macedo 2011[44] **Scale(s): RMDQ-24** (0-24), **RMDQ-18[a]** (0-18), **RMDQ-18[b]** (0-18), **RMDQ-11** (0-11) Time: 8-12 months						Internal responsiveness assessed using ES (84%CI): RMDQ-24: 0.67 (0.63-0.71) RMDQ-18[a]: 0.75 (0.71-0.79) RMDQ-18[b]: 0.78 (0.73-0.82) RMDQ-11: 0.65 (0.61-0.69) External responsiveness assessed using AUC values for patients classified as improved and not improved (based on GPE scale): RMDQ-24=0.78 (0.76-0.81) RMDQ-18[a]=0.78 (0.75-0.81) RMDQ-18[b]=0.78 (0.75-0.81) RMDQ-11=0.75 (0.72-0.78)	
Maughan 2010[45] **Scale(s): NRS** (intensity, 0-10), **ODI-2 (0-50), RMDQ-24 (0-24)** Time: 5 weeks	MCID (ROC approach) RMDQ-24: 3.5 ODI-2: 7.5 NRS: 4.0					AUC RMDQ-24: 0.64 ODI-2: 0.67 NRS: 0.5	
Merriwether 2016[46] **Scale(s): PROMIS-PI-SF (6b)** Time: 2nd visit							Internal Consistency Cronbach's alpha 0.90

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
MIkail 1993[47] **Scale(s): MPI** (interference and pain severity), **ODI** Time: Same day				MPI Interference correlated with: ODI: 0.66 MPI Pain Severity: 0.55			
Nilsdotter 2003[48] **Scale(s): HOOS** (40 items, 10 pain, 5 symptoms, 17 activity limitations [ADL], 4 sport/recreation function, 4 hip-related quality of life, 0-100), **WOMAC LK 3.0** (pain, function, 0-20), **SF-36 BPS** (0-100) Time: 6 months				Spearman's correlation HOOS (pain)/SF-36 BPS: ρ=0.61 HOOS (ADL)/SF-36 BPS; ρ=0.62		Responsiveness calculated as SRM after 6 months HOOS (pain)=2.11 WOMAC (pain)=1.83 HOOS (ADL)=1.70 WOMAC (function)=1.70	

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Parker 2012[49] **Scale(s): VAS-back pain** (severity, 0-10), **ODI** (0-100) Time: 2 years	**MCID thresholds (4 anchor-based approaches):** 1) Mean change approach: VAS-back pain=3.2 ODI=8.2 2) Minimum detectable change (95% CI) approach: VAS-back pain=2.2 ODI=2.0 3) Change difference approach: VAS-back pain=2.0 ODI=8.3 4) Receiving operating characteristic curve approach: VAS-back pain= 3.0, AUC=0.71 ODI=4.0, AUC= 0.90						
Pinsker 2015[50] **Scale(s): WOMAC** (pain, physical function, 0-20)**, NRS** (pain, 0-10) Time: NS		Mean of 15.5 days (range 4-35) between ICC WOMAC: Overall=0.90 Pain=0.90 Function=0.89 NOTE: limited to individuals who completed retest survey and reported condition to be stable on global change question		Spearman's rank correlations NRS pain/ WOMAC Overall=0.78 Pain=0.78 Function=0.73			Internal Consistency (Cronbach's α) WOMAC Overall= 0.97 Pain= 0.91 Function= 0.96

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Scott 2015[51] **Scale(s): PGIC** (pain, physical function, 1-7) Time: Single administration of scale						Effect sizes computed from pre- to post-treatment (Cohen's *d*) Pain=0.56 Physical Function=0.56	
Sindhu 2011[52] **Scale(s): VAS-D** (digital, pain level, 1-10), **VAS-P** (paper, pain level, 1-10), **NRS-V** (verbal, pain level, 1-10) Time: Administered twice on one visit, before and after grip tests (5-10 minutes apart) Up to 4 grip tests performed (1-minute apart)		ICC (pre-grip): VAS-P 0.96 VAS-D 0.96 NRS-V 1.00		Concurrent validity measured by Pearson's r: Pre-grip VAS-D/VAS-P =0.97 NRS-V/VAS-P= 0.84 NRS-V/VAS-D =0.84 Post-grip VASD/VAS-P=0.95 NRS-V/VAS-P=0.93 NRS-V/VAS-D =0.93		Mean score change between pre- and post-grip pain levels: VAS-P=0.40 VAS-D=0.48 NRS-V=0.54 Effect size coefficient (change score average/SD of pre-grip pain score): VAS-P=0.29 VAS-D=0.32 NRS-V=0.37 ANOVA on change scores showed no significant difference in responsiveness among scales: F= 1.36, P<=.25	

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Stewart, 2007[53] **Scale(s): NRS** (pain intensity [I], bothersomeness [B], 1-10), **SF-36 BPS** (0-100) Time: 6 weeks						**Internal responsiveness** *ES (84% CI):* NRS-I= 0.75 (0.61, 0.89) NRS-B=1.17 (1.02, 1.31) SF-36 BPS=0.49 (0.36, 0.61) Subpopulation* NRS-I=1.03 (0.88, 1.18) NRS-B=1.40 (1.24, 1.56) SF-36 BPS=0.72 (0.58, 0.86) *SRM (84% CI):* NRS-I= 0.64 (0.52, 0.77) NRS-B=0.98 (0.86, 1.10) SF-36 BPS=0.48 (0.35, 0.60) Subpopulation* NRS-I= 0.96 (0.82, 1.10) NRS-B=1.20 (1.06, 1.34) SF-36 BSP=0.71 (0.57, 0.85) *Subpopulation (n=101) participants who improved on GPE scale **External responsiveness Pearson's r for change score and AUC:** NRS-I=0.49 (0.68) NRS-B=0.47 (0.70) SF-36 BPS= 0.41 (0.73)	
Stroud 2004[54] **Scale(s): RMDQ-24** (0-24), **RMDQ-18** (0-18), **RMDQ-11** (0-11) Time: Single administration of scale				Intercorrelations among RMDQ 24-, 18-, and 11- item scales P<.01 24/18 =0.98 24/11 =0.93 18/11 =0.95			

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Tan 2004[55] **Scale(s): BPI** (intensity [S], interference [I], 0-10)**, RMDQ-24** (0-24) Time: Varied (upon follow up visits)				Concurrent validity (Pearson's r) BPI-I/RMDQ-24 r=0.57 BPI-S/RMDQ-24 r=0.40		Significant improvement with treatment confirms responsiveness of BPI intensity (S) and interference scales (I) (P<.001) Mean change (Visit 1 to Visit 3): BPI-S 0.93, t=5.33 (P<.001) BPI-I 0.96, t=4.66 (P<.001)	
Tong 2006[56] **Scale(s): VAS** (pain intensity, 0-100mm) Time: Administered at 2nd, 3rd, 4th, and final visits *Patients usually seen 2x/week for physical therapy							**Spearman's rank order correlation (r):** early responses at second (r=0.32, P=.02) third (r=0.34, P=.01), and fourth visits (r=0.62, P<.001) significantly correlated with discharge change in pain **Discriminant analysis:** early responses (2nd-4th visits) correctly predicted 80.4% of discharge outcomes (P<.001) defined by 30% improvement vs no improvement

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
Trudeau 2015[57] **Scale(s): NRS-24 hr** (pain intensity, 0-10), **NRS- 1 wk** (pain intensity, 0-10), **WOMAC** (pain 0-20) Time: 4 x daily, 24 hours, 48 hours, 1 week						Differences between treatment and placebo were measured using SES NRS-24hr=0.33, P=.02 WOMAC-48hr=0.54, P=.001 NRS-1 wk=0.38, P=.01	
Van der Roer 2006[58] **Scale(s): NRS** (pain intensity, 0-10) Time: 6, 12, 26, and 52 weeks	Chronic pain subgroup results Minimal Clinically Important Difference (MCIC) with NRS using 3 methods: 1) Δ= 3.7 (δ 2.1) 2) Minimal detectable change (95%CI= 4.5 (3.4-6.7) 3) Optimal cutoff point (sensitivity; specificity): 2.5 (77; 82) NRS sensitivity analysis showing range of MCIC results for lowest tertile baseline scores and highest tertile baseline scores: Low scores: 1.5-3.3 High scores 4.5-5.5						

Author Year/ Scale (range)/ Mean time between surveys	OUTCOMES REPORTED						
	Minimally Important Difference (Describe)	**Test-Retest Reliability**	**Inter-Rater Reliability**	**Concurrent and/or Criterion Validity**	**Discriminant and/or Predictive Validity**	**Responsive-ness (Describe)**	**Other Outcomes**
van Grootel 2007[59] **Scale(s): VAS** (pain intensity, 0-100mm) Time intervals: 1, 7 and 13 days (diary/SDD); 2-18 months (question-naire/ CID)	Smallest detectable difference (SDD) determined by calculating difference between duplicate VAS scores for each subject SDD=49 mm (for 13 days – longest interval)						
Wittink 2004[60] **Scale(s): SF-36 BPS** (0-100)**, MPI** (pain [S] 0-120; interference [I] 0-108), **ODI** (0-100) Time: After 3 visits				Overlap of the instruments measured using R^2 values (≥0.4 is high overlap) MPI-S/ODI=0.43 MPI-I/ODI=0.43 SF-36 BPS/MPI-S=0.58 SF-36 BPS/ODI =0.37		Responsiveness to change determined by ES from baseline to posttreatment (ES of <0.4 is small, >0.5 moderate, and >0.8 large) MPI-S=-0.41 MPI-I=-0.42 ODI=-0.39 SF 36 BPS=0.44	

δ=standard deviation; τ=Kendall's Tau; ρ=Spearman's rho; r=Pearson's r

ADL=activities of daily life; AUC=area under curve; BP=bodily pain; BPI=Brief Pain Inventory; CAT= computer adaptive testing; CDMD= chronic disabling musculoskeletal disorders; D=disability; ES=effect size; GCPS=Graded Chronic Pain Scale; GPE=global perceived effect; CPG=Chronic Pain Grade Questionnaire; I=Interference; ICC=Intraclass correlation coefficients; KOOS=Knee Osteoarthritis Outcome Score; MDC=minimal detectable change; MPQ=McGill Pain Questionnaire; NRS=numeric rating scale; ODI=Owestry Disability Index (also known as Oswestry Low Back Pain Disability Questionnaire); PEG=items assess average pain intensity (P), interference with enjoyment of life (E), and interference with general activity (G); PF= physical functioning; PGA=patient global assessment; PI=pain interference; PROMIS-PI=Patient-Reported Outcomes Measurement Information System-Pain Interference; RMDQ=Roland Morris Disability Questionnaire; ROC=receiver operating characteristic curve; S=severity/intensity; SCOPE=Stepped Care to Optimize Pain Care Effectiveness; SE=standard error; SES=standardized effect size; SF-36 BPS=Medical Outcomes Study short form-36 Bodily Pain Scale; SF-MPQ-2=Short Form McGill Pain Questionnaire; SRM=standardized response mean (SRM value 0.2-0.5 = small change, 0.5-0.8 = moderate, and >0.8 = large)

[a]William and Myers version

SUPPLEMENTAL TABLE 6. SUMMARY OF MINIMALLY IMPORTANT DIFFERENCE OUTCOMES

Study (ref)/ mode of administration/ (*version*)	n Condition of pain Time interval	MID equivalent	Approach(es) used to estimate MID equivalent
Studies Estimating MID for More Than One Scale			
Parker 2012[49] SAQ, on-site	47 Pseudoarthrosis (revision fusion patients) 2 years	*Oswestry Disability Index (range 0-100)* Average change approach 8.2 points Minimum detectable change 2.0 points Change difference approach 8.3 points ROC approach 4.0 points *VAS (range 0-10)* Average change approach 3.2 points Minimum detectable change 2.2 points Change difference approach 2.0 points ROC approach 3.0 points	Distribution and anchor-based Four approaches to MCID: 1) Average change approach: the average change score seen in the group defined by anchor to be responders 2) Minimum detectable change: the upper value of the 95% confidence interval for average change score seen in the cohort defined by anchor to be non-responders 3) Change difference approach: difference of the average change score for anchor-determined responders and non-responders 4) ROC approach: the change value that provides the greatest sensitivity and/or specificity for an anchor-determined positive response Two anchors produced the same responder/non-responder split: 1) SF-36 Health Transition Index, adapted: Patient rating of health before vs after surgery (markedly better or slightly better vs unchanged or worse) 2) Satisfied with results of surgery (yes vs no)
Krebs 2010[40] SAQ, on-site *Randomized trial*	205 Back, hip, or knee 12 months	*SEM* *BPI (range 0-10)* BPI-S: 0.7 BPI-I: 0.7 BPI total: 0.6 *PEG (range 0-10):* 1.8 CPG intensity (*range 0-100):* 9.0 CPG disability (*range 0-100):* 8.7 *RMDQ (range 0-24):* 1.0 *SF-36 BPS (range 0-100):* 9.8	Distribution and anchor-based minimal clinically important change (MCIC) Distribution: Change classified by one-SEM criteria as follows: better score improved ≥1 SEM from baseline, same score change <1 SEM from baseline, and worse score worsened ≥1 SEM. Anchor: Patient-reported retrospective global rating of change (better, about the same, worse) Agreement between anchor and SEM was then examined via weighted kappa statistics.
Krebs 2010[40]/ SAQ, on-site *Cohort study*	222 Back, hip, or knee 12 months	*SEM* *BPI (range 0-10)* BPI-S: 0.8 BPI-I: 0.8 BPI-total: 0.7 *PEG (range 0-10):* 1.9 CPG intensity (*range 0-100):* 9.9 CPG disability (*range 0-100):* 10.3	

Study (ref)/ mode of administration/ (*version*)	n Condition of pain Time interval	MID equivalent	Approach(es) used to estimate MID equivalent
		RMDQ (range 0-24): 1.2 *SF-36 BPS (range 0-100):* 11.8	
Maughan 2010[45] SAQ, on-site	63 (48) [a] Back 5 weeks	*Oswestry Disability Index (range 0-100)* Minimum detectable change 16.7 points ROC approach 7.5 points *RMDQ (range 0-24)* Minimum detectable change 4.9 points ROC approach 3.5 points *Numeric Rating Scale (range 0-10)* Minimum detectable change 2.4 ROC approach 4 points	Distribution and anchor-based minimal clinically important difference (MCID) Distribution: Minimal detectable change approach, estimated by 1.96 x square root of 2 x SEM test-retest. Anchor: Patient-reported global impression of change (much improved /completely better, unchanged, worse than ever). ROC analysis assessed the ability to distinguish patients who had and had not changed according to patient-reported global impression of change
Single Studies by Pain Scale			
Numeric Rating Scale for pain intensity (range 0 to 10)			
de Vet 2007[27] SAQ, on-site [see van der Roer, same study population]	135 (chronic) Lower back 12 weeks	ROC approach 3.5 points 95% limit cut-off approach 4.7 points	Distribution and anchor-based: Minimally important change (MIC) Distribution: distribution of the change in scores was plotted on anchor-based axes and 2 cut-points were applied, ROC and 95% limit Anchor: global perceived effect (completely recovered, much improved, slightly improved, no change, slightly worse, much worse). These were then clustered into 1) importantly improved, 2) not importantly changed, and 3) importantly deteriorated
van der Roer 2006[58] SAQ, on-site	138 (chronic) Lower back 12 weeks	Minimal detectable change approach 4.5 points Mean change approach 3.7 points Optimal cutoff point approach 2.5 points	Distribution and anchor-based: Minimal Clinically Important Change (MCIC) 1) Minimal detectable change approach: estimated by 1.96 x square root of 2 x SEM test-retest. 2) Mean change approach: mean change score of all patients who “improved” based on the GPE 3) Optimal cutoff point approach: point that yields the lowest overall misclassification, based on ROC curve. Anchor: global perceived effect (completely recovered, much improved, slightly improved, no change, slightly worse, much worse). These were then clustered into 1) improved, 2) unchanged, and 3) deteriorated
Oswestry Disability Index (range 0 to 100 points)			
Hicks 2009[35] SAQ, mail (*modified*)	107 (56)[a] Lower back 11 days	10.7 points	Distribution: minimum detectable change (MDC) SEM determined from participants classified as stable. The SEM was then used to calculate the 90% CI and then multiplied by the square root of 2, which

Study (ref)/ mode of administration/ (*version*)	n Condition of pain Time interval	MID equivalent	Approach(es) used to estimate MID equivalent
			resulted in an estimate of MID
Roland Morris Disability Questionnaire (range 0-24 points)			
Chansirinukor 2005[24] SAQ, on-site (*18-item*)	143 Lower back 3 months	$MDC_{95\%}$ 7.5 points	Distribution: minimal detectable change (MDC) 95% CI of the MDC was estimated by ± square root of 2 x $SEM_{test\text{-}retest}$ x 1.96
Visual Analog Scale (range 0 to 100 mm)			
van Grootel 2007[59] SAQ, on-site	118 (95-109)[a] Temporomandibular disorders 2 weeks	49 mm	Distribution: smallest detectable difference (SDD) Estimated by the standard deviation of the difference values x 1.96

BPI=Bodily Pain Index; BPS=Bodily Pain Scale: CPG=Chronic Pain Grade Questionnaire (also known as the Graded Chronic Pain Questionnaire); LBP=low back pain; NR=not reported; RMDQ=Roland Morris disability questionnaire; ROC=receiver operating characteristic curve; SAQ=self-administered questionnaire; SEM=standard error of measurement; VAS= visual analog scale;

[a] Post-treatment, no further details

SUPPLEMENTAL TABLE 7. SUMMARY OF RESPONSIVENESS OUTCOMES

Study (ref)/ Mode of administration (*version*)	N Condition of Pain Time interval	Responsiveness Results	Approach(es) used to estimate Responsiveness
Comparative Studies			
Kean 2016[38]/ Interview, on-site	250 (244)[a] Musculoskeletal (moderate) 3 months	AUC, any improvement BPI-S 0.73; BPI-I 0.68; BPI total 0.73 PEG 0.71 PROMIS-29-Profile PI 0.56; PROMIS-57-Profile PI 0.57; PROMIS-PI Short form 6b 0.61 SF-36 Bodily Pain 0.68 SRMs BPI-S: Worse -0.47; Same 0.13; Better 0.71 BPI-I: Worse 0.03; Same 0.38; Better 0.94 BPI total: Worse -0.22; Same 0.31; Better 0.93 PEG: Worse -0.14; Same 0.25: Better 0.86 PROMIS-29 Profile PI: Worse -0.11; Same 0.29; Better 0.33 PROMIS-57 Profile PI: Worse -0.16; Same 0.30; Better 0.37 PROMIS-PI Short form 6b: Worse -0.02; Same 0.27 Better 0.51 SF-36 Bodily Pain: Worse 0.17; Same -0.38; Better -0.71 SES (ES) BPI-S: 0.38 (Cohen's *d* 0.37) BPI-I: 0.37 (Cohen's *d* 0.33) BPI total: 0.42 (Cohen's *d* 0.38) PEG: 0.37 (Cohen's *d* 0.35) PROMIS-29 Profile PI: SES 0.17 (Cohen's *d* 0.14) PROMIS-57 Profile PI: SES 0.24 SES 0.42 (Cohen's *d* 0.38) PROMIS-PI Short form 6b: SES 0.28 (Cohen's *d* 0.21) SF-36 Bodily Pain: -0.25 (Cohen's *d* -0.24)	Based on SRM, SES and ES (0.2 is small, 0.5 is medium, and 0.8 is large) and ROC/AUC (0.5 is the same as chance to 1.0 is perfect discrimination). Anchored by patient-reported global change (much better, moderately better, a little better, no change, a little worse, moderately worse, and much worse)
Trudeau 2015[57] SAQ, on-site	47 Knee OA 1 week	SES NRS, 1 week: 0.38 NRS, 24 hours: 0.33 WOMAC-pain, 48 hours: 0.54	Based on SES of differences in pain scores between treatment and placebo
Burnham 2012[22] SAQ, on-site	67 Lower back 2-8 weeks	Responsiveness ratios Oswestry Disability Index 2.3; MPQ 1.9	Based on responsiveness ratio (RR). The RR evaluates intervention-related change over time while considering the between-subject variability in within-subject changes in stable subjects. Significant RR values should be >1.96
Sindhu 2011[52] SAQ, on-site (*paper and digital*)	33 Arm/hand Pre-post gripping	ES VAS-paper 0.29; VAS-digital 0.32 NRS 0.37	Based on ES of change scores between pre- and post-gripping pain levels

Study (ref)/ Mode of administration (*version*)	N Condition of Pain Time interval	Responsiveness Results	Approach(es) used to estimate Responsiveness
Krebs 2010[40] SAQ, on-site	427 Back, hip, or knee 12 months	<u>AUC, any improvement</u> BPI-S-cohort 0.83; BPI-S-RCT 0.81 BPI-I-cohort 0.70; BPI-I-RCT 0.78 BPI total-cohort 0.78; BPI total-RCT 0.81 PEG-cohort 0.73; PEG-RCT 0.78 CPG intensity-cohort 0.75, CPG intensity-RCT 0.78 CPG disability-cohort 0.65, CPG disability-RCT 0.75 RMDQ-cohort 0.70: RMDQ-RCT 0.81 SF-36 Bodily Pain-cohort 0.68: SF-36 Bodily Pain-RCT 0.72 <u>SMRs</u> BPI-S-cohort: Worse 0.75; Same 0.08: Better -1.07 BPI-S-RCT: Worse 0.29; Same -0.02; Better -0.99 BPI-I-cohort: Worse 0.43; Same -0.09: Better -0.69 BPI-I-RCT: Worse 0.06; Same -0.50; Better -1.06 BPI total-cohort: Worse 0.63; Same -0.04: Better -0.99 BPI total-RCT: Worse 0.15; Same -0.42; Better -1.15 PEG-cohort: Worse 0.35; Same -0.13; Better -0.83 PEG-RCT: Worse -0.05; Same -0.49; Better -1.14 CPG intensity-cohort: Worse 0.60; Same 0.07; Better -0.68 CPG intensity-RCT: Worse 0.56; Same -0.03; Better -0.73 CPG disability-cohort: Worse 0.37; Same -0.03; Better -0.57 CPG disability-RCT: Worse 0.14; Same -0.25; Better -0.94 RMDQ-cohort: Worse 0.57; Same -0.03; Better -0.67 RMDQ-RCT: Worse 0.35; Same -0.29; Better -1.09 SF-36 Bodily Pain-cohort: Worse -0.58; Same 0.17; Better 0.67 SF-36 Bodily Pain-RCT: Worse -0.17; Same 0.31; Better 0.76	Based on SRM and ROC/AUC. Anchored by global rating of change at 12 months (worse, same, or better)
Maughan 2010[45] SAQ, on-site	63 (48) [a] Back 5 weeks	<u>AUC</u> Oswestry Disability Index 0.67 RMDQ-24 0.64 NRS 0.5	Based on ROC/AUC. Anchored by patient-reported global impression of change (much improved /completely better, unchanged, worse than ever)
Krebs 2009[12] SAQ, on-site	210 Back, hip, or knee 6 months	<u>SRM (ES)</u> *Global rating of change* PEG Improved 1.20 (1.29); Unchanged 0.29 (0.26); Worse -0.06 (-0.06) BPI-severity Improved 1.04 BPI-interference Improved 1.13 *Chronic Pain Grade questionnaire* PEG Decreased by ≥1 level 0.99 (1.51); Baseline = follow-up 0.29 (0.25); Increased by ≥1 level 0.04 (0.05)	Based on SRM. Anchored by global rating of change (improved, unchanged, worse) and Chronic Pain Grade questionnaire grade (pain grade decreased by ≥1 level, pain grade at baseline = pain grade at follow-up, pain grade increased by ≥1 level) at 6 months

Study (ref)/ Mode of administration (*version*)	N Condition of Pain Time interval	Responsiveness Results	Approach(es) used to estimate Responsiveness
Stewart 2007[53] SAQ, on-site	134 Chronic whiplash 6 weeks	AUC (Pearson's r) NRS- pain intensity 0.68 (0.49) NRS- pain bothersomeness 0.70 (0.47) SF-36 Bodily Pain 0.73 (0.41) SRMs (ES) NRS- pain intensity Total cohort 0.64 (0.75); Improved 0.96 (1.03) NRS- pain bothersomeness Total cohort 0.98 (1.17); Improved 1.20 (1.40) SF-36 Bodily Pain Total cohort 0.48 (0.49); Improved 0.71 (0.72)	Based on SRMs, ES, ROC/AUC and Pearson's r. Anchored by global perceived effect scored on an 11-point numerical rating scale (-5 = vastly worse, 0 = unchanged, 5 = completely recovered) at 6 weeks
Keller 2007[39] SAQ	131 LBP First follow-up visit	Among improved patients BPI severity=-1.09 BPI interference=-1.13 GCPS intensity=-0.47 GCPS disability=-0.47 SF-36 BPS=0.69	Based on SRM
Wittink 2004[60] SAQ, on-site	87 Mostly back and neck NR[b]	ES MPI-S: -0.41; MPI-I: -0.42 Oswestry Disability Index -0.39 SF-36 Bodily Pain 0.44	Based on ES of differences between the baseline visit and post-treatment
Nilsdotter 2003[48] SAQ, on-site	62 OA, hip 6 months	SMRs HOOS pain: All patients 2.11; age ≤66 years 2.60; age >66 years 1.97 HOOS ADL: All patients 1.70; age ≤66 years 2.51; age >66 years 1.52 WOMAC pain: All patients 1.83; age ≤66 years 2.37; age >66 years 1.68 WOMAC function: All patients 1.70; age ≤66 years 2.51; age >66 years 1.52	Based on SRM
Gentelle-Bonnassie 2000[32] SAQ, on-site and mailed	80 Knee OA 6 months	SRMs *Intent-to-treat* VAS Month 1: -0.40; Month 3: -0.13; Month 6: -0.25 WOMAC Pain Month 1: -0.39; Month 3: -0.28; Month 6: -0.30 WOMAC Function Month 1: -0.37; Month 3: -0.15; Month 6: -0.09	Based on SRM
Single Studies by Pain Scale			
Brief Pain Inventory			
Chien 2013[25] SAQ, on-site	254 Chronic 10 days	AUC 0.71 BPI composite average SMRs BPI composite average All subjects 0.42; Improved 0.81; Unimproved 0.10	Based on SRM and ROC/AUC. Anchored by patient-reported rating of pain improvement ''Would you say that your pain has improved as a result of your treatment?'' (strongly disagree, disagree, neutral, agree, and strongly agree)

Study (ref)/ Mode of administration (*version*)	N Condition of Pain Time interval	Responsiveness Results	Approach(es) used to estimate Responsiveness
Tan 2004[55] SAQ, on-site	440 Chronic NR[b]	BPI-S, mean change P <.01 at all visits BPI-I, mean change P <.001 between visits 1 and 2, and visits 1 and 3. NS for visits 2 and 3.	Was assessed by using paired t tests to compare changes in the BPI scale scores across a span of 3 visits.
Numeric Rating Scale (range			
Godil 2015[33] Interview, off-site (*neck and arm versions*)	88 Neck and radicular arm 1 year	AUC NRS-neck pain 0.69; NRS-arm-pain 0.74 SRMs *Responders,* NRS-neck pain 0.95; NRS-arm-pain 0.97 *Non-responders* NRS-neck pain 0.49; NRS-arm-pain 0.38	Based on SRM and ROC/AUC. Anchored by Meaningful improvement versus not' (taken as the ''gold standard'' or the external criterion)
Oswestry Disability Index			
Anagnostis 2004[20] Unclear, on-site	230 Chronic disabled musculoskeletal disorder NR[b]	ES 0.95	Based on ES through comparison of pre- and post-treatment scores using paired t tests
Changulani 2009[23] SAQ, on-site	107 Lower back 6 weeks	SRM (ES) 0.84 (1.05)	Based on SRM and ES. Anchored by reported change in symptoms (much better, better, same, worse, much worse)
Patient Global Impression of Change			
Scott 2015[51] unclear, on-site	476 Back, upper body, other	ES Pain: 0.56 Physical function: 0.56	Based on within subject ES of differences between pre- and post-treatment means
PROMIS PI			
Askew 2016[21] SAQ, on-site	218 (175)[a] Lower back 3 months	SRMs Better -1.09; Same -0.26; Worse 0.44	Based on SRM. SRM ≥ 0.30 indicated responsiveness. Anchored by self-reported magnitude of changes (better, same, or worse) in overall pain scores
Deyo 2016[28] Interview, written survey	198 Musculoskeletal 3 months	SRM (ES) *Pain interference* Much better -1.07 (-1.03); Slightly better -0.29 (-0.28); Same -0.08 (-0.08); Slightly worse 0.18 (0.17); Much worse 0.74 (0.71)	Based on SRM and ES. Anchored by patient-reported global change (much better, slightly better, same, slightly worse, and much worse)
Roland Morris Disability Questionnaire			

Study (ref)/ Mode of administration (*version*)	N Condition of Pain Time interval	Responsiveness Results	Approach(es) used to estimate Responsiveness
Macedo 2011[44] SAQ, on-site (*24,18-item*[Williams and Myers (WM)], *18-item*[Stratford and Binkley (SB)], *11-item*)	461 Lower back Up to 1 year	AUC (cut-off of ≥3 global perceived effect units) 24-item 0.78; 18-item[WM] 0.78; 18-item[SB] 0.78; 11-item 0.75 ES 24-item 0.67; 18-item[WM] 0.75; 18-item[SB] 0.78; 11-item 0.65 GRI 24-item 1.55; 18-item[WM] 1.49; 18-item[SB] 1.52; 11-item 1.30	Based on ES, Guyatt's responsiveness index (GRI, calculated by dividing the mean change of patients who have improved by the standard deviation of change of patients reporting no improvement) and ROC/AUC. Anchored by global perceived effect (cut-off of 3 units was used to identify patients that improved and did not improve)
Chansirinukor 2005[24] SAQ, on-site (*18-item*)	143 Lower back 3 months	AUC 0.69 SRM (ES) 0.44 (0.44) SES 0.38	Based on SRM, SES, ES, and ROC/AUC Anchored by work status (working preinjury duties, full time; working preinjury duties, part-time or working other duties, full-time; working other duties, part time; and not working)

ES=effect size; HOOS=Hip Disability and Osteoarthritis Outcome Score; LBP=low back pain; MCID=minimum clinically important difference; MPQ=McGill Pain Questionnaire; NR=not reported; ROC=receiver operating characteristic curve (AUC area under the curve); SAQ=self-administered questionnaire; SEM=standard error of measurement; SES=standardized effect sizes; SRM=standardized response mean; VAS= visual analog scale; WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index

[a] Available at follow-up

[b] Post-treatment, no further details

REFERENCES

1. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. *Ann Acad Med Singapore.* 1994;23(2):129-138.

2. 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.

3. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res.* 2011;63 Suppl 11:S240-252.

4. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain.* 1992;50(2):133-149.

5. 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.

6. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes.* 2003;1:64.

7. Burckhardt CS, Jones KD. Adult Measures of Pain: The McGill Pain Questionnaire (MPQ), Rheumatoid Arthritis Pain Scale (RAPS), Short-Form McGill Pain Questionnaire (SF-MPQ), Verbal Descriptive Scale (VDS), Visual Analog Scale (VAS), and West Haven-Yale Multidisciplinary Pain Inventory (WHYMPI). *Arthritis Rheum.* 2003;49(5S):S96-S104.

8. McCaffery M, Beebe A. *Pain: Clinical Manural for Nursing Practice.* St. Louis, MO: Mosby, 1989.

9. Kerns RD, Turk DC, Rudy TE. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain.* 1985;23(4):345-356.

10. 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.

11. 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.

12. 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.

13. 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.

14. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol.* 2010;63(11):1179-1194.

15. 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.

16. 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.

17. 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.

18. 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.

19. Wong-Baker FACES® History. Available at: http://wongbakerfaces.org/us/wong-baker-faces-history/, Accessed 1 August 2017.

20. 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.

21. 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.

22. 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.

23. Changulani M, Shaju A. Evaluation of responsiveness of Oswestry low back pain disability index. *Arch Orthop Trauma Surg.* 2009;129(5):691-694.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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.

30. 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.

31. Gallasch CH, Alexandre NM. The measurement of musculoskeletal pain intensity: a comparison of four methods. *Rev Gaucha Enferm.* 2007;28(2):260-265.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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):414-421.

39. 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.

40. 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.

41. 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.

42. 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.

43. 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.

44. 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.

45. Maughan EF, Lewis JS. Outcome measures in chronic low back pain. *Eur Spine J.* 2010;19:1484-1494.

46. 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.

47. 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.

48. Nilsdotter 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.

49. 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.

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. 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.

52. 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.

53. 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.

54. 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.

55. 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.

56. 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.

57. 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.

58. 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.

59. 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.

60. 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.