HHS Public Access

Author manuscript

Med Care. Author manuscript; available in PMC 2017 April 01.

Published in final edited form as:

Med Care. 2016 April ; 54(4): 414–421. doi:10.1097/MLR.0000000000000497.

Comparative responsiveness of the PROMIS Pain Interference short forms, Brief Pain Inventory, PEG, and SF-36 Bodily Pain subscale

Jacob Kean[1,2], **Patrick Monahan**[3], **Kurt Kroenke**[4,5,6], **Jingwei Wu**[3], **Zhangsheng Yu**[3], **Tim Stump**[3], and **Erin E. Krebs**[7,8]

[1]VA Salt Lake City Health Care System, Salt Lake City, UT, United States

[2]Health System Innovation and Research, Department of Population Health Sciences, University of Utah School of Medicine, Salt Lake City, UT, United States

[3]Department of Biostatistics, Indiana University School of Medicine, Indianapolis, IN, United States

[4]Center for Health Information and Communication, Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service CIN 13-416, Richard L. Roudebush VA Medical Center, Indianapolis, IN, United States

[5]Regenstrief Institute, Indianapolis, IN, United States

[6]Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

[7]Center for Chronic Disease Outcomes Research, Minneapolis VA Health Care System, Minneapolis, MN, United States

[8]University of Minnesota Medical School, Minneapolis, MN, United States

Abstract

Purpose—To compare the sensitivity to change and the responsiveness to intervention of the PROMIS Pain Interference short forms, Brief Pain Inventory (BPI), 3-item PEG scale, and SF-36 Bodily Pain subscale in a sample of patients with persistent musculoskeletal pain of moderate severity.

Methods—Standardized response means, standardized effect sizes, and receiver operating curve (ROC) analyses were used to assess change between baseline and 3-month assessments in 250 participants who participated in a randomized clinical effectiveness trial of collaborative telecare management for moderate to severe and persistent musculoskeletal pain.

Results—The BPI, PEG, and SF-36 Bodily Pain measures were more sensitive to patient-reported global change than the PROMIS Pain Interference short forms, especially for the clinically improved group, for which the change detected by the PROMIS short forms was not

Correspondence to: Jacob Kean, PhD, VA Salt Lake City Health Care System, 500 Foothill Drive, Salt Lake City, UT 84148-0001. jacob.kean@gmail.com, (P) 801-213-3751..

The views expressed in this paper are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

statistically significant. The BPI was more responsive to the clinical intervention than the SF-36 Bodily Pain and PROMIS Pain Interference measures. Post-hoc analyses exploring these findings did not suggest that differences in content or rating scale structure (number of response options or anchoring language) adequately explained the observed differences in the detection of change.

Conclusions—In this clinical trial, the BPI and PEG measure were better able to detect change than the SF-36 Bodily Pain and PROMIS Pain Interference measures.

Introduction

Pain is the most common symptom reported in both the general population and in primary care.[1] Musculoskeletal pain is consistently the most common, disabling, and costly of all pain complaints, accounting for nearly 70 million outpatient visits in the U.S. each year.[1-3] **The** prevalence, associated disability and cost,[1, 3-5] large number of ongoing clinical trials, and pressure on clinicians to monitor pain as the "fifth vital sign" [6] all underscore the need for quality patient-reported outcome (PRO) measurement in pain.

The Patient-Reported Outcomes Measurement Information System (PROMIS) is a National Institutes of Health (NIH) Common Fund initiative to improve PRO measurement.[7, 8] The PROMIS assessment system is based on a comprehensive (i.e., physical, mental, social) self-reported health framework composed of many domains. Domains are represented as unidimensional hierarchies of dozens of items, called "item banks," which are scaled using Item Response Theory (IRT). Item banks include many items to represent fully the range of impairment in a given domain. Item banks can be administered adaptively or assembled as static "short forms" and compared meaningfully across populations. Pain outcome domains in the PROMIS assessment system include pain interference, pain severity, and pain behavior.

A critical psychometric property of pain outcome measures, particularly important in clinical settings and longitudinal studies, is their ability to detect change. Clinical change includes both responsiveness to treatment, or the ability of a measure to accurately detect change when it has occurred following a treatment of known efficacy, and sensitivity to change, which is sometimes used synonymously with responsiveness but which does not require a treatment to be present.[9] Studies comparing the responsiveness of legacy measures (i.e., older, established measures commonly developed in a classical test theory framework) have found comparable detection of change among pain-specific measures, including ultra-brief (≤ 4 items) measures, but greater responsiveness and sensitivity for pain-specific measures versus general health status measures.[10-17] Shahgholi et al.[18] compared the pre-post responsiveness of PROMIS short forms (pain behavior, pain interference and physical function) to that of legacy measures (Roland-Morris Disability Index, numerical rating scale) in a single-arm vertebroplasty intervention in a sample of 50 patients. Responsiveness was operationalized as a statistically significant change from baseline to 1 month in mean scores and similarity between measures on the categorization of patients into meaningful change groups (improvement, no change, decline) based on whether a participant's change exceeded a criterion of 1.96 * (√2 * SEM). Results indicated that mean score differences between baseline and 1-month assessment points were statistically significant for all

PROMIS and legacy measures ($p < 0.0001$) and that all measures were similar on the percentage of participants categorized into meaningful change groups.

Stepped Care to Optimize Pain Care Effectiveness (SCOPE),[19] a randomized clinical effectiveness trial of optimized analgesic therapy delivered by a collaborative telecare intervention in primary care, incorporated longitudinal assessment of pain severity, interference, and pain-related disability. We sought to assess pain interference measures for sensitivity to change (with respect to a retrospective global anchor item) in the combined control and intervention samples, and responsiveness to treatment (i.e., attributable to a multicomponent stepped care intervention provided in the SCOPE trial) compared to a control group. To our knowledge, this is the first study to compare responsiveness of PROMIS and legacy measures in the context of a randomized controlled trial.

Methods

Measures

PROMIS Pain Interference Short Forms—PROMIS Pain Interference short forms are composed of items from the unidimensional PROMIS Pain Interference item bank, one of eight PROMIS item banks that were designed to measure physical functioning as a component of overall self-reported health and validated in large clinical and community samples. The PROMIS Pain Interference items are scaled from 1-5, with **response options** "Not at all," "A little bit," "Somewhat," "Quite a bit," and "Very much," respectively. We assessed responsiveness of the 4-item pain interference short form included in the PROMIS-29 Profile instrument, the 8-item pain interference short form included on the PROMIS-57 Profile instrument, and the stand-alone 6-item PROMIS Short Form v1.0 – Pain Interference 6b. Short forms rather than adaptively administered item banks were chosen as these were more easily adopted into the study data collection process and may be more feasibly adopted in a diversity of clinical settings. Raw scores were converted to IRT-derived T-scores using the scoring manuals available at http://assessmentcenter.net

Brief Pain Inventory—The Brief Pain Inventory (BPI) is an 11-item pain measure validated initially in cancer populations [20] and later in other clinical populations.[13, 21] The BPI measures two domains of pain: a 4-item severity dimension and 7-item interference dimension. Both the BPI severity and interference items are scaled from 0-10, with **response options** at scale point 0 and 10, respectively, of "No pain" and "Pain as bad as you can imagine" for the severity items and of "Does not interfere" and "Interferes completely" for the interference items. We assessed responsiveness of the 4-item severity and the 7-item interference subscales separately and also together as an 11-item total.

PEG Scale: The ultra-brief 3-item PEG measure was derived from the BPI and validated in ambulatory care settings.[22] **The PEG** includes 1 severity item (average **pain**) and 2 interference items (**enjoyment** of life and **general** activity) and uses the same scales and scoring **(i.e., mean of item scores)** as the BPI.

SF-36 Bodily Pain—The 2-item Bodily Pain subscale of the Medical Outcomes Study SF-36 questionnaire has been validated in a wide range of populations.[23, 24] One item

assesses pain interference on a 1-5 scale, with **response options** "Not at all," "A little bit," "Moderately," "Quite a bit," and "Extremely." The other item assesses pain severity on a 1-6 scale, with **response options** "None," "Very mild," "Mild," "Moderate," "Severe," and "Very severe." We assessed the responsiveness of the two items together, as a scale. Responses are transformed into a 0 to 100 score.

Reference Standard – Patient-reported global change: A single item assessing patient-reported global change [25] from baseline to the 3-month assessment was used as the reference standard for responsiveness. The item asked, "How would you describe your pain now, compared to how you were when you started in our study?" The item was scaled from 1-7, with **response options** "Much better," "Moderately better," "A little better," "No change," "A little worse," "Moderately worse," and "Much worse," respectively.

Details of the SCOPE trial design and study measures have been published.[19] Briefly, the SCOPE trial enrolled veterans with moderate to severe and persistent musculoskeletal pain from one of five primary care clinics at the VA Medical Center in Indianapolis. Musculoskeletal pain was defined as regional (joints, limbs, back, neck) or more generalized (fibromyalgia or chronic widespread pain). Nonmusculoskeletal single-site pain syndromes (headache, facial pain, chest pain, abdominal pain) and inflammatory types of arthritis that would be treated with disease-modifying drugs (e.g., rheumatoid arthritis) were excluded. Moderate to severe pain was defined as a Brief Pain Inventory item score ≥5 for either the patient's average or worst pain in the past week. Persistent was defined as pain that continued for ≥3 months despite trying at least one analgesic medication. The study excluded those who did not speak English, had moderately severe cognitive impairment or psychoses, had a pending pain-related VA or Social Security disability claim, were actively suicidal, had current illicit drug use, or had an anticipated life expectancy of less than 12 months. The SCOPE study protocol was approved by the Indiana University Institutional Review Board and the VA Research Review Committee.

A total of 250 participants were enrolled in the SCOPE study. The study patients had a mean age of 55.1 years (range, 28 to 65); 83% were men; 77% were white, 19% black, and 4% other race. The mean baseline BPI total pain score was 5.2 **(1.8)**, representing a moderate level of pain. The duration of pain was 1 year or longer in 98% of participants. The results reported here represent the data from the 244 patients who completed both baseline and 3-month assessments.

Analyses

Sensitivity to change: Standardized Response Means and Area Under the Curve—Standardized response means (SRM) are effect size measures of within-group change that are computed by standardizing the difference between mean scores within a group at two time points.[26] SRMs compare the average change to the variability in change and thus show the size of the change relative to a unit of 1.0 SD:

$$\frac{\text{Mean}\,T1 - \text{Mean}\,T2}{\text{SD of change score}}$$

SRMs were used as the effect size to measure the magnitude of the sensitivity to change between baseline and 3 months on each scale in each of three global change reference groups: those who were worse (global change item categories 5-7), those who were the same (category 4), and those who were better (categories 1-3).

In addition, the ability of the measures to detect any improvement (global change < 4) and moderate improvement (global change < 3) was investigated using receiver operating characteristic (ROC) analyses. Area under the curve (AUC) values can range from 0.5 (the same as chance) to 1.0 (perfect discrimination) and are interpreted as the probability of a measure correctly discriminating between patients who have improved and those who have not.

Responsiveness to intervention: Standardized Effect Size—Standardized effect sizes (SES) are effect size measures of between-group difference of change that are computed by standardizing the difference between change-score means of two independent groups:[26]

$$\frac{(\text{Intervention Group Mean Change - Control Group Mean Change})}{\text{Pooled SD of change}}$$

SESs were used as the effect size to measure the magnitude of the responsiveness of the scales to the SCOPE collaborative telecare intervention. Primary results of the SCOPE trial showed the intervention was effective in reducing persistent musculoskeletal pain and pain-related functional limitations.[19] Cohen's rule of thumb interpretative guidelines for two-group standardized mean difference effect sizes (i.e., 0.2 is small, 0.5 is medium, and 0.8 is large)[27] were used to assess the magnitude of the effects.

Results

Sensitivity to change: Standardized Response Means and Area Under the Curve

Table 1 shows the measure-level sensitivity to change for pain measures in the SCOPE study, including baseline and 3-month scores for each measure and each (i.e., Worse, Same, Better) group, the change in scores between baseline and 3-months, the statistical significance of the change between reference (i.e. Same) and other (Worse, Better) groups, and corresponding SRM values. The independent-group *t*-test comparison between groups on the change scores from baseline to 3-months revealed that the Worse group showed statistically significant greater change on all measures relative to the Same group change scores. These *t*-tests showed that the Better group demonstrated statistically significant greater change than the Same group on all BPI variants (Severity, Interference, Total), the PEG scale, the SF-36 Bodily Pain subscale, and the **PROMIS** Pain Interference Short Form 6b, but not on the PROMIS Profile short forms (Profile-29, Profile-57).

Across all measures, effect size estimates of change **(**Table 1**)** for the Same (−0.38 to 0.38) and Worse (−0.47 to 0.17) groups were small. In the Better group, effect size estimates for all BPI variants (0.71 to 0.94), PEG (0.86), and SF-36 Bodily Pain (−0.71; reverse-scaled) were medium to large, whereas effect size estimates for all PROMIS variants were small to

medium (0.34 to 0.51). The overlapping confidence intervals indicate that few differences are statistically significant.

Table 2 shows the area under the ROC curve (AUC) for each measure. Consistent with the SRM analysis, all BPI variants and the PEG showed greater responsiveness than all PROMIS variants and the SF-36 Bodily Pain subscale. The responsiveness of BPI variants and the PEG to any global improvement (0.677 to 0.727) was similar to their responsiveness to moderate improvement (0.694 to 0.743). The responsiveness to any improvement (0.679) was also similar to that of moderate improvement (0.644) for the SF-36 Bodily Pain subscale, as were the ranges of PROMIS short form AUC values to any (0.562-0.612) and moderate improvement (0.597-0.661).

Responsiveness to intervention: Standardized Effect Sizes

Table 3 shows measure-level responsiveness to intervention for treatment and control groups, including group change and corresponding standardized effect size estimates for all pain measures. All BPI variants (0.37 to 0.42) and the PEG (0.37) were more responsive to the SCOPE intervention than PROMIS variants (0.17 to 0.28) and the SF-36 Bodily Pain subscale (−0.25; reverse-scaled).

Post-hoc analyses of item-level responsiveness and BPI rating scales

Post-hoc analyses of sensitivity to change and responsiveness to intervention at the individual item level were undertaken to investigate whether the differences in item content could help explain the measure-level findings. Item-level results were generally consistent with the measure-level results (item results available from authors). Across all items, effect size estimates of sensitivity to change for the Same (−0.36 to 0.41) and Worse (−0.40 to 0.19) groups ranged between small and medium. In the group that reported overall improvement in pain at 3 months, effect size estimates for all BPI items (0.32 to 0.92) and SF-36 Bodily Pain (−0.58 to −0.66; reverse-scaled) generally ranged from medium to large, whereas effect size estimates for all PROMIS items ranged from small to medium (0.16 to 0.40).

Table 4 displays BPI, PROMIS, and SF pain interference short form item content, ordered from most to least similar, to facilitate comparison of item-level sensitivity to change. Comparison of SRM values shows that BPI and SF items were each more sensitive to change than any PROMIS item, regardless of item content. Unique BPI items (i.e., mood, walking ability, sleep) were no more sensitive to change than BPI items with other content. Similarly, the unique PROMIS item (i.e., ability to concentrate) was not less sensitive to change than PROMIS items with other content. Conceptually similar BPI and SF items were more responsive than corresponding PROMIS items.

Table 5 shows item-level responsiveness to intervention for treatment and control groups, including group change and corresponding standardized effect size estimates for all items. The intervention effect size ranged from small to medium for all items. However, BPI items (0.11 to 0.56) were generally more responsive to the SCOPE intervention than PROMIS (−0.05 to 0.27) and the SF-36 Bodily Pain items (−0.17 to −0.25; reverse-scaled).

Discussion

Using longitudinal data collected in the SCOPE effectiveness trial, we found that PROMIS Pain Interference short forms were less sensitive to change (with respect to a global retrospective anchor) and less responsive to treatment than BPI severity, interference, total, and PEG scales, especially in the group who demonstrated global improvement. The differences in sensitivity to change and responsiveness to treatment held at the item level, as well, and did not appear to result from content differences between the short forms. The overlapping confidence intervals indicated that few differences were statistically significant, which given the magnitude of the differences in effect size, is likely a function of the size of the study sample. The performance of PROMIS Pain Interference short forms relative to the SF-36 Bodily Pain subscale depended on the outcome: the PROMIS short forms were less sensitive to change than the SF-36 Bodily Pain subscale (i..e, SRM analyses using global change categories) but similarly responsive to the SCOPE intervention in both measure- and item-level analyses (i.e., SES analyses using between-group treatment effects). Decreased responsiveness and less sensitivity to change renders measures less capable of registering improvement or decline.

The poorer sensitivity to change of PROMIS pain interference measures in the study may be attributable to the duration of the study interview and to the order of presentation of measures. The BPI was administered near the beginning of the 30-45 minute interview, whereas the PROMIS and SF-36 Bodily Pain items were administered near the end of the interview. Respondent fatigue arising from the length of interview overall or the large number of pain items in particular could have led to less discrimination in participants' responses to PROMIS pain items. However, subjects were informed at the beginning of the interview that some questions may seem redundant but was done intentionally for research purposes. Moreover, the respondent burden was similar to that in several previous trials showing comparable responsiveness of pain measures.[13, 14] Additionally, the respondent burden in the current study was certainly not greater than that in the item bank validation studies undertaken by the PROMIS group. Perhaps most importantly, the SF-36 Bodily Pain subscale items were adjacent to the PROMIS items near the end of the interview but showed greater sensitivity to change than PROMIS items. It is also important to note that the structure of the interview, in which the global change item fell immediately after PROMIS and SF-36 Bodily Pain items, would seem to promote concordance and, thus, greater sensitivity to change rather than less. Taken together, the interview duration and structure do not appear to be viable explanations for the differences in sensitivity to change found in this study. Although the pain measures in SCOPE were not self-administered bur rather completed by interviewer-administration, the available evidence suggests that mode of administration has minimal, if any, effect on the psychometric properties of PROMIS measures.[28]

Another potential explanation for differences in sensitivity to change is the possibility of an artifactual relationship based upon similarities between the response structures of the study measures and the reference standard for change. However, the opposite was true: there was a greater degree of association between scores of the global change (7 verbally anchored response options) and the BPI (11 numeric response options anchored only at the tails)

despite the similarities between the response structures of the global change, PROMIS (5 verbally anchored response options) and SF-36 Bodily Pain (5 or 6 verbally anchored response options) measures. Thus, differences in response structures cannot explain differences found in the sensitivity of the measures to change.

The greater responsiveness to treatment of BPI may have resulted from the use of two BPI items (worst pain and average pain) for baseline eligibility and use of the PEG, a 3-item scale derived from the BPI, in the automated symptom monitoring (ASM) in the intervention arm. More responsiveness could arguably be expected in the questionnaire that was used to constrain baseline eligibility. In addition, treatment may have been adjusted to optimize the PEG score and the BPI by virtue of the shared content between the measures, thus optimizing its change relative to other pain scales. Also, it is possible that the intervention group who received automated symptom monitoring as a part of treatment were "trained" on the 0-10 scale. However, in previous trials where automated symptom monitoring was not done (i.e., SCAMP trial) or was done with a 5-item BPI (i.e., INCPAD trial), the differences in BPI and non-BPI pain scale responsiveness were **subtler than** those found in the present study.[29, 30] Nevertheless, we acknowledge that this study was not designed to compare these scales head-to-head, with random ordering; thus the issues considered above render the results somewhat less robust to alternative interpretations.

The poorer responsiveness to treatment of the PROMIS pain interference short forms relative to the BPI, PEG, and SF measures contrast with findings reported by Shahgholi et al. Some degree of difference between studies might be expected between the present study and that by Shahgholi et al. as a result of differences in design (randomized, controlled trial vs. consecutive cohort), differences in the type and chronicity of pain (i.e., chronic and commonly multisite musculoskeletal pain vs. vertebral fracture), differences in the indicators of change (patient global impression vs. statistical thresholds), and difference in sample size (244 vs. 50). However, it seems more likely that these elements could explain between-study rather than within-study differences.

The data in this study strongly suggest that the unique item content of the BPI (i.e., mood, sleep and walking) does not explain the differences in sensitivity to change and responsiveness to treatment between PROMIS short forms and BPI interference items. **A**ll BPI items, not just the unique items (i.e., mood, sleep and walking) had larger SRM values than PROMIS short form items for the clinically improved group. Conceptual differences may be a significant driver of the differences in responsiveness. It may be the case that instruments measuring symptom severity are more responsive to treatment effects than instruments measuring the impact/interference of the symptom on day-to-day activities as the latter are more distal to what is actually changing.

PROMIS pain interference measures were validated in a general population sample augmented with a sample of participants with chronic pain to ensure reporting higher levels of pain. Still, it may be possible that these PROMIS short form items exhibit less heterogeneity of responsiveness in clinical populations. Short forms composed of items that maximize precision of measurement across the full continuum of pain could potentially contribute to a lack of heterogeneity if change occurs in a restricted range of the continuum.

Direct item-level comparison of PROMIS, BPI and SF-36 Bodily Pain content suggests the differences in responsiveness to intervention and sensitivity to change are difficult to attribute to differences in the conceptual targets of items. Special factors about clinical samples might also make a difference in responsiveness: receiving a pain intervention if in a trial; higher rates of important pain comorbidities such depression; pain type, bodily location, or chronicity of pain; and clinical population (e.g., primary care, cancer, specialty pain clinics, etc.). It is important to note that the effect sizes reported here are based on small sample sizes and are dependent on the variances of the measures in this particular sample. Interpretation and conclusions about the generalizability of the findings should be made cautiously.

As the detection of sensitivity to change is important for cohort studies, and the detection of responsiveness to treatment is essential for clinical trials, future studies of PROMIS pain interference scales are sorely needed prior to adoption of these measures as primary research outcomes. Future studies should consider counter-balancing the order of administration of measures and including both static and adaptive item administration. In the future, alternatives to the effect size estimates reported here, such as an SEM-based approach or the assessment of change at the individual level in the context of an IRT model, could be considered.

References

1. Medicine, Io. Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research. National Academies Press; Washington, DC: 2011.
2. Schappert, SM. National Ambulatory Medical Care Survey: 1992 Summary. US Department of Health and Human Services, Public Health Service, Center for Disease Control and Prevention, National Center for Health Statistics; 1994.
3. Magni G, Marchetti M, Moreschi C, et al. Chronic musculoskeletal pain and depressive symptoms in the National Health and Nutrition Examination I. Epidemiologic follow-up study. Pain. 1993; 53(2):163–168. [PubMed: 8336986]
4. Disorders NRCPoM, Workplace t, Medicine Io. Musculoskeletal Disorders and the Workplace: Low Back and Upper Extremities. Natl Academy Pr; 2001.
5. Turk DC, Dworkin RH, Revicki D, et al. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. Pain. 2008; 137(2):276–285. [PubMed: 17937976]
6. Lorenz KA, Sherbourne CD, Shugarman LR, et al. How reliable is pain as the fifth vital sign? The Journal of the American Board of Family Medicine. 2009; 22(3):291–298. [PubMed: 19429735]
7. 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. Journal of clinical epidemiology. 2010; 63(11):1179–1194. [PubMed: 20685078]
8. Gershon RC, Rothrock N, Hanrahan R, et al. The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. Journal of applied measurement. 2010; 11(3):304. [PubMed: 20847477]
9. Löwe B, Unützer J, Callahan CM, et al. Monitoring depression treatment outcomes with the patient health questionnaire-9. Medical Care. 2004; 42(12):1194–1201. [PubMed: 15550799]
10. Angst F, Aeschlimann A, Steiner W, et al. Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. Annals of the rheumatic diseases. 2001; 60(9):834–840. [PubMed: 11502609]

11. Bronfort G, Bouter LM. Responsiveness of general health status in chronic low back pain: a comparison of the COOP charts and the SF-36. Pain. 1999; 83(2):201–209. [PubMed: 10534591]
12. Dworkin R, Peirce-Sandner S, Turk D, et al. Outcome measures in placebo-controlled trials of osteoarthritis: responsiveness to treatment effects in the REPORT database. Osteoarthritis and Cartilage. 2011; 19(5):483–492. [PubMed: 21396467]
13. Krebs EE, Bair MJ, Damush TM, et al. Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. Medical Care. 2010; 48(11):1007–1014. [PubMed: 20856144]
14. Kroenke K, Theobald D, Wu J, et al. Comparative Responsiveness of Pain Measures in Cancer Patients. The Journal of Pain. 2012
15. Nilsdotter AK, Lohmander LS, Klässbo M, et al. Hip disability and osteoarthritis outcome score (HOOS)–validity and responsiveness in total hip replacement. BMC Musculoskeletal Disorders. 2003; 4(1):10. [PubMed: 12777182]
16. Nilsdotter AK, Roos EM, Westerlund JP, et al. Comparative responsiveness of measures of pain and function after total hip replacement. Arthritis Care & Research. 2001; 45(3):258–262. [PubMed: 11409667]
17. Wiebe S, Guyatt G, Weaver B, et al. Comparative responsiveness of generic and specific quality-of-life instruments. Journal of clinical epidemiology. 2003; 56(1):52–60. [PubMed: 12589870]
18. Shahgholi L, Yost K, Kallmes D. Correlation of the National Institutes of Health Patient Reported Outcomes Measurement Information System scales and standard pain and functional outcomes in spine augmentation. American Journal of Neuroradiology. 2012; 33(11):2186–2190. [PubMed: 22700753]
19. Kroenke K, Krebs E, Wu J, et al. Stepped Care to Optimize Pain care Effectiveness (SCOPE) trial study design and sample characteristics. Contemporary clinical trials. 2013; 34(2):270–281. [PubMed: 23228858]
20. Cleeland C, Ryan K. Pain assessment: global use of the Brief Pain Inventory. Annals of the Academy of Medicine, Singapore. 1994; 23(2):129–138.
21. Tan G, Jensen MP, Thornby JI, et al. Validation of the Brief Pain Inventory for chronic nonmalignant pain. The Journal of Pain. 2004; 5(2):133–137. [PubMed: 15042521]
22. Krebs EE, Lorenz KA, Bair MJ, et al. Development and initial validation of the PEG, a three-item scale assessing pain intensity and interference. Journal of general internal medicine. 2009; 24(6): 733–738. [PubMed: 19418100]
23. McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Medical Care. 1993; 31(3):247–263. JR. [PubMed: 8450681]
24. Ware JE, Gandek B. The SF-36 Health Survey: development and use in mental health research and the IQOLA Project. International Journal of Mental Health. 1994
25. Fischer D, Stewart AL, Bloch DA, et al. Capturing the patient's view of change as a clinical outcome measure. JAMA. 1999; 282(12):1157–1162. [PubMed: 10501119]
26. McDowell, I. Measuring health: a guide to rating scales and questionnaires. 3rd ed.. Oxford University Press; New York: 2006.
27. Cohen J. A power primer. Psychological bulletin. 1992; 112(1):155. [PubMed: 19565683]
28. Bjorner JB, Rose M, Gandek B, et al. Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. Journal of clinical epidemiology. 2014; 67(1):108–113. [PubMed: 24262772]
29. Kroenke K, Bair MJ, Damush TM, et al. Optimized antidepressant therapy and pain self-management in primary care patients with depression and musculoskeletal pain: a randomized controlled trial. JAMA. 2009; 301(20):2099–2110. [PubMed: 19470987]
30. Kroenke K, Theobald D, Wu J, et al. Effect of telecare management on pain and depression in patients with cancer: a randomized trial. JAMA. 2010; 304(2):163–171. [PubMed: 20628129]

Table 1

Measure-level sensitivity to change for pain measures in SCOPE trial (overall n = 244: worse n = 51; same n = 96; better n = 97)

Pain measure Global category	Baseline score (SD)	3 month score (SD)	Change score (SD)*	P-value†	SRM‡ (95% CI)
BPI Severity					
Worse	5.10 (1.42)	5.76 (1.50)	−0.66 (1.40)	<0.001	−0.47 (−0.75, −0.19)
Same	5.19 (1.77)	5.02 (1.89)	0.17 (1.35)	--	0.13 (−0.08, 0.33)
Better	5.12 (1.78)	3.89 (2.09)	1.23 (1.72)	<0.001	0.71 (0.51, 0.92)
BPI Interference					
Worse	5.45 (2.21)	5.40 (2.25)	0.05 (1.82)	0.044	0.03 (−0.25, 0.31)
Same	5.38 (2.26)	4.73 (2.29)	0.70 (1.86)	--	0.38 (0.18, 0.58)
Better	5.18 (2.18)	3.48 (2.26)	1.70 (1.81)	<0.001	0.94 (0.74, 1.14)
BPI Total					
Worse	5.28 (1.59)	5.58 (1.63)	−0.30 (1.36)	0.002	−0.22 (−0.50, 0.06)
Same	5.28 (1.86)	4.85 (1.99)	0.44 (1.41)	--	0.31 (0.11, 0.51)
Better	5.15 (1.82)	3.69 (2.02)	1.47 (1.57)	<0.001	0.93 (0.73, 1.14)
PEG					
Worse	5.33 (1.96)	5.56 (2.00)	−0.23 (1.68)	0.028	−0.14 (−0.42, 0.15)
Same	5.41 (2.13)	4.92 (2.20)	0.48 (1.92)	--	0.25 (0.05, 0.45)
Better	5.36 (2.12)	3.70 (2.28)	1.66 (1.93)	<0.001	0.86 (0.66, 1.06)
SF-36 Bodily Pain					
Worse	33.78 (15.89)	31.33 (14.35)	2.45 (14.83)	0.002	0.17 (−0.12, 0.45)
Same	33.78 (16.01)	38.95 (16.52)	−5.18 (13.76)	--	−0.38 (−0.58, −0.17)
Better	35.80 (17.23)	47.65 (20.12)	−11.85 (16.60)	0.003	−0.71 (−0.92 −0.51)
PROMIS®-29 Profile Pain Interference Short Form					
Worse	60.72 (7.33)	61.35 (7.05)	−0.63 (5.45)	0.020	−0.11 (−0.40, 0.17)
Same	60.47 (7.24)	58.80 (7.25)	1.67 (5.71)	--	0.29 (0.09, 0.49)
Better	59.00 (6.99)	57.05 (7.44)	1.95 (5.87)	0.730	0.33 (0.13, 0.53)
PROMIS®-57 Profile Pain Interference Short Form					
Worse	60.41 (7.56)	61.23 (7.37)	−0.82 (5.22)	0.010	−0.16 (−0.44, 0.13)
Same	60.26 (7.75)	58.64 (7.55)	1.63 (5.50)	--	0.30 (0.09, 0.50)
Better	59.00 (7.04)	56.79 (7.69)	2.20 (5.92)	0.483	0.37 (0.17, 0.57)
PROMIS® Pain Interference Short Form 6b					
Worse	60.70 (7.79)	60.83 (6.59)	−0.14 (5.55)	0.010	−0.02 (−0.31, 0.26)
Same	60.01 (7.39)	58.67 (7.21)	1.34 (4.95)	--	0.27 (0.07, 0.47)
Better	59.08 (6.73)	56.08 (7.60)	3.01 (5.92)	0.036	0.51 (0.31, 0.71)

[*] change calculated as (baseline - 3 month).

[†] p-value derived from 2-sided independent t-test comparing change scores between reference (i.e. “Same) and other groups

[‡] Standardized response mean, an effect size of the magnitude of within group change, calculated as (baseline – 3 month) / SD change score.

Table 2

Area under the receiver operating characteristic curve (AUC) for pain measures (n=244) in SCOPE trial

	Accuracy for detecting any improvement		Accuracy for detecting moderate improvement	
	AUC	**SE**	**AUC**	**SE**
BPI Severity	.727	.033	.737	.041
BPI Interference	.677	.035	.694	.042
BPI Total	.727	.033	.743	.041
PEG	.707	.034	.716	.043
SF-36 Bodily Pain	.679	.036	.644	.047
PROMIS®-29 Profile Pain Interference Short Form	.562	.038	.597	.044
PROMIS®-57 Profile Pain Interference Short Form	.570	.038	.608	.045
PROMIS® Pain Interference Short Form 6b	.612	.037	.661	.044

AUC is probability of correctly discriminating between patients who have improved and those who have not. Any improvement ≥ “a little better”; moderate improvement ≥ “moderately better”

Table 3
Measure-level responsiveness to SCOPE trial intervention

Pain scale	Intervention change* (SD) (n = 120)	Control change* (SD) (n = 124)	Standardized Effect Size (SES)†	Cohen's d Effect Size**	Cronbach's Alpha - Full Sample
BPI severity	0.74 (1.83)	0.11 (1.46)	0.38	0.37	0.872
BPI interference	1.33 (1.94)	0.61 (1.87)	0.37	0.33	0.880
BPI total	1.04 (1.70)	0.36 (1.45)	0.42	0.38	0.909
PEG	1.18 (2.07)	0.44 (1.89)	0.37	0.35	0.764
SF-36 pain	−8.24 (16.13)	−4.29 (15.76)	−0.25	−0.24	0.717
PROMIS®-29 Profile Pain Interference Short Form	1.81 (5.67)	0.81 (5.88)	0.17	0.14	0.888
PROMIS®-57 Profile Pain Interference Short Form	2.05 (5.54)	0.67 (5.81)	0.24	0.19	0.946
PROMIS® Pain Interference Short Form 6b	2.48 (5.27)	0.94 (5.79)	0.28	0.21	0.919

*change calculated as (baseline - 3 month).

†SES calculated as (intervention group change – control group change) / pooled change score SD.

**Cohen's d calculated as (intervention group change – control group change) / pooled baseline SD.

Table 4

BPI, PROMIS® and SF-36 pain interference Items, ordered by content from identical to moderately related to unique with corresponding standardized response mean values for the clinically improved group

BPI Interference		**PROMIS® Pain Interference**		**SF-36 Bodily Pain (reverse-scaled)**	
What number best describes how, in the last week, pain has interfered with your …	SRM‡	In the past 7 days, how much did pain interfere with your …	SRM‡	During the past 4 weeks, how much did pain interfere with your	**Standardized Response Mean** (SRM)‡
14. enjoyment of life	0.66	4. enjoyment of life	0.16		
		10. enjoyment of recreational activities	0.32		
		5. the things you usually do for fun	0.25		
8. general activity	0.71	1. day to day activities	0.34		
11. normal work (includes both work outside the home as well as housework)	0.76	2. work around the home	0.26	8. normal work, including both work outside the home and housework	−0.66
		7. household chores	0.37		
		11. tasks away from home (e.g., getting groceries, running errands)	0.40		
12. relations with other people	0.47	3. ability to participate in social activities	0.28		
		6. enjoyment of social activities	0.17		
		8. family life	0.29		
		12. How often did pain keep you from socializing with others? (frequency)	0.33		
9. mood	0.65				
10. walking ability	0.56				
13. sleep	0.42				
		9. your ability to concentrate	0.33		

‡**SRM calculated as (baseline – 3 month) / SD change score; change calculated as (baseline - 3 month).**

Table 5
Item-level responsiveness to SCOPE trial intervention

Pain item		Intervention change* (SD) (n = 120)	Control change* (SD) (n = 124)	Standardized Effect Size (SES)†
BPI Severity				
bpi1	*worst*	1.68 (2.26)	0.46 (1.92)	0.56
bpi2	*least*	0.18 (2.02)	−0.04 (1.96)	0.11
bpi3	*average* ‡	0.62 (2.10)	0.06 (1.69)	0.29
bpi4	*right now*	0.38 (2.42)	−0.06 (2.04)	0.20
BPI Interference				
bpi8	*general activity* ‡	1.40 (2.97)	0.62 (2.78)	0.27
bpi9	*mood*	1.44 (3.02)	0.70 (2.72)	0.26
bpi10	*walking*	1.57 (3.09)	0.52 (2.36)	0.37
bpi11	*normal work*	1.19 (2.93)	0.57 (2.75)	0.22
bpi12	*relations with other people*	0.88 (3.10)	0.44 (2.59)	0.15
bpi13	*sleep*	1.42 (2.70)	0.80 (2.86)	0.22
bpi14	*enjoyment life* ‡	1.45 (2.79)	0.64 (2.82)	0.29
PROMIS® pain				
promp1	*day to day activities*	0.34 (1.02)	0.17 (1.01)	0.17
promp2	*work around home*	0.26 (1.16)	0.10 (1.06)	0.14
promp3	*participation in social activity*	0.20 (1.27)	0.01 (1.22)	0.15
promp4	*enjoyment of life*	0.15 (1.09)	0.20 (1.11)	−0.05
promp5	*things you usually do for fun*	0.28 (1.14)	0.12 (1.24)	0.13
promp6	*enjoyment of social activities*	0.24 (1.31)	−0.01 (1.11)	0.21
promp7	*household chores*	0.34 (1.13)	0.06 (0.91)	0.27
promp8	*family life*	0.23 (1.08)	0.11 (1.02)	0.11
promp9	*ability to concentrate*	0.28 (1.09)	0.20 (1.22)	0.07
promp10	*enjoyment recreational activity*	0.40 (1.23)	0.06 (1.28)	0.27
promp11	*tasks away from home*	0.33 (1.15)	0.17 (1.19)	0.13
promp12	*socializing with others*	0.13 (1.07)	−0.02 (1.14)	0.14
SF pain				
sf8	*pain interfere with normal work*	−0.36 (0.94)	−0.19 (1.03)	−0.17
sf9	*how much bodily pain*	−0.48 (0.95)	−0.24 (0.86)	−0.25

*change calculated as (baseline - 3 month).

†SES calculated as (intervention group change – control group change) / pooled SD.

‡Items included in the PEG scale.