



Published in final edited form as:

J Clin Epidemiol. 2016 May ; 73: 103–111. doi:10.1016/j.jclinepi.2015.08.035.

Clinical Validity of PROMIS® Pain Interference and Pain Behavior in Diverse Clinical Populations

Robert L. Askew¹, Karon F. Cook², Dennis A. Revicki³, David Cella², and Dagmar Amtmann¹

¹Department of Rehabilitation Medicine, University of Washington School of Medicine, Seattle, WA, USA; Present address: Institute for Public Health and Medicine Northwestern University Feinberg School of Medicine, Chicago, IL, USA

²Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

³Outcomes Research, Evidera, Bethesda, MD, USA

Abstract

Objective—To evaluate clinical validity, including responsiveness, of PROMIS® Pain Interference (PROMIS-PI) and Pain Behavior (PROMIS-PB) T-scores.

Study Design and Setting—Data were aggregated from longitudinal studies of cancer, chronic low back pain (cLBP), rheumatoid arthritis (RA), chronic obstructive pulmonary disease (COPD), and major depressive disorder (MDD). Linear mixed-effects models were used to compare baseline score differences and score changes over time. We calculated standardized response means (SRMs) for subgroups defined by self-reported change in general health and pain.

Results—1357 individuals participated at baseline and 1225 at follow-up. Hypotheses of significant change in PROMIS-PI and PROMIS-PB scores were supported in the intervention groups (cLBP and MDD). Differences in baseline scores for COPD-exacerbators compared to stable-COPD patients were in the hypothesized direction but were not statistically significant. Subgroups reporting better health showed corresponding negative SRM values supporting responsiveness of T-scores to improvement. Responsiveness to decrements was supported in some but not all clinical groups and varied by anchor. More congruent values were obtained when using a pain-specific anchor.

Address correspondence to: Dagmar Amtmann, Department of Rehabilitation Medicine, Box 357920, University of Washington School of Medicine, Seattle, WA 98195, Phone: 206-543-4741, Fax: 206-685-9224, dagmara@u.washington.edu.

CONFLICT OF INTEREST

Robert L. Askew: None

Karon F. Cook is an unpaid officer of the PROMIS Health Organization

Dennis A. Revicki: None

David Cella is an unpaid member of the board of directors and officer of the PROMIS Health Organization

Dagmar Amtmann: None

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusion—This study provides evidence that PROMIS-PI and –PB scores are sensitive to changes in pain in studies of interventions expected to impact pain. The results inform estimation of meaningful change and support power analyses for comparative effectiveness research.

Keywords

chronic pain; PROMIS; pain interference; pain behavior; cancer; back pain

INTRODUCTION

Approximately 100 million people in the United States experience chronic pain, and pain costs up to \$635 billion dollars annually in treatment and lost productivity [1] In order to develop effective treatments for pain and to evaluate treatment effectiveness, researchers and clinicians need psychometrically sound and clinically valid instruments for measuring different aspects of pain. Quality criteria for the measurement properties of health status measures have been developed that can inform selection of outcome measures in clinical research. Critical features include a measure's reliability, reproducibility, and validity evidence (including responsiveness). [2] Recent guidelines for clinical pain research underscore the importance of these psychometric characteristics and emphasize the need for comprehensive health assessment involving multiple facets of patient health-status. [3]

Pain is a multidimensional construct. Two important aspects of pain are pain interference and pain behavior. Pain interference is the degree to which pain interferes with an individual's daily activities,[4] and it is increasingly recognized as an important facet of patients' pain experiences. Pain interference has been recommended as a core outcome in clinical trials of pain treatments. [3] Pain behavior, defined as behavior that typically indicates to others that an individual is experiencing pain, [4–6] can include both verbal (e.g., asking for help, sighing, moaning) and non-verbal (e.g., grimacing, resting, guarding) behaviors. Pain behaviors can be protective by eliciting assistance or support after a precipitating event (e.g. trauma, surgery), but when they are maintained beyond rehabilitation and recovery, pain behaviors can contribute to subsequent physical and psychosocial disability, [7] making them useful targets for behavioral interventions. [6, 8]

The NIH-funded Patient Reported Outcomes Measurement Information System (PROMIS®) has developed a family of instruments that can be used to measure different aspects of physical, mental, and social health. [9] More information on the development, validation, and implementation of all PROMIS measures can be found at www.nihpromis.org, and options regarding fixed-length and customized short forms and CAT administration can be found at www.assessmentcenter.net. PROMIS Pain Interference (PROMIS-PI) [10] and Pain Behavior (PROMIS-PB) [11] item banks were developed using modern psychometric methods, [12, 13] and the psychometric properties of these measures have been previously evaluated in a large cross-sectional sample that included both healthy people and people with various chronic conditions. [10, 11] PROMIS-PI measures also have been evaluated in ambulatory cancer care, [14] individuals with disabilities, [15] inflammatory bowel disease, [16] and arthritis, [17] among others. [18–20] The PROMIS-PB also has been evaluated in a range of clinical populations, but in fewer than the PROMIS-PI. [20, 21]

Though PROMIS pain measures have potential advantages ensuing from the advanced methods with which they were developed and are administered and scored, they are relatively new compared to other, more extensively evaluated pain measures such as Medical Outcomes Short Form Bodily Pain scale, [22] Brief Pain Inventory, [23] and the 3-item pain scale (PEG).[24] If the PROMIS measures are to be considered for clinical trials and comparative effectiveness research, the clinical validity of the scores in a wide variety of clinical contexts and for different purposes needs to be established. The current paper reports psychometric evaluations of PROMIS-PI and -PB scores in the context of five longitudinal studies of participants with one of five chronic conditions: chronic low back pain (cLBP), cancer, chronic obstructive pulmonary disease (COPD), major depressive disorder (MDD), and rheumatoid arthritis (RA).

Significant changes in PROMIS-PI and -PB scores were hypothesized for those receiving intervention (cLBP and MDD). [25] We also expected that baseline scores would be higher for patients with COPD who had experienced an exacerbation compared to scores of those with stable COPD. We expected standardized response means (SRMs) to distinguish among participants by self-reported clinical status (worse, about same, better).

METHODS

Data for this study were collected in longitudinal validation studies conducted by PROMIS investigators with samples of participants with cLBP, cancer, COPD, MDD, and RA. Time between baseline and follow-up varied by clinical sample (i.e., cLBP: 3 months, cancer: 2 months, COPD: 3 months, MDD: 3 months, RA: 12 months).

Measures

PROMIS Pain Interference and Pain Behavior—Two PROMIS pain measures were used in the current study, PROMIS-PI and PROMIS-PB. Like other PROMIS measures, these measures are based on banks of items calibrated using the graded response model that estimates item location (severity) and discrimination (ability to distinguish among people with different levels of the pain outcome. [26] PROMIS item banks were developed using qualitative and quantitative methods. [10–12, 27–29] The items have a 7-day time-frame.

The use of item banks for measures allows the development of short forms or more flexible administration using computer adaptive testing (CAT), a tailored approach in which the items administered are selected based on individuals' responses to previous items. [12] Scores from short forms and those generated using CAT are on a common mathematical metric. For the current study, CAT was used in the cLBP, COPD, and MDD cohorts. PROMIS short forms (Version 1) were used to assess PROMIS-PI and -PB in the arthritis and cancer cohorts. Full PROMIS-PI and -PB item banks are reported in the online appendix.

Clinical Anchors—In addition to standard clinical and demographic descriptors, participants also rated single items related to general health and pain, which served as anchors for subgroup comparisons between those classified as “better”, “about the same” and “worse”.

Samples

Detailed descriptions of recruitment, eligibility criteria, and treatments are available in the accompanying introductory article in this issue. [30] Briefly, participants with cLBP were recruited from the University of Washington Spine Center in Seattle and local recruitment sites. All had cLBP for at least six weeks and received a spinal injection. Participants with MDD were recruited from outpatient treatment clinics at Western Psychiatric Institute and Clinic, Pittsburgh, PA and its affiliates and received treatment in the form of antidepressants, psychotherapy, or both. Participants with COPD had a 10 pack/year history of smoking (i.e., packs per day multiplied by number of years) and met the Global Initiative for Chronic Obstructive Lung Disease clinical criteria for COPD. [31] They were recruited from multiple participating institutions including the University of North Carolina, North Shore University Health System, The University of Pittsburgh, and Duke University. Participants with COPD-related exacerbations at baseline were included and compared to those with stable COPD who had been exacerbation-free for 2 or more months prior to enrollment. Participants in the cancer study were recruited from North Shore University Health System in Chicago, IL and were administered PROMIS measures in an observational study setting with multiple heterogeneous treatment modalities. Participants were enrolled irrespective of treatment status (i.e. before or after starting treatment regimen). The RA study also was observational. Participants were recruited from multiple sources including the Aging Medical Information System (ARAMIS) and the Stanford Rheumatoid Arthritis Registry. Administration of PROMIS measures was intended to evaluate longitudinal changes in this clinical population, and even though RA is known to worsen over time, all participants received routine clinical care that at times included intervention. Like the cohort with cancer, some participants with RA were expected to improve while others were expected to deteriorate due to the heterogeneity of clinical characteristics and treatment status. No specific hypotheses were developed for the RA and cancer samples, but responsiveness was calculated based on change status at follow-up.

Statistical Analyses

The PROMIS-PI was administered at baseline and follow-up in studies of patients with cLBP, cancer, COPD, MDD, and RA. The PROMIS-PB was administered only in the studies of patients with cLBP, COPD, and MDD (baseline and follow-up). We used linear mixed effects models to estimate average change in scores over time. To account for similarity of repeated measurements within individuals, the models were estimated with random subject effects. [32, 33] Risk of bias from problematic missing data was examined in each clinical sample by assessing relationships between scores on baseline measures and attrition. This risk was found to be minimal, and the data were considered missing at random (MAR) for all analyses. This is advantageous because, when data are MAR, single-time-point data can be included in model estimation. [34, 35] Least square means, standard errors, and 95% confidence intervals were derived from the model.

We stratified each clinical sample into subgroups based on change in self-reported health status using general health anchors and, additionally, using pain anchors. These change anchors were either based on self-reported magnitude of changes or calculated changes in self-reported general health or overall pain scores. Details of how individuals were classified

as “better”, “same” or “worse” are reported in Table 1. Standardized response means (SRM; ratio of mean change to the standard deviation of that change) were estimated. We judged a SRM of $|\geq 0.30|$ to indicate responsiveness. [36–38] Data management and preparation was carried out using SAS 9.3 for Windows (Copyright 2002–2010 SAS Institute Inc.), and all statistical analyses were carried out using STATA/IC 12.1 (Copyright 1985–2011 Stata Corp LP).

RESULTS

The demographic and clinical profiles of each cohort are detailed in the accompanying introductory article in this issue. [30] Pooling across all studies, most participants were non-Hispanic whites (82%), roughly half of all participants were 60 years old or older (54%). Most participants were female (68%), with the exception of the sample with COPD (42%). The majority of the cohort with stable COPD ($n=79$) was white (72%), male (56%), and had a Medical Research Council (MRC) Breathlessness Rating of 1 or 2 (56%). The majority of the cohort with COPD-related exacerbations ($n=46$) was white (73%), male (61%), 50 or more years of age (91%), and had an MRC Breathlessness Rating of 3 or higher (63%). The majority of the cLBP cohort ($n=218$) was white (84%), female (56%), 50 or more years of age (62%) with largely moderate to severe back pain (74% ≥ 8 on 0–10 numeric pain scale for worst back pain). For this cohort, spinal injections were administered an average of 3.9 days after baseline assessment ($sd = 6.6$). The majority of the MDD cohort ($n=196$) was white (78%), female (74%), 18–49 years of age (52%) and had a Center for Epidemiologic Studies-Depression score of 22 or greater (73%). [39] The majority of the cancer cohort ($n=310$) was white (81%), female (61%), 50 or more years of age (76%), and had an Eastern Cooperative Oncology Group Performance Status Rating (ECOG-PSR) of 0 or 1 (77%). The majority of rheumatoid arthritis sample ($n=521$) was white (88%), female (81%), 50 or more years of age (88%) and had a HAQ Disability Index of 0–1 (57%). From the combined sample of 1370 participants, less than 1% ($n=13$) did not have sufficient data available to score PROMIS-PI and -PB measures.

COPD Clinical Severity

COPD participants defined as exacerbators had mean PROMIS-PI scores of 59.9 compared to a mean of 57.4 for stable participants at baseline. This difference, though in the expected direction, was not statistically significant ($p = 0.23$). The mean PROMIS-PB score for exacerbators was 55.0 compared to 53.7 for those classified as stable. This difference also was non-significant ($p = 0.50$)

Longitudinal Change

Baseline, follow-up, and change scores from the mixed effects models are presented in Table 2. Least squares means by clinical group are presented in Figure 1. The largest changes in PROMIS-PI and -PB scores were observed for the cohort with cLBP. These changes, both in magnitude and direction (lower scores at follow-up) were consistent with expectations. Pre- to post-treatment change scores for both PROMIS-PI ($\Delta = -5.4$, 95% CI: -6.6 to -4.3 , $p < 0.001$) and PROMIS-PB ($\Delta = -3.2$, 95% CI: -4.0 to -2.4 , $p < 0.001$) were statistically significant. PROMIS-PI and -PB score changes were smaller but statistically significant for

participants undergoing treatment for depression ($\beta = -1.5$ [95%CI: -2.8 to -0.2], $p=0.027$ and $\beta = -1.8$ [95%CI: -3.1 to -0.5], $p=0.005$, respectively).

The smallest changes in PROMIS-PI scores were observed for participants with RA; the magnitude of change ($\beta = 0.1$) was not statistically different from zero. Changes in PROMIS-PI scores for the cohort with cancer were also small but were statistically significant ($\beta = -1.1$, 95%CI: -2.1 to -0.2 , $p=0.023$). Both of these studies were observational in nature and PRO assessment was not tied to treatment status. Accordingly, separate estimates are presented in Figure 1 for individuals with RA reporting improved (baseline mean=58.4 [sd=7.6]; followup mean=55.6 [sd=7.4]; $\beta = -2.8$; $p<0.001$) or worsening (baseline mean=53.1 [sd=8.7]; followup mean=55.3 [sd=8.7]; $\beta = 2.2$; $p=0.005$) general health over time and for those with cancer reporting improved (baseline mean=52.8 [sd=8.5]; followup mean=49.2 [sd=9.1]; $\beta = -3.6$; $p<0.001$) or worsening (baseline mean=50.7 [sd=8.7]; followup mean=51.3 [sd=9.7]; $\beta = 0.6$; NS) general health over time.

Clinically Anchored Subgroups

Based on our criterion reference of SRM values $\geq .30$, PROMIS-PI and PROMIS-PB measures were responsive to improvement, but there were substantial differences by clinical anchor—general health vs. pain specific (Tables 3 and 4). With the exception of values for those with Cancer and those with COPD exacerbation, SRM values for improvement were larger when a pain-specific anchor was used. On the whole, PROMIS-PI scores were more responsive to improvement than were PROMIS-PB scores; the only exception was in the study of individuals with depression.

The PROMIS-PI and -PB scores proved less responsiveness to *decrements* in pain scores when estimated based on general health anchors. In fact, when general health anchors were used to classify individuals as “worse”, resulting SRM values often were in the wrong direction (negative value indicating improvement in scores). SRM values for “worse” based on pain-specific anchors were more consistent with expectations. A salient example of this finding can be observed in Table 3. PROMIS-PI scores for those with cLBP defined as “worse” had a SRM of 0.44 when calculated based on the pain-specific anchor. In contrast, when using the general health anchor, the estimate was -0.47, a moderately large SRM value, but in the wrong direction. This reversal of direction in some SRM values anchored to worse general health was observed in other studies reported in this issue; that is, some SRM values for groups defined by general health anchors as “worse”, actually had score improvements. Also of note in the results are the relatively large magnitudes of some SRM values for those classified as being the “same” using the general health anchor (e.g., in cLBP, PROMIS-PI = -0.58 and PROMIS-PB SRM = -0.59).

DISCUSSION

The PROMIS-PI and -PB item banks were developed using state-of-the-art techniques including extensive qualitative evaluations [40] and modern psychometric methods.[12, 13] The psychometric properties of these measures have been previously evaluated in a large cross-sectional, community and a number of other clinical samples. [10, 11, 14] The

findings reported here extend the body of knowledge about how PROMIS-PI and –PB scores function, particularly with respect to changes over time.

Some of our *a priori* hypotheses were upheld and some were not. We expected baseline PROMIS-PI and –PB scores to be higher, on average in persons with COPD exacerbation compared to those defined as stable, but the observed differences were non-significant. This may have been due, in part, to low statistical power; this cohort was the smallest in the study. However, if the differences had been statistically significant, this would not have altered the fact that the differences in PROMIS-PI and –PB scores were small in magnitude—differences of 2.2 and 1.2, respectively. In our *a priori* search for clinical comparisons by which to evaluate the PROMIS-PI and –PB scores, we found it intuitively attractive to compare by COPD severity. However, though pain is prevalent in COPD (estimates range 32% to 60%), [41] it is not used for diagnosing the condition, classifying its severity, or defining an exacerbation. It is possible to have an exacerbation without increase in pain.

A priori hypotheses regarding changes in intervention groups' scores were upheld. We expected and found statistically significant changes in PROMIS-PI and –PB scores for those receiving intervention for cLBP and for MDD. These changes were substantially larger for the cLBP group than for the MDD group. SRM values for clinically anchored groups supported PROMIS-PI and -PB scores responsiveness to change. Of note, however, were substantial differences in SRM values for “worse” when the sample was classified based on a general health anchor. Not only were SRMs values generally lower when based on the general health anchor, in some cases they were counter-intuitive, which calls into question the appropriateness of using a general health anchor. It also underscores the impact choice of anchor has on psychometric assessments. Further, though anchoring change estimates on patients' perceptions of change is appealing intuitively, it has limits. Retrospective global ratings of change have been criticized because of their vulnerability to response bias, [42] and these ratings have been found to be more strongly associated with current status than with change in status. [43, 44]

This study had additional limitations. As stated in the introductory paper, [30] the decision to conduct the analyses reported in this issue was opportunistic. Studies were conducted around the same time using many of the same measures providing the opportunity to evaluate PROMIS measures by domain across several clinical populations. However, there were methodological differences across studies (e.g., differences in wording of anchors), and this was a limitation. Another limitation is more subtle. Most of the authors of this study were involved in the original work in developing the PROMIS-PI and –PB item banks. We recognize that this “pride of ownership” could influence how results are presented and interpreted.

Despite these limitations, the results of this study add to a growing body of evidence supporting the usefulness of PROMIS-PI and –PB scores in pain studies. Scores were responsive to interventions, especially to an intervention that specifically targeted pain. Psychometric validation, however, is never a completed task, and evaluations of the PROMIS-PI and –PB scores are warranted in other clinical samples and contexts. Of particular interest would be intervention trials in which responsiveness of PROMIS-PI and –

PB scores are compared to that of other pain measures. Also useful would be studies that relate scores to other benchmarks of interest to pain clinicians and researchers.

Acknowledgments

PROMIS® was funded with cooperative agreements from the National Institutes of Health (NIH) Common Fund Initiative (Northwestern University, PI: David Cella, PhD, U54AR057951, U01AR052177, R01CA60068; Northwestern University, PI: Richard C. Gershon, PhD, U54AR057943; American Institutes for Research, PI: Susan (San) D. Keller, PhD, U54AR057926; State University of New York, Stony Brook, PIs: Joan E. Broderick, PhD and Arthur A. Stone, PhD, U01AR057948, U01AR052170; University of Washington, Seattle, PIs: Heidi M. Crane, MD, MPH, Paul K. Crane, MD, MPH, and Donald L. Patrick, PhD, U01AR057954; University of Washington, Seattle, PI: Dagmar Amtmann, PhD, U01AR052171; University of North Carolina, Chapel Hill, PI: Harry A. Guess, MD, PhD (deceased), Darren A. DeWalt, MD, MPH, U01AR052181; Children's Hospital of Philadelphia, PI: Christopher B. Forrest, MD, PhD, U01AR057956; Stanford University, PI: James F. Fries, MD, U01AR052158; Boston University, PIs: Alan Jette, PT, PhD, Stephen M. Haley, PhD (deceased), and David Scott Tulskey, PhD (University of Michigan, Ann Arbor), U01AR057929; University of California, Los Angeles, PIs: Dinesh Khanna, MD (University of Michigan, Ann Arbor) and Brennan Spiegel, MD, MSHS, U01AR057936; University of Pittsburgh, PI: Paul A. Pilkonis, PhD, U01AR052155; Georgetown University, PIs: Carol M. Moinpour, PhD (Fred Hutchinson Cancer Research Center, Seattle) and Arnold L. Potosky, PhD, U01AR057971; Children's Hospital Medical Center, Cincinnati, PI: Esi M. Morgan DeWitt, MD, MSCE, U01AR057940; University of Maryland, Baltimore, PI: Lisa M. Shulman, MD, U01AR057967; and Duke University, PI: Kevin P. Weinfurt, PhD, U01AR052186). NIH Science Officers on this project have included Deborah Ader, PhD, Vanessa Ameen, MD (deceased), Susan Czajkowski, PhD, Basil Eldadah, MD, PhD, Lawrence Fine, MD, DrPH, Lawrence Fox, MD, PhD, Lynne Haverkos, MD, MPH, Thomas Hilton, PhD, Laura Lee Johnson, PhD, Michael Kozak, PhD, Peter Lyster, PhD, Donald Mattison, MD, Claudia Moy, PhD, Louis Quatrano, PhD, Bryce Reeve, PhD, William Riley, PhD, Peter Scheidt, MD, Ashley Wilder Smith, PhD, MPH, Susana Serrate-Sztejn, MD, William Phillip Tonkins, DrPH, Ellen Werner, PhD, Tisha Wiley, PhD, and James Witter, MD, PhD. The contents of this article uses data developed under PROMIS. These contents do not necessarily represent an endorsement by the US Federal Government or PROMIS. See www.nihpromis.org for additional information on the PROMIS® initiative.

References

1. Committee on Advancing Pain Research Care and Education, Institute of Medicine. *Relieving pain in america: A blueprint for transforming prevention, care, education, and research*. Washington: National Academies Press; 2011.
2. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007 Jan; 60(1):34–42. [PubMed: 17161752]
3. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: Immpact recommendations. *Pain*. 2005; 113(1):9–19. [PubMed: 15621359]
4. Riley WT, Rothrock N, Bruce B, Christodolou C, Cook K, Hahn EA, et al. Patient-reported outcomes measurement information system (PROMIS) domain names and definitions revisions: Further evaluation of content validity in IRT-derived item banks. *Qual Life Res*. 2010; 19(9):1311–21. [PubMed: 20593306]
5. PROMIS Health Organization and PROMIS Cooperative Group. PROMIS adult self-reported health-domain frameworks. 2011. [cited 2014 August 20]; Available from: <http://www.nihpromis.org/measures/domainframework1>
6. Fordyce, WE. *Behavioral methods for chronic pain and illness*. St. Louis: Mosby; 1976.
7. Prkachin KM, Schultz IZ, Hughes E. Pain behavior and the development of pain-related disability: The importance of guarding. *Clin J Pain*. 2007; 23(3):270–7. [PubMed: 17314588]
8. Hadjistavropoulos T, Herr K, Turk DC, Fine PG, Dworkin RH, Helme R, et al. An interdisciplinary expert consensus statement on assessment of pain in older persons. *Clin J Pain*. 2007; 23:S1–S43. [PubMed: 17179836]
9. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The patient-reported outcomes measurement information system (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. *Med Care*. 2007; 45(5 Suppl 1):S3–S11. [PubMed: 17443116]

10. Amtmann D, Cook KF, Jensen MP, Chen W-H, Choi S, Revicki D, et al. Development of a PROMIS item bank to measure pain interference. *Pain*. 2010; 150(1):173–82. [PubMed: 20554116]
11. Revicki DA, Chen WH, Harnam N, Cook K, Amtmann D, Callahan L, et al. Development and psychometric analysis of the PROMIS pain behavior item bank. *Pain*. 2009; 146(1–2):158–69. [PubMed: 19683873]
12. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, 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):1179–94. [PubMed: 20685078]
13. Embretson, SE.; Reise, SP. *Item response theory for psychologists*. Mahwah, N.J: Lawrence Erlbaum Associates; 2000.
14. Wagner LI, Schink J, Bass M, Patel S, Diaz MV, Rothrock N, et al. Bringing PROMIS to practice: Brief and precise symptom screening in ambulatory cancer care. *Cancer*. 2014 Epub ahead of print.
15. Salem R, Bamer AM, Alschuler KN, Johnson KL, Amtmann D. Obesity and symptoms and quality of life indicators of individuals with disabilities. *Disabil Health J*. 2014; 7(1):124–30. [PubMed: 24411517]
16. Kappelman MD, Long MD, Martin C, DeWalt DA, Kinneer PM, Chen W, et al. Evaluation of the patient-reported outcomes measurement information system in a large cohort of patients with inflammatory bowel diseases. *Clin Gastroenterol Hepatol*. 2014 Aug; 12(8):1315–23.e2. [PubMed: 24183956]
17. Broderick JE, Schneider S, Junghaenel DU, Schwartz JE, Stone AA. Validity and reliability of patient-reported outcomes measurement information system instruments in osteoarthritis. *Arthritis Care Res*. 2013; 65(10):1625–33.
18. Schneider S, Choi SW, Junghaenel DU, Schwartz JE, Stone AA. Psychometric characteristics of daily diaries for the patient-reported outcomes measurement information system (PROMIS®): A preliminary investigation. *Qual Life Res*. 2013; 22(7):1859–69. [PubMed: 23180166]
19. Simon JE, Docherty CL. Current health-related quality of life is lower in former division I collegiate athletes than in non-collegiate athletes. *Am J Sports Med*. 2013; 42(2):423–9. [PubMed: 24318608]
20. Burke JR. Feasibility of using the patient-reported outcomes measurement information system in academic health centers: Case series design on pain reduction after chiropractic care. *J Chiropr Med*. 2014; 13(3):168–77. [PubMed: 25225465]
21. Shahgholi L, Yost KJ, Kallmes DF. Correlation of the national institutes of health patient reported outcomes measurement information system scales and standard pain and functional outcomes in spine augmentation. *Am J Neuroradiol*. 2012 Dec 1; 33(11):2186–90. [PubMed: 22700753]
22. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992; 30(6):473–83. [PubMed: 1593914]
23. Cleeland C. Measurement of pain by subjective report. *Adv Pain Res Ther*. 1989; 12:391–403.
24. Krebs EE, Lorenz KA, Bair MJ, Damush TM, Wu J, Sutherland JM, 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):733–8. [PubMed: 19418100]
25. Attal N, Cruccu G, Baron R, Haanpää M, Hansson P, Jensen T, et al. Efn guidelines on the pharmacological treatment of neuropathic pain: 2010 revision. *Eur J Neurol*. 2010; 17(9):1113–e88. [PubMed: 20402746]
26. Samejima, F. Estimation of latent ability using a response pattern of graded scores. Richmond, VA: Psychometric Society; 1969. Available from: <http://www.psychometrika.org/journal/online/MN17.pdf>
27. Cook KF, Bamer AM, Amtmann D, Molton IR, Jensen MP. Six patient-reported outcome measurement information system short form measures have negligible age- or diagnosis-related differential item functioning in individuals with disabilities. *Arch Phys Med Rehabil*. 2012 Epub ahead of print.

28. Kim J, Chung H, Amtmann D, Revicki DA, Cook KF. Measurement invariance of the PROMIS pain interference item bank across community and clinical samples. *Qual Life Res.* 2013; 22(3): 501–7. [PubMed: 22552604]
29. Chung H, Kim J, Cook K, Askew R, Revicki D, Amtmann D. Testing measurement invariance of the patient-reported outcomes measurement information system pain behaviors score between the us general population sample and a sample of individuals with chronic pain. *Qual Life Res.* 2014; 23(1):239–44. [PubMed: 23824606]
30. Cook KF, Jensen SE, Schalet BD, Beaumont JL, DA, Czaikowski S. PROMIS measures of pain, fatigue, negative affect, physical function and social function demonstrate clinical validity across a range of chronic conditions. *J Clin Epidemiol.* Submitted.
31. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Nhlbi/who global initiative for chronic obstructive lung disease (gold) workshop summary. *Am J Respir Crit Care Med.* 2001; 163(5):1256–76.
32. Verbeke, G.; Molenberghs, G. Linear mixed models for longitudinal data. New York, NY: Springer-Verlag; 2000.
33. Hedeker, DR.; Gibbons, RD. Longitudinal data analysis. Hoboken, N.J: Wiley-Interscience; 2006.
34. Little, RJA.; Rubin, DB. Statistical analysis with missing data. Hoboken, NJ: John Wiley & Sons, Inc; 2002.
35. Troxel AB, Fairclough DL, Curran D, Hahn EA. Statistical analysis of quality of life with missing data in cancer clinical trials. *Stat Med.* 1998; 17(5–7):653–66. [PubMed: 9549814]
36. Yost KJ, Cella D, Chawla A, Holmgren E, Eton DT, Ayanian JZ, et al. Minimally important differences were estimated for the functional assessment of cancer therapy-colorectal (fact-c) instrument using a combination of distribution- and anchor-based approaches. *J Clin Epidemiol.* 2005; 58(12):1241–51. [PubMed: 16291468]
37. Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six patient-reported outcomes measurement information system-cancer scales in advanced-stage cancer patients. *J Clin Epidemiol.* 2011; 64(5):507–16. [PubMed: 21447427]
38. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes.* 2006; 4:1–5. [PubMed: 16393335]
39. Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. *Appl Psychol Meas.* 1977; 1(3):385–401.
40. DeWalt DA, Rothrock N, Yount S, Stone AA. PROMIS Cooperative Group. Evaluation of item candidates: The PROMIS qualitative item review. *Med Care.* 2007; 45(5 Suppl 1):S12–S21. [PubMed: 17443114]
41. van Isselt, EFvD; Groenewegen-Sipkema, KH.; Spruit-van Eijk, M.; Chavannes, NH.; de Waal, MW.; Janssen, DJ., et al. Pain in patients with COPD: A systematic review and meta-analysis. *BMJ open.* 2014; 4(9):e005898.
42. Ross M. Relation of implicit theories to the construction of personal histories. *Psychol Rev.* 1989; 96(2):341.
43. Schmitt J, Di Fabio RP. The validity of prospective and retrospective global change criterion measures. *Arch Phys Med Rehabil.* 2005; 86(12):2270–6. [PubMed: 16344022]
44. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: The lesson of cronbach. *J Clin Epidemiol.* 1997; 50(8):869–79. [PubMed: 9291871]

What is new?

Key Findings

- PROMIS[®] Pain Interference (PROMIS PI) and Pain Behavior (PROMIS-PB) scores were responsive to treatment and self-reported clinical change.

What this adds to what is known?

- These findings add to the body of evidence regarding the clinical validity of PROMIS pain measures.

What is the implication and what should change now?

- When selecting pain measures for clinical trials and comparative effectiveness research, the current findings could be considered along with the growing body of validity evidence for PROMIS PI and PROMIS-PB scores.

Change in PROMIS Pain Interference and Pain Behavior T-scores Across Clinical Samples

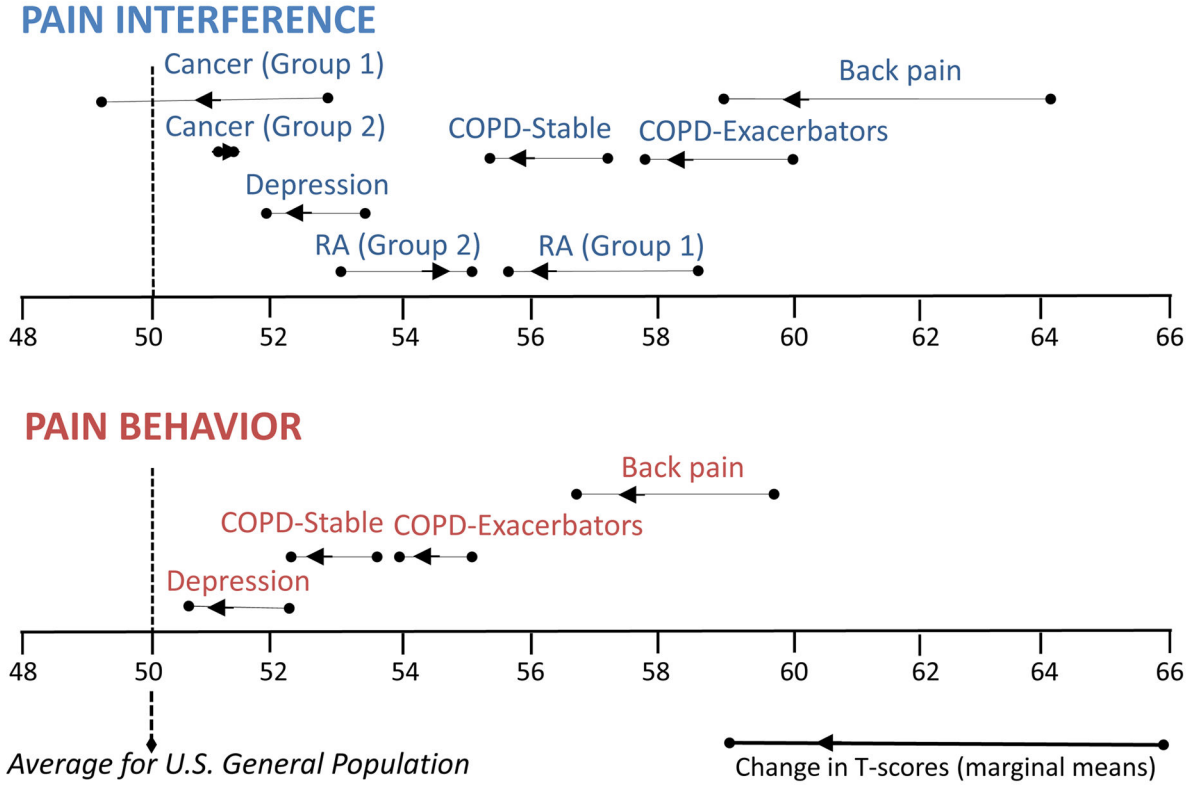


Figure 1. Change in PROMIS Pain Interference and Pain Behavior scores by clinical sample
 *Cancer (group 1, n=51): patients reporting improved global health at follow-up
 *Cancer (group 2, n=85): patients reporting worsened global health at follow-up
 *RA (group 1, n=60): patients reporting improved global health at follow-up
 *RA (group 2, n=92): patients reporting worsened global health at follow-up

Table 1

Anchor items, responses, and change classifications by clinical sample

Disease or Condition	Pain Anchor Text	Responses	Classification	General Health Anchor Text	Responses	Classification
Chronic Low Back Pain	"Compared to your first appointment how would you rate your current back pain?"	Much better	Better	"Is your health [(Excellent, Very good, Good, Fair, Poor (5 to 1))]"	Baseline minus follow-up < 0	Better
		Somewhat better	About the same		Baseline minus follow-up = 0	About the same
		No change	Worse		Baseline minus follow-up > 0	Worse
		Somewhat worse				
Cancer	"Since the last time you filled out a questionnaire your level of pain is..."	Much worse	Better	"In general, would you say your health is [(Excellent, Very good, Good, Fair, Poor (5 to 1))]"	Baseline minus follow-up < 0	Better
		Very much better	About the same		Baseline minus follow-up = 0	About the same
		Moderately better	Worse		Baseline minus follow-up > 0	Worse
		A little better				
COPD	"How would you rate your pain on average?" (0-10)	About the same	Better	"Has there been any change in your overall health since you started the study?"	Very much better	Better
		A little worse	About the same		Moderately better	
		Moderately worse	Worse		A little better	
		Very much worse			About the same	About the same
Depression	"When you had back pain, how would you rate your average pain?" (0-10)	Got a lot better	Better	"Is your health [(Excellent, Very good, Good, Fair, Poor (5 to 1))]"	Baseline minus follow-up < 0	Better
		Got a little better	About the same		Baseline minus follow-up = 0	About the same
			Worse		Baseline minus follow-up > 0	Worse
Rheumatoid Arthritis	"How has your pain changed?"	Got a lot better	Better	"In general, would you say your 'current' health is [(Excellent, Very good, Good, Fair, Poor (5 to 1))]"	Baseline minus follow-up < 0	Better
		Got a little better			Baseline minus follow-up = 0	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Disease or Condition	Pain Anchor Text	Responses	Classification	General Health Anchor Text	Responses	Classification
		Stayed the same	About the same		Baseline minus follow-up > 0	About the same
		Got a little worse	Worse		Baseline minus follow-up > 0	Worse
		Got a lot worse				

Table 2
Baseline, follow-up, and change in PROMIS Pain Interference and Pain Behavior estimates by clinical sample

	Back Pain		Cancer		COPD-Exacerbation		COPD-Stable		Depression		Rheumatoid Arthritis	
	n=218/175	95% CI	n=305/277	95% CI	n=44/44	95% CI	n=74/70	95% CI	n=196/187	95% CI	n=520/472	95% CI
PAIN INTERFERENCE												
Baseline	64.2	(63.2, 65.2)	51.9	(50.8, 52.9)	59.9	(56.6, 63.1)	57.4	(55.1, 59.7)	53.4	(51.9, 54.9)	55.3	(54.6, 56.0)
Follow-up	58.8	(57.6, 59.9)	50.7	(49.6, 51.8)	57.7	(54.4, 61.0)	55.2	(52.9, 57.5)	51.9	(50.4, 53.4)	55.4	(54.7, 56.2)
Change	-5.4**	(-6.6, -4.3)	-1.1*	(-2.1, -0.2)	-2.2	(-5.6, 1.3)	-2.2*	(-4.2, -0.1)	-1.5*	(-2.8, -0.2)	0.12	(-0.5, 0.7)
PAIN BEHAVIOR												
Baseline	59.7	(59.1, 60.4)			55.0	(52.0, 58.1)	53.7	(51.5, 55.9)	52.2	(50.7, 53.7)		
Follow-up	56.6	(55.8, 57.3)			53.8	(50.8, 56.8)	52.3	(50.1, 54.5)	50.4	(48.9, 51.9)		
Change	-3.2**	(-4.0, -2.4)			-1.2	(-4.1, 1.7)	-1.4	(-3.4, 0.5)	-1.8*	(-3.1, -0.5)		

COPD, Chronic Obstructive Pulmonary Disease; CI, Confidence Interval;

* denotes statistical significance at $p < 0.05$;

** denotes statistical significance at $p < 0.001$

Sample sizes are listed as baseline/follow-up, and differences between sample sizes from this paper and the related papers in this issue reflect individual differences in item-level response rates.

Table 3
Standardized response means for PROMIS Pain Interference scores by clinical subgroup and change anchor

Clinical Subgroup	Change by general health anchor			Change by pain anchor				
	n	SRM score	SD ()	n	SRM score	SD ()		
Back Pain								
Better	51	-0.94	-7.50	7.96	112	-1.09	-8.09	7.43
Same	94	-0.58	-4.62	7.97	32	-0.26	-1.64	6.33
Worse	24	-0.47	-3.36	7.18	27	0.44	2.18	4.95
Cancer								
Better	51	-0.52	-3.60	6.98	79	-0.43	-3.16	7.37
Same	130	-0.19	-1.43	7.74	109	-0.34	-2.78	8.08
Worse	85	0.06	0.56	9.95	53	0.57	4.44	7.82
COPD-Exacerbators								
Better	7	-0.95	-11.87	12.51	15	-0.87	-8.66	9.98
Same	14	0.49	4.34	8.90	12	-0.37	-4.11	11.19
Worse	13	-0.09	-0.86	9.75	14	0.63	6.09	9.63
COPD-Stable								
Better	13	-1.02	-8.20	8.02	21	-1.14	-8.41	7.37
Same	38	-0.15	-1.02	6.99	22	-0.50	-3.23	6.52
Worse	14	0.08	0.90	11.73	25	0.51	3.94	7.69
Depression								
Better	43	-0.40	-4.29	10.66	66	-0.73	-6.44	8.86
Same	113	-0.15	-1.14	7.76	73	-0.21	-1.70	8.11
Worse	30	0.06	0.77	12.06	46	0.67	5.29	7.90
Rheumatoid Arthritis								
Better	60	-0.49	-2.75	5.67	75	-0.52	-3.31	6.42
Same	297	0.01	0.07	6.13	215	0.00	0.02	6.61
Worse	92	0.30	2.22	7.46	159	0.34	1.97	5.75

SRM, Standardized Response Mean; COPD, Chronic Obstructive Pulmonary Disease; SD, standard deviation

Standardized response means for PROMIS Pain Behavior scores by clinical subgroup and change anchor

Table 4

Clinical Subgroup	Change by general health anchor			Change by pain anchor				
	n	SRM	SD ()	n	SRM	SD ()		
Back Pain								
Better	51	-0.69	4.89	7.11	112	-0.81	-4.88	5.98
Same	95	-0.59	2.74	4.62	32	-0.60	-1.53	2.56
Worse	24	-0.52	2.57	4.93	28	0.31	1.01	3.28
COPD-Exacerbators								
Better	7	-0.85	8.30	9.80	15	-0.85	-7.32	8.63
Same	14	0.39	3.75	9.54	12	-0.37	-2.48	6.71
Worse	13	-0.06	0.55	8.69	14	0.73	6.52	8.92
COPD-Stable								
Better	13	-0.70	6.09	8.67	21	-0.81	-6.96	8.64
Same	38	-0.02	0.11	7.07	22	-0.32	-1.76	5.44
Worse	14	-0.08	0.90	10.82	25	0.47	3.48	7.37
Depression								
Better	43	-0.48	4.33	8.99	66	-0.65	-5.73	8.75
Same	113	-0.18	1.52	8.41	73	-0.17	-1.49	8.85
Worse	30	0.03	0.37	10.85	46	0.33	2.65	7.93

SRM, Standardized Response Mean; COPD, Chronic Obstructive Pulmonary Disease; SD, standard deviation