Development and Psychometric Analysis of the PROMIS Pain Behavior Item Bank

Dennis A. Revicki, PhD[1], **Wen-Hung Chen, PhD**[1], **Neesha Harnam, MPH**[1], **Karon F. Cook, PhD**[2], **Dagmar Amtmann, PhD**[2], **Leigh F. Callahan, PhD**[3], **Mark P. Jensen, PhD**[2], and **Francis J. Keefe, PhD**[4]

[1]Center for Health Outcomes research, United BioSource Corporation, Bethesda, MD

[2]Department of Rehabilitation Medicine, University of Washington, Seattle, WA

[3]Department of Medicine, University of North Carolina, Chapel Hill, NC

[4]Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC

Abstract

The measurement of pain behavior is a key component of the assessment of persons with chronic pain; however few self-reported pain behavior instruments have been developed. We developed a pain behavior item bank as part of the Patient Reported Outcome Measurement Information System (PROMIS). For the Wave I testing, because of the large number of PROMIS items, a complex sampling approach was used where participants were randomly assigned to either respond to two full item banks or to multiple 7-item blocks of items. A web-based survey was designed and completed by 15,528 members of the general population and 967 individuals with different types of chronic pain. Item response theory (IRT) analysis models were used to evaluate item characteristics and to scale both items and individuals on the pain behavior domain. The pain behavior item bank demonstrated good fit to a unidimensional model (Comparative Fit Index = 0.94). Several iterations of IRT analyses resulted in a final 39 item pain behavior bank, and different IRT models were fit to the total sample and to those participants who experienced some pain. The results indicated that these items demonstrated good coverage of the pain behavior construct. Pain behavior scores were strongly related to pain intensity and moderately related to self-reported general health status. Mean pain behavior scores varied significantly by groups based on pain severity and general health status. The PROMIS pain behavior item bank can be used to develop static short-form and dynamic measures of pain behavior for clinical studies.

Keywords

Pain behavior; item response theory analysis; patient reported outcomes; psychometric analysis; chronic pain; item banks

CORRESPONDING AUTHOR: Dennis Revicki, PhD, Center for Health Outcomes Research, United BioSource Corporation, 7101 Wisconsin Ave., Suite 600, Bethesda, MD 20814, Voice: 1-301-654-9729, Fax: 1-301-654-9864, Dennis.Revicki@unitedbiosource.com .

INTRODUCTION

People with pain engage in behaviors that communicate to others that they are experiencing pain [6,13,31]. Pain behaviors, may include verbal complaints of pain and suffering, non-language sounds, facial expressions, body posturing and gesturing, and limitations in activities. There is growing recognition that assessment of pain behavior is a key outcome of persons suffering from chronic pain [7,26]. Careful measurements of pain behavior can be helpful in several ways. First, it can provide clues about the existence, intensity and causes of pain. Second, knowledge about pain behaviors can provide insights into a person's attempts to cope with or manage pain [33]. This information may help clinicians identify and reinforce adaptive pain coping efforts. Information on pain behaviors can also reveal maladaptive coping efforts that can be targeted in treatment interventions. Finally, pain behaviors may provide a behavioral marker of enhanced risk for the development of chronic pain and disability [11].

Direct observation provides the most objective approach to assessing pain behavior [13]. Two basic methods have been developed, sampling pain behavior during a standardized task [12] or in a naturalistic setting [23]. Although direct observation methods are objective, they have limitations (e.g., need for ongoing observer training, costs, intrusiveness) that limit their utility in clinical settings.

Self-report provides an alternative strategy for measuring pain behavior [14]. Individuals with pain are often aware of pain's effect on their behavior and can report on their experience and activities. Given the advantages of self-report (e.g., low cost, ease of administration), it is surprising that self-report pain behavior measures are not more widely used in pain assessment. Among the self-report measures of pain behavior that have been developed and validated, the Pain Behavior Check List (PBCL) is the most widely used in research studies [14,19]. The PBCL assesses four categories of pain behaviors (distorted ambulation, affective distress, facial/audible expressions, and help seeking behavior). Although re is evidence for the reliability and validity of PBCL scores [14], the scale is brief and does not capture potentially important categories of pain behaviors such as isolating oneself from others, crying, and massaging a painful area, among others.

A new effort to develop a comprehensive pain behavior item bank was undertaken through the NIH Patient-Reported Outcomes Measurement Information System (PROMIS) [3,21]. The goal of the current study was to develop a self-report measure that can be integrated into clinical studies and that reliably assesses a full array of pain behaviors that can be integrated into clinical studies. This report focuses on the development and psychometric evaluation of the PROMIS pain behavior item bank. The psychometric characteristics of the final pain behavior item bank were evaluated based both on classical test theory and modern measurement theory methods [8,21,22,28]. Modern measurement theory approaches include application of item response theory (IRT) models to evaluate item characteristics and to scale both items and individuals on a unidimensional construct (i.e., pain behavior). IRT methods allow for more comprehensive understanding of item responses and their relationships to the targeted domain. Items with good psychometric properties can be assembled into short-form scales or computerized adaptive tests for more targeted patient assessment.

METHODS

Study Design

A web-based survey was developed and conducted to field test 14 candidate PROMIS item banks, including pain behavior. This field test was completed during 2006 using an internet panel maintained by YouGov/Polimetrix (see www.polimetrix.com). The 14 item banks represented five domains (i.e., pain, fatigue, physical functioning, social activities, and

emotional distress) were administered to selected participants [3]. Because of the large number of items to be tested, a complex sampling approach was used. Participants were randomly assigned either to respond to two full item banks or to multiple 7-item blocks of items. Item banks were comprised of the full set of candidate items for a particular domain; for example, the pain impact bank had 56 items and the pain behavior bank had 52 items. Those administered item blocks responded to 14 7-item blocks, one for each of the 14 PROMIS sub-domains (e.g., fatigue experience and impact, physical function, pain behavior, impact and quality, social function, depression, anxiety, anger).

Study Participants

The PROMIS field test sample was developed to be generally comparable to distributions of gender, age, race/ethnicity (White/Black/Hispanic/Other) and education (High School or less versus more than High School) based on the 2000 US census data. These participants were identified from the YouGov/Polimetrix internet panel. A final scale-setting, community sample was identified and weighted to reflect the 2000 United States General Census. For the current study, the study subjects included those who completed either all 52 items of the pain behavior item bank or a subset of 7 items from this bank.

Wave I Sample—Because of the number of item banks tested in Wave I, a complex data collection strategy was employed (see Figure 1). This strategy included two arms and a total sample size of 21,133. A total of 19,601 were recruited by Polimetrix, with the remaining 1,532 recruited from selected PROMIS primary research sites. In the full bank testing arm, 7,005 persons from the general population were administered 2 of the 14, sub-domain-specific PROMIS item banks. In the second arm, 14,128 individuals were administered randomly-selected 7-item blocks measuring each of 14 PROMIS-targeted sub-domains (Figure 1). The primary research sites and the Polimetrix sample included both community and clinical samples. The clinical samples included persons with heart disease (n = 1,156), cancer (n = 1,754), rheumatoid arthritis (n = 557), osteoarthritis (n = 918), psychiatric illness (n = 1,193), COPD (n = 1,214), spinal cord injury (n = 531), and other conditions (n = 560).

Secondary Data Collection

After initial calibrations of WAVE I data, it was noted that, because most of the participants were from the general community, there were relatively few persons who had moderate to severe pain. We recruited an additional sample to increase the number of participants with more severe pain intensity, including members of American Chronic Pain Association (ACPA), a non-profit organization focused on education and support for people with chronic pain.

Nine hundred and sixty-seven participants with chronic pain were recruited through the ACPA. An invitation to complete the PROMIS pain survey was posted on the ACPA website. To be eligible, participants had to be 21 years of age or older and have at least one chronic pain condition for at least 3 months prior to participating in the survey. Those who met eligibility criteria were asked to complete an informed consent form. After obtaining informed consent, participants immediately began the survey. The survey was posted on the website of the ACPA from September 2007 to March 2008.

Measures

Pain Behavior Item Bank—A draft pain behavior item bank was developed based on the published literature, clinician review, and qualitative research with patients experiencing various kinds of pain. Existing observer-rated and self-reported pain behavior instruments were reviewed [5,10,12,14,15,20,23,24,27,29,30,32], and we summarized the range and type of item content covered by the existing published measures. A comprehensive set of draft pain behavior items were generated, and then these items were reviewed by six pain and outcomes research

experts. After several iterations of item review and revision, a final draft set of pain behavior items were available for field testing. Content of the final set of pain behavior items was revised based on the results of cognitive debriefing interviews [4]. Cognitive debriefing interviews are designed to determine whether respondents understand the instructions, content of items, recall period, and response scales of a patient-reported instrument [34]. Based on the cognitive debriefing interviews, several of the items were modified to increase respondent understanding. The final pain behavior item bank consisted of 52 items covering movement (i.e., move slowly, stiffness), affective (i.e., irritable, angry), social interaction (i.e., ask for help, withdraw), and facial/verbal (i.e., groan, grimace) behaviors. As can be seen in Table 1, the items cover a very wide range of pain behaviors. Respondents rated how frequently they engaged in each pain behavior using a 6-point Likert-type response scale, ranging from 1 (had no pain) to 6 (always). The recall period for the pain behavior items was the past 7 days.

Other Measures—Demographic information was collected for the study participants (i.e., age, gender, race/ethnicity, education). In addition, all participants completed the PROMIS average pain intensity item based on a 0 to 10 numeric rating scale and verbal pain rating scale ("none" to "very severe"). Information was also collected on a number of chronic medical conditions in the Wave I sample and on different chronic pain conditions in the ACPA sample.

All participants in the Wave I sample completed nine additional global rating items covering physical health, mental health, physical function, fatigue, emotional distress, social satisfaction, social activities, general health perceptions, and quality of life [9]. The recall period for the global items was the past 7 days.

Statistical Analysis

Descriptive statistics for the demographic characteristics and average pain intensity were summarized for the entire study sample. We report means and standard deviations for the pain behavior bank items. The psychometric analyses were guided by the PROMIS analysis plan [21]. Item response theory (IRT) analyses were applied to evaluate the measurement qualities of the pain behavior items [8,21,22,28]. IRT makes full use of data provided by participants' responses to each item of an instrument or item bank. Simultaneous estimates of the properties of items and persons' levels of the trait being measured are used to develop a probabilistic model that places individuals' scores and item properties on the same metric [22]. IRT models estimate a set of properties for each item in the scale (i.e., item calibrations). The item properties represent an item's relationship with a measured construct (i.e., pain impact, pain behavior) [8,21,22]. The item parameters from an IRT analysis allows instrument developers to select the most effective items to target a patient's level of functioning.

We examined item-total correlations and Cronbach's alpha for the pain behavior items. Exploratory and confirmatory factor analyses (CFA) were conducted to assess the unidimensionality of the pain behavior items. For the exploratory analysis, we did not specify any number of factors. A one-factor solution was specified and goodness of fit tests were evaluated in the confirmatory analysis. The fit statistics examined included: Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). MPLUS was used to perform the CFA [16].

The graded response IRT model was fit to the 52 items using MULTILOG [25] to estimate item parameters. The graded response model is a flexible IRT model that supports examination of item and scale properties, estimation of item characteristics, and estimating persons' pain behavior scores based on responses to items of the pain behavior item bank. The item characteristics that are estimated with the graded response model are slope and threshold parameters. The slope parameter indicates the strength of the relationship between an item and the measured construct (i.e., pain behavior). The threshold parameters provide information on

item difficulty or severity, and locate the item along the measured construct. IRTFIT [1] was used to assess IRT model fit for each item. IRTFIT computes the extension of $S-X^2$ and $S-G^2$ [17,18] for items with more than two responses. These statistics estimate the fit of the item responses to the IRT model, that is, whether the responses follow the pattern predicted by the model. Statistically significant differences indicate poor fit. The $S-X^2$ (a Pearson X^2 statistic) and $S-G^2$ (a likelihood ratio G^2 statistic) are fit statistics that use the sum score of all items and compare the predicted and observed response frequencies for each level of the scale sum score.

The development of an item bank starts with more items than will be used in the final bank. The item selection process for the pain behavior bank was iterative. First, the IRT calibration was conducted to estimate item parameters, and IRT model fit was estimated. Items were excluded from the next iteration if the p-value of either $S-X^2$ or $S-G^2$ was less than 0.001. The process continued until no items exhibited significant misfit (alpha = 0.001).

At the conclusion of the iterative IRT and fit analyses, 39 items were selected for the PROMIS pain behavior bank. However, there was a concern regarding the slope parameters of these items. The slope parameters were higher than generally found in other IRT studies. We hypothesized that the high slope parameters were an artifact caused by a large proportion of subjects who reported experiencing "no pain". This no-pain population created either non-normality or mixed sample (general healthy and those with some pain) that affected the IRT parameter estimation. The high slopes were thought to result from the large variance in responses and the mixture of two distinct populations: people with pain and people without pain.

Several options were proposed to address the high slope findings. One was to exclude subjects who answered 0 or 1 on an 11-point pain intensity scale (0='No pain', 10='worst pain imaginable'). The other solution was to fit the pain behavior items with a constrained nominal model [2]. For a nominal model, each response option has its own slope parameter. In our constrained nominal model, two slopes were estimated, one slope for the "had no pain" option, and one slope for all other response options. We decided that excluding subjects reported no pain was the most feasible approach and to have item parameters consistent with other PROMIS domains. Therefore, IRT analysis was again conduced for the 39 selected items with the "no-pain" subjects excluded. This resulted in lower slope parameter estimates.

For IRT the concept of reliability is conceptualized as "information" and examines measurement precision can differ across the levels of a construct. The relationship between information and standard error (SE) is defined by the formula, $SE(\theta)=1/\sqrt{I(\theta)}$, where θ is estimated trait level, SE is the standard error of θ, and I is information. As the formula indicates, increased scale information is associated with smaller SE's and, therefore, greater precision.

Differential item functioning (DIF) was examined by gender, age and education. DIF exists when an item functions differently for respondents from different subgroups [21,22]. DIF can be consistent across the range of the trait being measured (uniform DIF), or its impact can vary depending on trait level (non-uniform DIF). Logistic regression as described in Zumbo [33] was used to for detecting DIF. The estimated IRT theta was used as the conditioning variable. DIF was assessed for gender, education (high school or less vs. high school and above), and age (< 65 vs. >=65).

The concurrent validity of the pain behavior bank items was further tested by correlating the pain behavior scores with measures of pain intensity and the global PROMIS items. We evaluated known groups validity by comparing mean pain behavior scores by groups defined by pain intensity (none, mild, moderate, severe, very severe) and general health status (poor, fair, good, very good, excellent). Analysis of variance was used to compare the mean pain

behavior scores by pain intensity and general health status group. Scheffe's adjustment was used to control for multiple comparisons.

RESULTS

Demographic Characteristics

The overall sample (n = 21,133) was 52% female. The sample median age was approximately 50 years; 12% were 18–29 years; 12% were 30–39 years; 16% were 40–49 years, 32% were 50–64 years, and 28% were 65 years or older. The racial/ethnic breakdown was 82% white; 9% black; 8% multi-racial; and 1% Asian/Pacific Islanders or Native Americans. Nine percent of the sample was Hispanic or Latino. Educational attainment was categorized as less than high school (3%), high school diploma (16%), some college (39%), college degree (24%), and advanced degree (19%).

Across the full bank and block design arms, a total of 15,528 persons met inclusion criteria and responded to full item banks (N=881) or 7-item subsets (N=13,695) of the items of the pain behavior bank. In addition, we excluded 125 subjects who gave repetitive strings of ten or more identical responses or had a response time less than or equal to one second per item.

There were 967 participants who responded to 43 pain behavior and a pain intensity item through the ACPA survey. Average age was 48.2 years (SD = 11.1). Eighty-one percent of the respondents were female, and 91% were Caucasian, 1.5% were African-American, and 5% were of Hispanic origin. Eighty-one percent of the participants had an education equal to or greater than high school. The data were combined with Wave 1 full-bank test data to complete the IRT analysis calibrations.

Item Descriptive Statistics

Table 1 summarizes the descriptive statistics for the pain behavior items. Although there was a total of 15,792 subjects in the pooled data (WAVE1 and ACPA), because some respondents were administered 7-item subsets, not the whole pain behavior item bank (Wave 1 block testing) individual items were responded to by approximately 3,000 subjects. The sample sizes were smaller for those items that were not included in the ACPA survey.

As can be seen in Table 1, for many of the items approximately 20% or more of subject responses are at the floor of the response distribution (indicating no pain) (range of 17.0% to 26.7%). Very few responses were at the ceiling of the response distribution (range 0.7% to 16.6%); most items had less than 5% at the ceiling.

Before the IRT analyses, the item response distributions of each item were examined. The results show that items 40 ("flung my arms and limbs around") and 41 ("screamed") each had less than 5 observations in the highest response category (6=Always). A decision was made to collapse the two highest response categories for these two items.

Evaluation of Unidimensionality

We evaluated item-total correlations and internal consistency reliability (Cronbach's alpha) for the total set of 52 pain behavior items. Item-total correlations ranged from 0.44 to 0.87 and internal consistency reliability was 0.99.

The exploratory factor analyses found that the first factor explained 90% of the variance in the pain behavior items. The CFA findings for the one factor solution for the 52 pain behavior items resulted in a CFI of 0.902 and TLI of 0.991, with RMSEA equal to 0.156 and SRMR equal to 0.035.

Based on the item-total correlations and the CFA results, we determined that there was sufficient evidence supporting unidimensionality of the pain behavior item pool.

IRT Analyses

Originally, all subjects who met the inclusion criteria were included in the IRT analysis. The graded response model was fit to these data. Items were excluded from the item list if the p-value of either $S-X^2$ and $S-G^2$ were less than 0.001. After removal of misfitting items, the IRT calibration repeated until there was no item that was significantly misfit. In the first iteration, items 1, 7, 36, 53, and 54 were excluded. In subsequent iterations, eight more items were excluded (items 4, 5, 12, 15, 19, 20, 30, and 52). The item parameters of the final 39 items are shown in Table 2. The slope parameters ranged from 3.55 ("I limped because of pain") to 5.94 ("When I was in pain I screamed"). The slopes parameters are considered larger than normally observed. The threshold values ranged from −0.72 to 2.15 reflecting a somewhat narrow coverage of the range of the pain behavior domain.

To address the unexpectedly large slope parameters we explored excluding subjects who self-reported having no pain. A total of 6,035 subjects who reported 0 or 1 to the global average pain item were excluded, and the item parameters of the 39 items were re-calibrated. The new item parameters were noticeably lower and are shown in Table 3. The slope parameters ranged from 1.81 ("I limped because of pain") to 3.41 ("I had pain so bad it made me cry"). The threshold values ranged from −2.52 to 3.21 indicating good coverage across the range of the pain behavior domain.

The 39 pain behavior items were also calibrated with the nominal model. The nominal model was constrained to have a slope parameter for the "Had no pain" response option, and equal slope parameters for all other response options. Table 4 shows the item parameters for the nominal model. The item parameters for the nominal model cannot be directly compared to those for the graded response model. However, one can gauge the magnitude of the slope by subtracting adjacent parameter estimates. For example, for item 2, the slope for the item characteristic curve of the response option "Had no pain" is the difference between the parameter $a2$ and $a1$, 6.24. The slopes for the item characteristic curves for other response options are the difference between $a3$ and $a2$, or $a4$ and $a3$, etc. Because these are constrained to be equal, thus they are all 2.98. Slopes for the response option "Had no pain" were much larger than those for the combined other response options.

Descriptive Statistics and Reliability

The derived IRT theta scores were transformed into T-scores with a mean of 50 and standard deviation (SD) of 10 based on the PROMIS general population sub-sample. The mean T-score for the overall sample was 51.2 (SD=9.4), with median of 50.8, and range from 10 to 90. Figure 2 summarizes the standard errors over the range of pain behavior scores for the total item bank, a 7-item short form, and a 7-item computerized adaptive test (CAT). For the full item bank, reliability is 0.90 or greater across most of the score distribution, and the short form and CAT have reliabilities exceeding 0.80 across the majority of the score distribution. Chronbach's alpha for the full item bank was 0.98.

Validity

To assess the concurrent validity of the pain behavior measure, we correlated the IRT theta score with the global health rating and pain intensity rating. The correlation between pain behavior scores and self-reported pain intensity ratings was 0.69 (p<0.001). The correlation between pain behavior scores and self-reported global health ratings was −0.48 (p<0.001).

We also tested the difference between pain behavior theta scores by groups defined based on the pain intensity ratings and global health ratings. Analysis of variance was used to compare mean theta scores among groups based on pain intensity rating ('none', 'mild', 'moderate', 'severe', 'very severe') (Figure 3). There were significant differences among mean pain behavior scores (p<0.0001), and all pair-wise comparisons were statistically significant after Scheffe's adjustment (all p<0.001). Analysis of variance was used to compare pain behavior scores among subjects grouped by global health ratings ('poor', 'fair', 'good', 'very good', 'excellent') (Figure 4). We found significant differences by global health groups (p<0.0001), and all pair-wise comparisons were significant with Scheffe's adjustment (all p<0.001).

Differential Item Functioning

Ordinal logistic regression [35] was used to assess items for DIF. DIF was assessed for gender, education (high school or less vs. high school and above), and age (< 65 vs. >=65). The IRT theta score was used as the conditioning variable. Non-uniform DIF was first examined followed by uniform DIF. DIF was determined to be significant when the p-value for the chi-square was less then 0.001. No items had non-uniform DIF or uniform DIF detected due to education levels (data not shown). No items had non-uniform DIF due to gender (data not shown). Item 18 ("ask for help doing things needed to be done") had uniform DIF due to gender. No non-uniform DIF for the pain behavior items was detected due to age (data not shown). However, there were five items that showed uniform DIF due to age. After reviewing the content of these items, four of these five items were associated with age-related movement: "moved extremely slowly" (item 8), "bend over while walking" (item 22)," "use cane or something for support" (item 29)," and "move my limbs protectively" (item 50). The other item that showed uniform DIF due to age was item 48 ("gritted my teeth or clenched my jaw"). In all cases, the older participants demonstrated more age related movement problems (i.e., the items were easier to endorse for older participants).

DISCUSSION

As part of the larger PROMIS project, we developed a self-report pain behavior measure and 39 item bank covering a wide array of pain behaviors. This comprehensive pain behavior item bank can be used to develop static short-forms and focused measures of pain behavior for different clinical studies and patient populations. The pain behavior item bank can also be administered as a computerized adaptive test (CAT). CAT allows for more individualized assessment of pain behavior where the items administered are tailored to individual respondents and their severity level on the domain of pain behavior. CAT enables the assessment of the health domain with very few items (see Figure 2). The set of calibrated pain behavior items also allows the researcher to intelligently design a short-form pain behavior scale which targets a specific region of the pain behavior construct.

Based on the results of this study, the PROMIS pain behavior item bank has good evidence supporting unidimensionality, model fit, and coverage of the pain behavior construct, and preliminary evidence supporting construct validity. For the final selected items, a one-factor CFA resulted in excellent fit and good item-total correlations indicating unidimensionality of these pain behavior items. Although the item bank covers different aspects of pain behavior, this set of items appears to collectively measure a single construct, pain behavior. Our findings differ from previous research on pain behavior which indicates multiple pain behavior sub-domains [], however previous research has not used a modern measurement theory approach to pain behavior assessment.

We were challenged by the large slopes from the IRT analysis of the PROMIS general population sample. These large slopes were likely attributable to the significant proportion of subjects who reported no pain and, therefore, for whom the pain behavior items were not

relevant. When the IRT analysis was performed including the sample from the ACPA and including only those subjects who reported some pain, the item slopes were lower and more consistent with those expected in an IRT calibration. Also, a nominal IRT model, which estimated slopes for responses of "no pain" and for all other responses resulted in slopes more consistent with expectation. We recommend that for future applications of the pain behavior item banks, researchers use the item calibrations from the sample of subjects with some pain (Table 3). We anticipate that the pain behavior item bank will be used primarily in samples of individuals experiencing some pain.

Reliability of a new patient-reported outcome is most often evaluated using internal consistency reliability coefficients (i.e., Cronbach's alpha). In IRT analyses, reliability or measurement precision is assessed across different ranges of scores and provides information as to the reliability of the scores at different levels of the targeted domain. We found evidence that the full pain behavior item bank and two short-form measures have acceptable reliability for group comparisons across the range of the pain behavior construct.

We have some preliminary evidence supporting the validity of the pain behavior scores. The pain behavior scores were strongly related to pain intensity scale scores and moderately related to self-ratings of general health status. Mean pain behavior scores vary significantly by groups based on verbal ratings of pain severity and by general health status. Those subjects in the moderate or severe pain severity groups were more likely to report greater pain behaviors. Clearly, additional evidence supporting the validity of the pain behavior scores is needed, especially from clinical populations and patients with different types of chronic pain. Future studies will also need to compare observer rated pain behavior scales with the PROMIS self-reported pain behavior scores to further demonstrate construct validity.

We identified little evidence of differential item functioning among the pain behavior items. None of the items showed DIF by education level, and only one item demonstrated uniform DIF by gender (asking for help). Several items, related to mobility and movement demonstrated uniform DIF related to age. When selecting subsets of PROMIS pain behavior items for specific patient populations, the recommendation is to select those items with no evidence of DIF.

The PROMIS pain behavior item bank overlaps in content with other patient rated pain behavior instruments [5,10,12,14,15,20,23,24,27,29,30,32]. For example, the PBCL contains items covering distorted ambulation, affective distress, facial/audible expressions, and help seeking behavior. The current PROMIS item bank also includes items representing these aspects of pain behavior as well as items assessing other pain behaviors (as reclining, trying to minimize movement, isolating oneself from others, crying, screaming, increased body tension, avoiding physical contact with others, or massaging a painful area.). Future research is needed to evaluate the relationship between the PROMIS pain behavior measure and the PBCL. Future studies should also examine how self-reports of pain behavior collected using the PROMIS measures relate to direct observations of pain behavior.

There are several limitations which should be considered when interpreting the results of these psychometric analyses. First, the psychometric analysis of the self-reported pain behaviors in the PROMIS item bank included general population, clinic samples, and members of the ACPA. A significant proportion of the general population did not report experiencing pain, and this required us to perform several additional IRT analyses to address this problem. For PROMIS, our interest was to have the scores for all the item banks in a metric normed to the general population. However, the inclusion of a general population sample may have affected the item parameter estimates; in fact such an effect was evidenced in the reduced slope estimates when those with very low pain behavior scores were excluded from the calibration. Second, although the ACPA sample reported various chronic pain conditions and moderate to severe

pain, differences between the demographic characteristics of the ACPA sample (e.g., 81% female, 91% Caucasian) may result in less generalizability of the current findings to the general population of people with chronic pain. Future research is needed to examine whether this resulted in any significant impact on the pain behavior scores, and further research is needed in patients with various types of chronic pain conditions (e.g., osteoarthritis, low back pain, neuropathic pain, fibromyalgia, etc.). Finally, we were unable to clearly distinguish individuals experiencing acute versus chronic pain in the general population sample, and additional research is needed to examine the PROMIS pain behavior items in these different pain populations.

In conclusion, we have developed a patient-reported pain behavior item bank based on careful and systematic item generation, patient interviews and the application of modern measurement theory methods. Based on the results of this study, we have evidence supporting satisfactory measurement precision across a range of the pain behavior construct, and some evidence supporting construct validity. Additional research is now needed to further investigate the construct validity, test-retest reliability and responsiveness of the new pain behavior item bank in longitudinal studies. Short-form and CAT pain behavior scales need to be developed and evaluated in clinical populations. The PROMIS pain behavior item bank is now available for applications in clinical studies of patients with chronic diseases and chronic pain (see www.nihpromis.org). This new patient-reported pain behavior measure has the potential to improve measurement of pain behavior and to increase our understanding of the effects of pain on patients' lives and of the effects of interventions.

Acknowledgments

REFERENCES

1. Bjorner, JB.; Smith, KJ.; Orlando, M.; Stone, C.; Thissen, D.; Sun, X. IRTFIT: A Macro for Item Fit and Local Dependence Tests under IRT Models. Lincoln, RI: QualityMetric Incorporated; 2006.

2. Bock RD. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika 1972;39:29–51.

3. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap cooperative group through its first two years. Med Care 2007;45:S3–S11. [PubMed: 17443116]

4. DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: the PROMIS qualitative item review. Med Care 2007;45:S12–S21. [PubMed: 17443114]

5. Dirks JF, Wunder J, Kinsman R, McElhinny J, Jones NF. A Pain Rating Scale and a Pain Behavior Checklist for clinical use: development, norms, and the consistency score. Psychother Psychosom 1993;59:41–49. [PubMed: 8441794]

6. Fordyce, WE. Behavioral Methods for Chronic Pain and Illness. St. Louis, MO: C.V. Mosby; 1976.

7. Hadjistavropoulos T, Herr K, Turk DC, Fine PG, Dworkin R, Helme R, Jackson K, Parmalee P, Rudy T, Beattie B, Chibnall J, Craig K, Ferrell B, Ferrell B, Fillingim R, Gagliese L, Gallagher R, Gibson S, Harrison E, Katz B, Keefe FJ, Lieer S, Lussier D, Schmader K, Tait R, Weiner D, Williams J. An interdisciplinary expert consensus statement on assessment of pain in older persons. Clin J Pain 2007;23:S1–S43. [PubMed: 17179836]

8. Hambleton, RK. Applications of item response theory to improve health outcomes assessment: developing item banks, linking instruments, and computer adaptive testing. In: Lipscomb, J.; Gotay, CC.; Snyder, C., editors. Outcomes Assessment in Cancer: Measures, Methods, and Applications. Cambridge: Cambridge University Press; 2005. p. 445-464.

9. Hays RD, Bjorner J, Revicki DA, Spritzer K, Cella D. Development and psychometric evaluation of the PROMIS short-form health status measure. Qual Life Res. in press

10. Husebo BS, Strand LI, Moe-Nilssen R, Husebo SB, Snow AL, Ljunggren AE. Mobilization-Observation-Behavior-Intensity-Dementia Pain Scale (MOBID): Development and validation of a nurse-administered pain assessment tool for use in dementia. J Pain Symptom Manage 2007;34:67–80. [PubMed: 17509814]

11. Jensen, MP. Validity of self-report and observational measures. In: Jensen, TS.; Turner, JA., editors. Proceedings of the 8th World Congress on Pain: Progress in Pain Research and Management. Seattle, WA: IASP Press; 1997.

12. Keefe FJ, Block AR. Development of an observation method for assessing pain behavior in chronic low back pain patients. Behavior Therapy 1982;13:363–375.

13. Keefe, FJ.; Williams, DA.; Smith, SJ. Assessment of pain behaviors. In: Turk, DC.; Melzack, R., editors. Handbook of Pain Assessment. New York, NY: Guilford Press; 2001. p. 170-187.

14. Kerns RD, Hayhornthwaite J, Rosenberg R, Southwick S, Giller EL, Jacob MC. The Pain Behavior Check List (PBCL): factor structure and psychometric properties. J Behav Med 1991;14:155–167. [PubMed: 1880794]

15. McDanieI LK, Anderson KO, Bradley LA, Young LD, Turner RA, Agudelo CA, Keefe FJ. Development of an observation method for assessing pain behavior in rheumatoid arthritis patients. Pain 1986;24:165–184. [PubMed: 3960569]

16. Muthén, LK.; Muthén, B. Mplus User's Guide. Vol. Third Edition. Los Angeles, CA: Muthén & Muthén; 2004.

17. Orlando M, Thissen D. Likelihood-based item-fit indices for dichotomous item response theory models. Appl Psychol Measure 2000;24:50–64.

18. Orlando M, Thissen D. Further investigation of the performance of S - X2: An item fit index for use with dichotomous item response theory models. Appl Psychol Measure 2003;27:289–298.

19. Osman A, Barrios FX, Kopper B, Osman JR, Grittmann L, Troutman JA, Panak WJ. The Pain Behavior Check List (PBCL): psychometric properties in a college sample. J Clin Psychol 1995;51:775–782. [PubMed: 8778125]

20. Prkachin KM, Hughes E, Schultz I, Joy P, Hunt D. Real-time assessment of pain behavior during clinical assessment of low back pain patients. Pain 2002;95:23–30. [PubMed: 11790464]

21. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Honghu L, Gershon R, Reise SP, Lai J, Cella D. Psychometric evaluation and calibration of health-related quality of life items banks: plans for the patient-reported outcome measurement information system (PROMIS). Med Care 2007;45(5 Suppl 1):S22–S31. [PubMed: 17443115]

22. Reise, SP. Item response theory and its applications for cancer outcomes measurement. In: Lipscomb, J.; Gotay, CC.; Snyder, C., editors. Outcomes Assessment in Cancer: Measures, Methods, and Applications. Cambridge: Cambridge University Press; 2005. p. 425-444.

23. Richards JS, Nepomuceno C, Riles M, Suer A. Assessing pain behavior: the UAB pain behavior scale. Pain 1982;14:393–398. [PubMed: 7162841]

24. Terstegen C, Koot HM, de Boer JB, Tibboel D. Measuring pain in children with cognitive impairment: pain response to surgical procedures. Pain 2003;103:187–198. [PubMed: 12749973]

25. Thissen, D.; Chen, WH.; Bock, D. MULTILOG. Lincolnwood, IL: Scientific Software International; 2002.

26. Turk DC, Dworkin RH, Revicki DA, Harding G, Burke LB, Cella D, Cleeland CS, Cowan P, Farrar JT, Hertz S, Max MB, Rappaport BA. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. Pain 2008;137:276–285. [PubMed: 17937976]

27. Turk DC, Wack JT, Kerns RD. An empirical examination of the "pain-behavior" construct. J Behav Med 1985;8:119–130. [PubMed: 4032472]

28. Van der Linden, WJ.; Hambleton, RK., editors. Handbook of Modern Item Response Theory. New York: Springer; 1997.

29. Vlaeyen JW, Pernot DF, Kole-Snijders AM, Schuerman JA, Van Eek H, Groenman NH. Assessment of the components of observed chronic pain behavior: the Checklist for Interpersonal Pain Behavior (CHIP). Pain 1990;43:337–347. [PubMed: 2293144]

30. Waddell G, et al. Non-organic physical signs in low back pain. Spine 1980;5:117–125. [PubMed: 6446157]

31. Waters, SJ.; Dixon, KE.; Keefe, FJ. Ayers, S.; Baum, A.; McManus, C.; Newman, S.; Wallston, K.; Weinman, J.; West, R., editors. Cambridge Handbook of Psychology, Health and Medicine. 2nd EditionCambridge UK: Cambridge University Press; 2007. Pain Assessment. p. 300-303.

32. Weiner D, Pieper C, McConnell E, Martinez S, Keefe F. Pain measurement in elders with chronic low back pain: traditional and alternative approaches. Pain 1996;67:461–467. [PubMed: 8951942]

33. Wilkie DJ, Keefe FJ, Dodd MJ, Copp LA. Behavior of patients with lung cancer: Description and associations with oncologic and pain variables. Pain 1992;51:231–240. [PubMed: 1484719]

34. Willis, G. Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, CA: Sage Publications; 2004.

35. Zumbo, BD. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense; 1999. A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores.
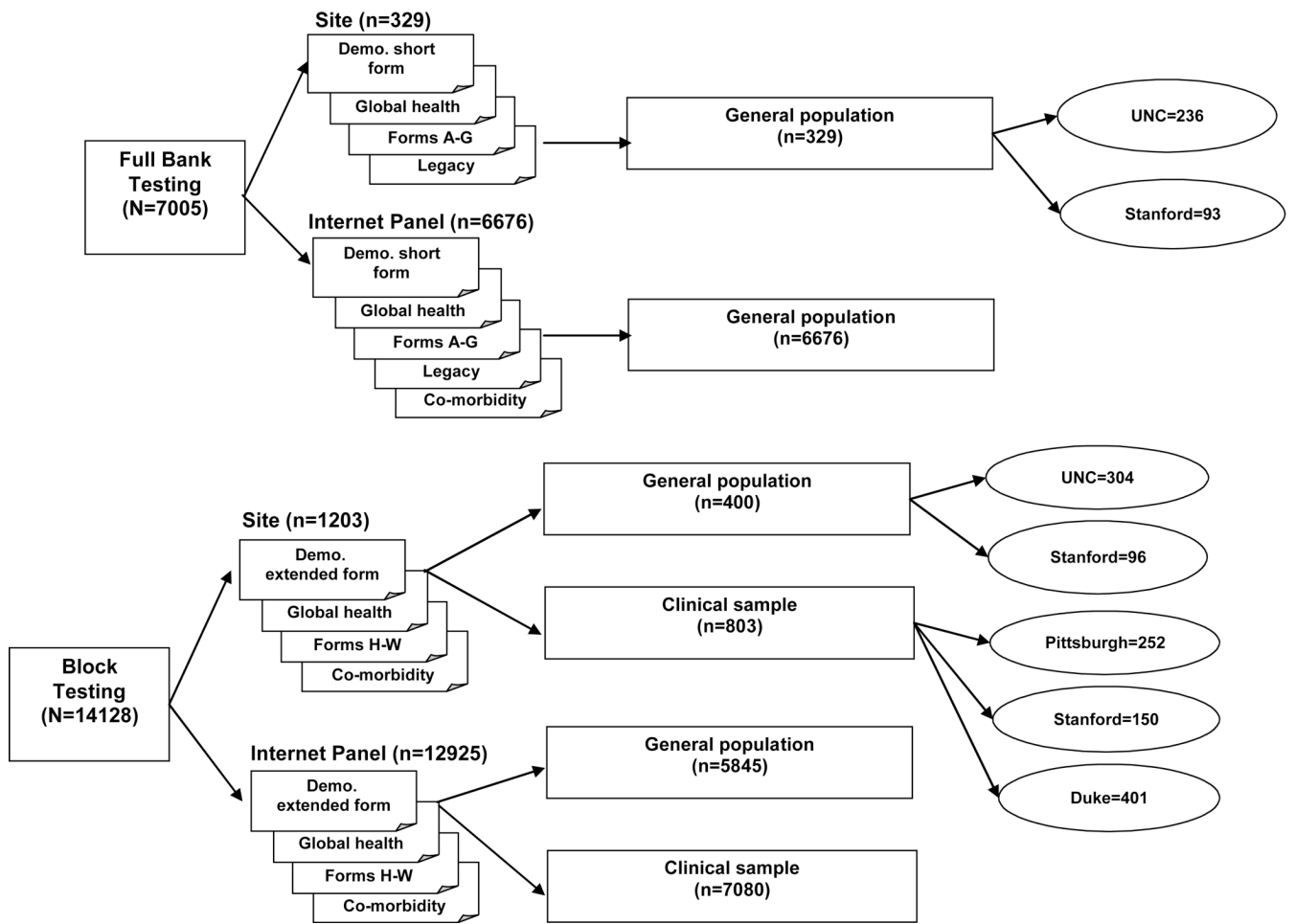
Figure 1.
PROMIS Wave 1 Sample
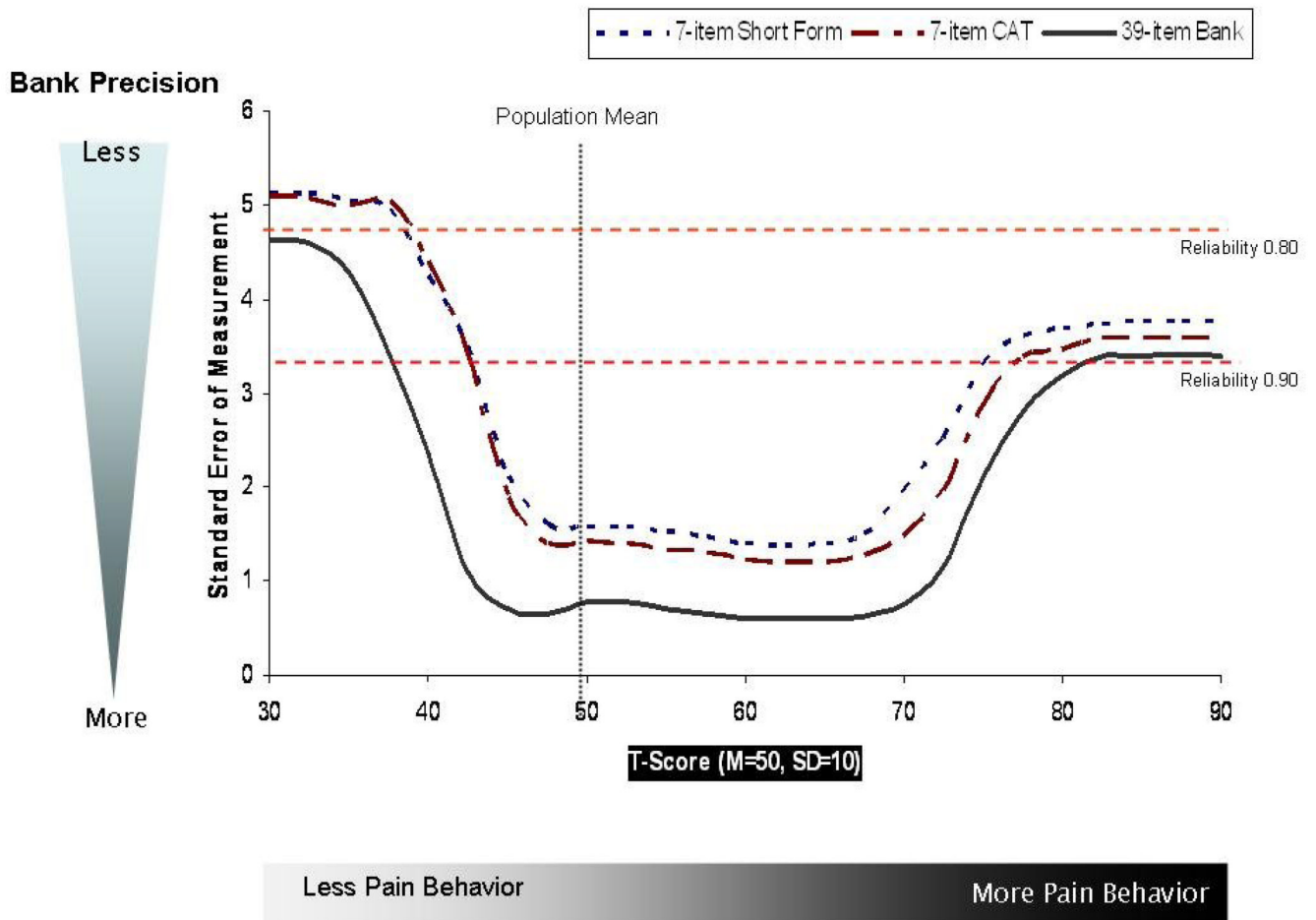
Figure 2.
Standard Error of Measurement for Pain Behavior Bank, Short Form, and 7-item CAT
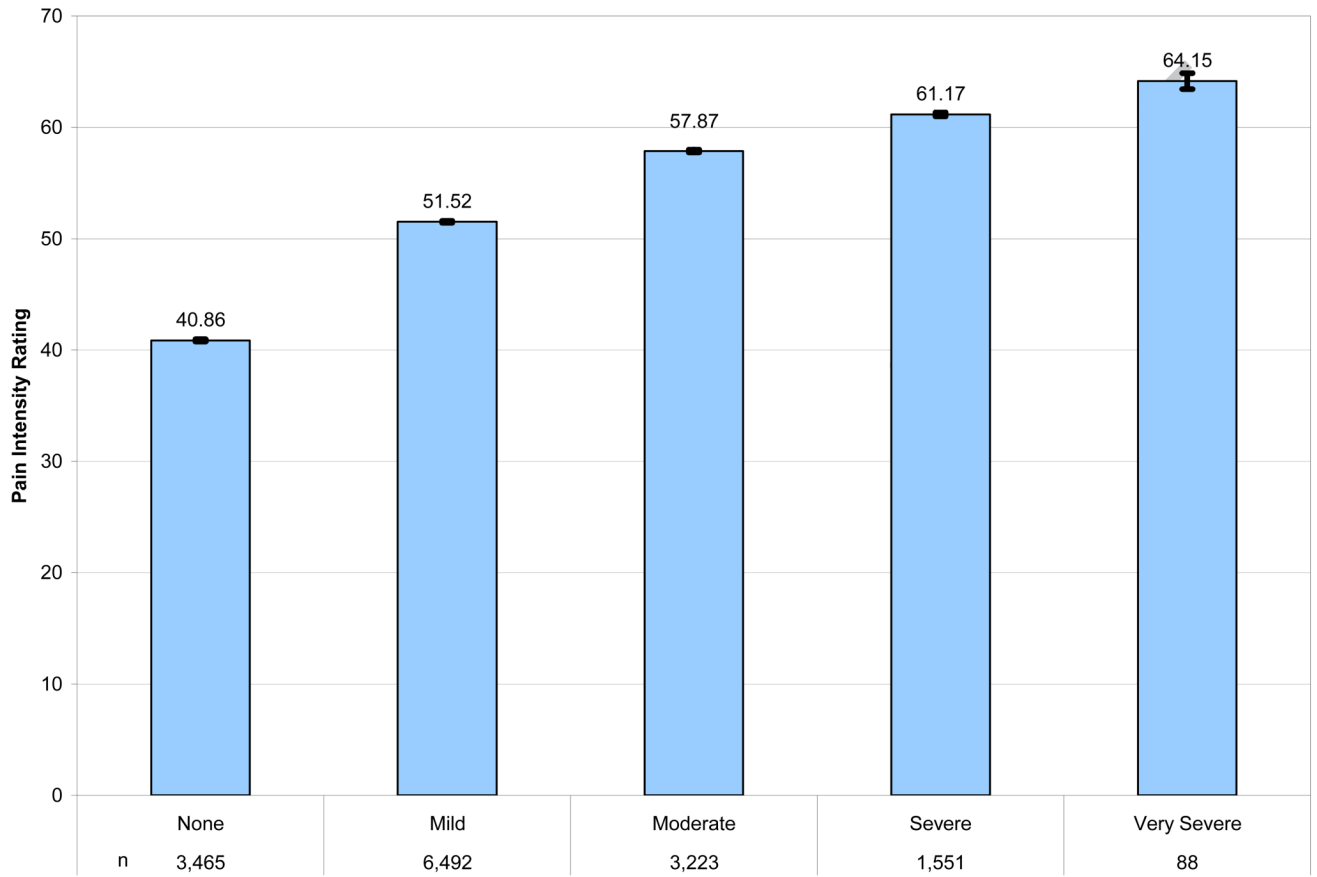
All groups are statistically significantly different from each other at $P < 0.0001$

Figure 3.
Mean Pain Behavior T-Scores by Pain Intensity
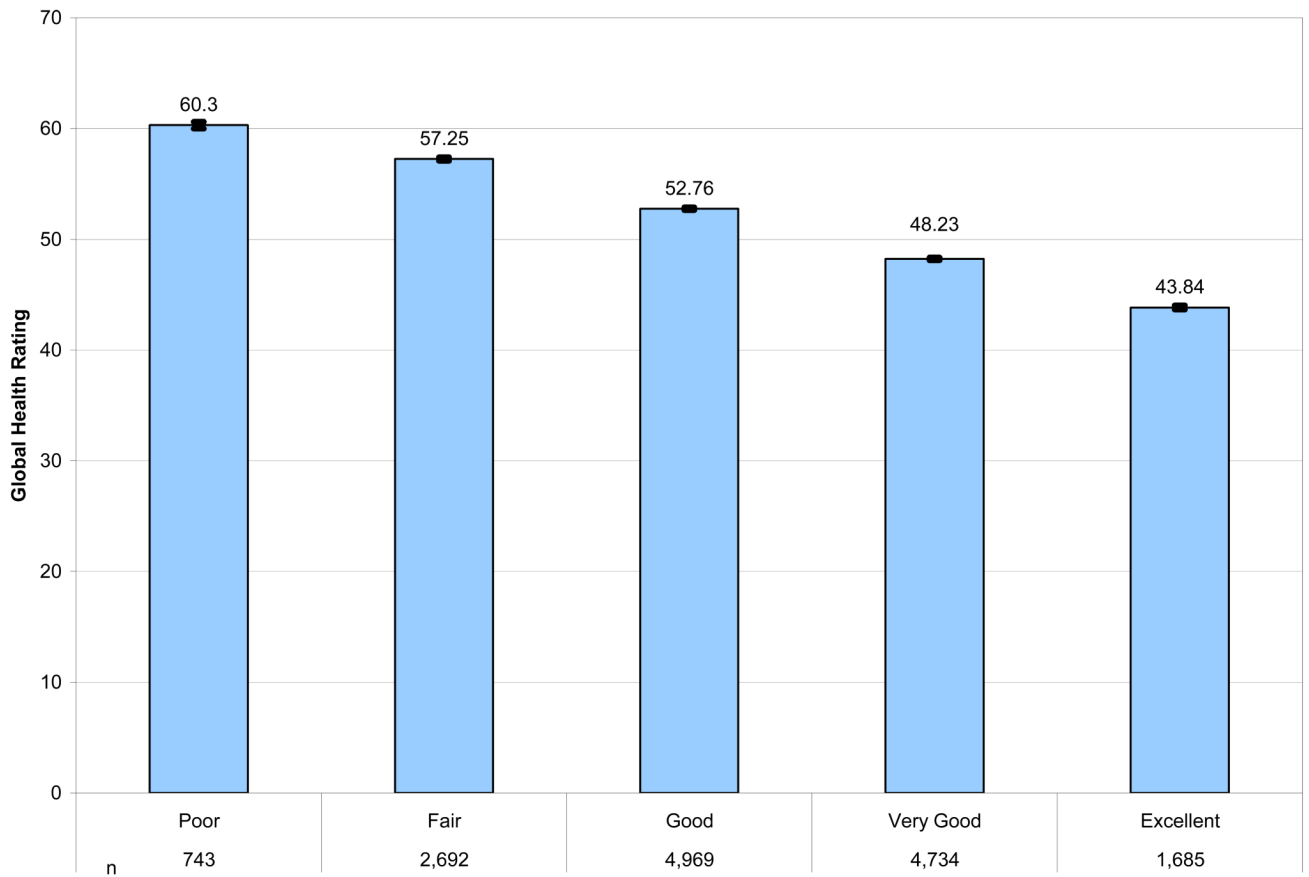
All groups are statistically significantly different from each other at *P* < 0.0001

Figure 4.
Mean Pain Behavior T-Scores by Global Health Status

Table 1

Descriptive Statistics for PROMIS Pain Behavior Items: Pooled WAVE1 and ACPA Data

ID	Item	N	Mean (SD)	Median	Range	Floor (%)	Ceiling (%)
PAINBE1	When I was in pain I walked with a limp or distorted gait	3594	3.2 (1.65)	3.0	1.0–6.0	676 (18.8%)	422 (11.7%)
PAINBE2	When I was in pain I became irritable	3606	3.3 (1.51)	3.0	1.0–6.0	612 (17.0%)	260 (7.2%)
PAINBE3	When I was in pain I grimaced	3598	3.2 (1.43)	3.0	1.0–6.0	615 (17.1%)	156 (4.3%)
PAINBE4	When I was in pain I took medication for the pain	3431	3.5 (1.72)	4.0	1.0–6.0	647 (18.9%)	570 (16.6%)
PAINBE5	When I was in pain I talked about the pain	3434	3.0 (1.33)	3.0	1.0–6.0	656 (19.1%)	98 (2.9%)
PAINBE6	When I was in pain I would lie down	3430	3.1 (1.48)	3.0	1.0–6.0	667 (19.4%)	156 (4.5%)
PAINBE7	When I was in pain I rubbed the site of pain	3439	3.5 (1.57)	4.0	1.0–6.0	641 (18.6%)	284 (8.3%)
PAINBE8	When I was in pain I moved extremely slowly	2661	2.8 (1.43)	3.0	1.0–6.0	691 (26.0%)	87 (3.3%)
PAINBE9	When I was in pain I became angry	3464	2.7 (1.32)	2.0	1.0–6.0	702 (20.3%)	85 (2.5%)
PAINBE11	When I was in pain I clenched my teeth	3399	2.7 (1.39)	2.0	1.0–6.0	665 (19.6%)	117 (3.4%)
PAINBE12	When I was in pain I clutched the part that hurt	3407	3.0 (1.42)	3.0	1.0–6.0	643 (18.9%)	122 (3.6%)
PAINBE13	When I was in pain I tried to stay very still	3403	3.0 (1.44)	3.0	1.0–6.0	665 (19.5%)	112 (3.3%)
PAINBE15	When I was in pain I walked carefully to keep from making my pain worse	3036	3.4 (1.70)	4.0	1.0–6.0	606 (20.0%)	412 (13.6%)
PAINBE16	When I was in pain I appeared upset or sad	3035	3.0 (1.48)	3.0	1.0–6.0	588 (19.4%)	168 (5.5%)
PAINBE17	When I was in pain I gasped	3041	2.6 (1.23)	2.0	1.0–6.0	600 (19.7%)	40 (1.3%)
PAINBE18	When I was in pain I asked for help doing things that needed to be done	3056	2.8 (1.41)	2.0	1.0–6.0	650 (21.3%)	103 (3.4%)
PAINBE19	When I was in pain I tried to get relief by changing my posture	3062	3.7 (1.71)	4.0	1.0–6.0	607 (19.8%)	456 (14.9%)
PAINBE20	When I was in pain I gave a detailed description of the pain to others	2258	2.4 (1.17)	2.0	1.0–6.0	603 (26.7%)	16 (0.7%)
PAINBE21	When I was in pain it showed on my face (squinching eyes, opening eyes wide, frowning)	3056	3.1 (1.53)	3.0	1.0–6.0	609 (19.9%)	201 (6.6%)
PAINBE22	Pain caused me to bend over while walking	3110	2.4 (1.25)	2.0	1.0–6.0	687 (22.1%)	64 (2.1%)
PAINBE23	When I was in pain I asked one or more people to leave me alone	3112	2.5 (1.25)	2.0	1.0–6.0	663 (21.3%)	54 (1.7%)

ID	Item	N	Mean (SD)	Median	Range	Floor (%)	Ceiling (%)
						310 (10.0%)	
PAINBE24	When I was in pain I moved stiffly	3109	3.3 (1.63)	4.0	1.0–6.0	631 (20.3%)	
PAINBE25	When I was in pain I called out for someone to help me	3075	2.2 (1.03)	2.0	1.0–6.0	733 (23.8%)	22 (0.7%)
PAINBE26	Pain caused me to curl up in a ball	3067	2.3 (1.13)	2.0	1.0–6.0	705 (23.0%)	22 (0.7%)
PAINBE27	I had pain so bad it made me cry	3071	2.4 (1.28)	2.0	1.0–6.0	713 (23.2%)	46 (1.5%)
PAINBE28	When I was in pain I squirmed	3067	2.7 (1.43)	2.0	1.0–6.0	704 (23.0%)	101 (3.3%)
PAINBE29	When I was in pain I used a cane or something else for support	3489	2.4 (1.35)	2.0	1.0–6.0	765 (21.9%)	192 (5.5%)
PAINBE30	When I was in pain I talked about the pain with one or more people	2694	2.6 (1.32)	3.0	1.0–6.0	715 (26.5%)	54 (2.0%)
PAINBE31	I limped because of pain	3488	3.0 (1.59)	3.0	1.0–6.0	722 (20.7%)	301 (8.6%)
PAINBE32	When I was in pain I became quiet and withdrawn	2879	2.7 (1.36)	2.0	1.0–6.0	645 (22.4%)	66 (2.3%)
PAINBE33	When I was in pain I frowned	3674	3.3 (1.47)	4.0	1.0–6.0	643 (17.5%)	186 (5.1%)
PAINBE34	When I was in pain I asked for help when walking or changing positions	2880	2.0 (0.87)	2.0	1.0–6.0	672 (23.3%)	8 (0.3%)
PAINBE35	When I was in pain I groaned	3683	2.8 (1.32)	3.0	1.0–6.0	648 (17.6%)	90 (2.4%)
PAINBE36	When I was in pain I walked in a way that protected me from the pain	2878	3.0 (1.57)	3.0	1.0–6.0	714 (24.8%)	166 (5.8%)
PAINBE37	When I was in pain I isolated myself from others	3690	3.0 (1.50)	3.0	1.0–6.0	668 (18.1%)	173 (4.7%)
PAINBE38	When I was in pain I drew my knees up	2892	2.2 (1.04)	2.0	1.0–6.0	690 (23.9%)	16 (0.6%)
PAINBE39	When I was in pain I moaned, whined or whimpered	3708	2.6 (1.28)	2.0	1.0–6.0	720 (19.4%)	75 (2.0%)
PAINBE40	When I was in pain I flung my arms or limbs around	2913	1.9 (0.71)	2.0	1.0–6.0	749 (25.7%)	3 (0.1%)
PAINBE41	When I was in pain I screamed	2912	1.9 (0.67)	2.0	1.0–6.0	730 (25.1%)	4 (0.1%)
PAINBE42	When I was in pain my upper body would tense up	3711	3.0 (1.56)	3.0	1.0–6.0	730 (19.7%)	261 (7.0%)
PAINBE43	When I was in pain I walked carefully	3235	3.5 (1.75)	4.0	1.0–6.0	653 (20.2%)	528 (16.3%)
PAINBE44	When I was in pain I bit or pursed my lips	3246	2.7 (1.34)	2.0	1.0–6.0	653 (20.1%)	72 (2.2%)
PAINBE45	When I was in pain I thrashed	3236	2.2 (1.03)	2.0	1.0–6.0	663 (20.5%)	24 (0.7%)
PAINBE46	When I was in pain I protected the part of my body that hurt	2436	3.0 (1.59)	3.0	1.0–6.0	603 (24.8%)	162 (6.7%)

ID	Item	N	Mean (SD)	Median	Range	Floor (%)	Ceiling (%)
PAINBE47	When I was in pain my body became stiff	2447	2.7 (1.45)	2.0	1.0–6.0	599 (24.5%)	90 (3.7%)
PAINBE48	When I was in pain I clenched my jaw or gritted my teeth	2450	2.5 (1.33)	2.0	1.0–6.0	601 (24.5%)	61 (2.5%)
PAINBE49	When I was in pain I winced	3241	3.3 (1.49)	4.0	1.0–6.0	591 (18.2%)	160 (4.9%)
PAINBE50	When I was in pain I moved my limbs protectively	3230	3.3 (1.62)	3.0	1.0–6.0	623 (19.3%)	320 (9.9%)
PAINBE51	When I was in pain I avoided physical contact with others	3243	3.0 (1.55)	3.0	1.0–6.0	612 (18.9%)	222 (6.8%)
PAINBE52	When I was in pain I shifted positions to make it better	2437	3.5 (1.70)	4.0	1.0–6.0	574 (23.6%)	240 (9.8%)
PAINBE53	When I was in pain I moved my arms or legs stiffly	3274	3.3 (1.60)	3.0	1.0–6.0	637 (19.5%)	278 (8.5%)
PAINBE54	When I was in pain I massaged the part of my body that hurt	2483	3.3 (1.54)	4.0	1.0–6.0	543 (21.9%)	133 (5.4%)

Table 2

IRT Item Parameters and Fit Statistics for PROMIS Pain Behavior Items: Pooled Wave1, CORE Cancer, and ACPA Data (N = 15,834)

ID	Item	Slope[1]	Category Threshold[1]					Item Fit Statistics[2]				
		a	b1	b2	b3	b4	b5	S-G2	Prob_G2	S-X2	Prob_X2	df
PAINBE2	When I was in pain I became irritable	5.00	−0.65	0.06	0.51	1.06	1.55	84.01	0.1575	79.97	0.2431	72
PAINBE3	When I was in pain I grimaced	4.56	−0.66	0.06	0.51	1.16	1.75	101.7	0.0315	93.11	0.1020	77
PAINBE6	When I was in pain I would lie down	4.20	−0.66	0.15	0.54	1.10	1.69	71.62	0.6208	65.54	0.7984	76
PAINBE8	When I was in pain I moved extremely slowly	5.19	−0.58	0.07	0.47	0.96	1.41	107.6	0.0081	104.0	0.0151	75
PAINBE9	When I was in pain I became angry	4.62	−0.61	0.50	0.89	1.36	1.83	94.46	0.0328	84.73	0.1270	71
PAINBE11	When I was in pain I clenched my teeth	3.92	−0.72	0.50	0.83	1.31	1.80	90.52	0.0691	76.90	0.3246	72
PAINBE13	When I was in pain I tried to stay very still	3.92	−0.71	0.22	0.62	1.18	1.80	127.8	0.0002	108.4	0.0086	76
PAINBE16	When I was in pain I appeared upset or sad	5.15	−0.69	0.29	0.69	1.15	1.60	56.48	0.9105	54.02	0.9438	72
PAINBE17	When I was in pain I gasped	5.12	−0.67	0.52	0.92	1.44	1.95	47.79	0.9635	47.79	0.9634	67
PAINBE18	When I was in pain I asked for help doing things that needed to be done	3.72	−0.63	0.40	0.81	1.30	1.82	105.4	0.0145	81.35	0.3163	76
PAINBE21	When I was in pain it showed on my face (squinching eyes, opening eyes wide, fro	5.13	−0.66	0.13	0.56	1.06	1.52	84.56	0.1673	72.84	0.4833	73
PAINBE22	Pain caused me to bend over while walking	3.62	−0.64	0.70	1.03	1.48	1.98	91.68	0.0499	71.80	0.4510	71
PAINBE23	When I was in pain I asked one or more people to leave me alone	4.50	−0.64	0.70	1.00	1.41	1.93	65.43	0.5315	60.06	0.7135	67
PAINBE24	When I was in pain I moved stiffly	4.11	−0.70	−0.04	0.33	0.89	1.37	109.7	0.0128	91.85	0.1528	79
PAINBE25	When I was in pain I called out for someone to help me	4.59	−0.59	0.85	1.25	1.70	2.07	46.44	0.9677	46.48	0.9674	66
PAINBE26	Pain caused me to curl up in a ball	5.51	−0.64	0.81	1.10	1.49	1.95	28.38	0.9997	26.72	0.9999	59
PAINBE27	I had pain so bad it made me cry	5.90	−0.65	0.73	0.99	1.35	1.80	23.30	1.0000	23.62	1.0000	64
PAINBE28	When I was in pain I squirmed	4.59	−0.65	0.41	0.73	1.22	1.72	76.11	0.3477	70.11	0.5410	72
PAINBE29	When I was in pain I used a cane or something else for support	3.79	−0.64	0.87	1.10	1.38	1.65	70.03	0.3762	54.51	0.8635	67
PAINBE31	I limped because of pain	3.55	−0.72	0.27	0.59	1.07	1.49	97.01	0.0614	87.43	0.1954	77
PAINBE32	When I was in pain I became quiet and withdrawn	4.58	−0.71	0.26	0.62	1.13	1.67	81.04	0.2690	73.77	0.4858	74
PAINBE33	When I was in pain I frowned	4.73	−0.71	0.04	0.44	1.08	1.61	90.34	0.0709	74.75	0.3891	72
PAINBE34	When I was in pain I asked for help when walking or changing	4.31	−0.66	0.94	1.29	1.70		32.02	0.9992	29.98	0.9997	61

ID	Item	Slope[1]	Category Threshold[1]					Item Fit Statistics[2]				
		a	b1	b2	b3	b4	b5	S-G2	Prob_G2	S-X2	Prob_X2	df
PAINBE35	positions When I was in pain I groaned	4.11	−0.72	0.39	0.84	1.36	1.89	72.10	0.5077	54.80	0.9448	73
PAINBE37	When I was in pain I isolated myself from others	4.88	−0.70	0.42	0.70	1.09	1.64	99.16	0.0125	84.46	0.1146	70
PAINBE38	When I was in pain I drew my knees up	4.19	−0.67	0.72	1.00	1.52	2.05	61.75	0.6583	60.40	0.7024	67
PAINBE39	When I was in pain I moaned, whined or whimpered	4.72	−0.67	0.53	0.91	1.36	1.84	61.97	0.7421	51.67	0.9507	70
PAINBE40	When I was in pain I flung my arms or limbs around	4.76	−0.61	1.11	1.39	1.80		36.95	0.9818	35.54	0.9885	57
PAINBE41	When I was in pain I screamed	5.94	−0.63	1.08	1.38	1.76		27.44	0.9995	33.36	0.9930	56
PAINBE42	When I was in pain my upper body would tense up	4.26	−0.67	0.33	0.62	1.06	1.50	84.67	0.1861	79.20	0.3183	74
PAINBE43	When I was in pain I walked carefully	4.04	−0.64	0.04	0.35	0.82	1.23	90.31	0.2483	84.19	0.4123	82
PAINBE44	When I was in pain I bit or pursed my lips	4.26	−0.65	0.56	0.88	1.35	1.92	85.27	0.1359	77.80	0.2993	72
PAINBE45	When I was in pain I thrashed	4.52	−0.63	0.97	1.27	1.67	2.15	54.75	0.7884	52.09	0.8568	64
PAINBE46	When I was in pain I protected the part of my body that hurt	3.86	−0.68	0.05	0.38	0.89	1.39	111.1	0.0100	95.92	0.0946	79
PAINBE47	When I was in pain my body became stiff	3.93	−0.68	0.30	0.59	1.06	1.60	86.52	0.2145	73.33	0.5972	77
PAINBE48	When I was in pain I clenched my jaw or gritted my teeth	4.91	−0.66	0.43	0.74	1.17	1.64	104.0	0.0065	89.09	0.0721	71
PAINBE49	When I was in pain I winced	4.52	−0.69	0.07	0.46	1.09	1.66	96.59	0.0649	83.93	0.2755	77
PAINBE50	When I was in pain I moved my limbs protectively	4.17	−0.62	0.12	0.45	0.98	1.44	90.68	0.1544	78.05	0.4770	78
PAINBE51	When I was in pain I avoided physical contact with others	4.92	−0.63	0.42	0.73	1.10	1.58	62.71	0.7745	55.65	0.9230	72

[1] Using IRT software MULTILOG [25]

[2] Using SAS macros IRTFIT [1]

* Two highest response categories are collapsed for items 34, 40, and 41

Table 3

IRT Item Parameters and Fit Statistics for PROMIS Pain Behavior Items, Excluding Subjects with 0 or 1 Global Pain Responses: Pooled WAVE-1 and ACPA Data (N=9,589)

ID	Item	Slope[1]	Category Threshold[1]					Item Fit Statistics[2]				
		a	b1	b2	b3	b4	b5	S-G2	Prob_G2	S-X2	Prob_X2	df
PAINBE2	When I was in pain I became irritable	2.90	−2.16	−0.82	−0.07	0.81	1.64	55.97	0.2936	52.29	0.4236	51
PAINBE3	When I was in pain I grimaced	2.55	−2.24	−0.86	−0.03	1.02	2.03	46.83	0.6770	47.95	0.6338	52
PAINBE6	When I was in pain I would lie down	2.26	−2.38	−0.69	0.00	0.93	1.96	48.13	0.6270	45.34	0.7314	52
PAINBE8	When I was in pain I moved extremely slowly	2.67	−2.22	−0.93	−0.18	0.68	1.49	59.31	0.2266	57.11	0.2911	52
PAINBE9	When I was in pain I became angry	2.55	−2.22	−0.12	0.56	1.35	2.20	48.56	0.4099	46.29	0.5018	47
PAINBE11	When I was in pain I clenched my teeth	2.31	−2.36	−0.06	0.48	1.26	2.12	51.73	0.3679	52.52	0.3395	49
PAINBE13	When I was in pain I tried to stay very still	2.11	−2.52	−0.56	0.14	1.13	2.27	88.90	0.0011	75.33	0.0189	52
PAINBE16	When I was in pain I appeared upset or sad	3.17	−2.18	−0.40	0.23	1.00	1.70	32.19	0.9510	30.35	0.9716	47
PAINBE17	When I was in pain I gasped	2.81	−2.19	0.00	0.68	1.59	2.46	33.91	0.9237	34.44	0.9134	47
PAINBE18	When I was in pain I asked for help doing things that needed to be done	1.94	−2.37	−0.30	0.43	1.29	2.30	72.07	0.0276	58.28	0.2252	51
PAINBE21	When I was in pain it showed on my face (squinching eyes, opening eyes wide, fro	3.18	−2.10	−0.63	0.05	0.84	1.59	56.85	0.2059	53.26	0.3138	49
PAINBE22	Pain caused me to bend over while walking	1.99	−1.97	0.30	0.88	1.65	2.54	58.90	0.1143	50.57	0.3343	47
PAINBE23	When I was in pain I asked one or more people to leave me alone	2.60	−1.97	0.29	0.79	1.45	2.35	42.43	0.7347	37.50	0.8848	49
PAINBE24	When I was in pain I moved stiffly	2.15	−2.37	−1.11	−0.42	0.57	1.48	73.67	0.0317	67.45	0.0875	53
PAINBE25	When I was in pain I called out for someone to help me	2.51	−2.01	0.58	1.25	2.07	2.83	35.09	0.8792	35.73	0.8625	46
PAINBE26	Pain caused me to curl up in a ball	2.99	−2.04	0.49	0.97	1.64	2.57	18.23	0.9995	17.39	0.9997	42
PAINBE27	I had pain so bad it made me cry	3.41	−1.98	0.28	0.71	1.33	2.20	12.42	1.0000	12.50	1.0000	47
PAINBE28	When I was in pain I squirmed	2.46	−2.20	−0.19	0.35	1.16	2.07	56.26	0.2523	54.64	0.3026	50
PAINBE29	When I was in pain I used a cane or something else for support	1.89	−2.23	0.60	1.01	1.52	2.02	45.04	0.5539	37.68	0.8322	47
PAINBE31	I limped because of pain	1.81	−2.41	−0.47	0.06	0.93	1.71	55.35	0.3859	55.70	0.3735	53
PAINBE32	When I was in pain I became quiet and withdrawn	2.73	−2.16	−0.46	0.15	1.00	1.90	51.54	0.4524	50.51	0.4932	51
PAINBE33	When I was in pain I frowned	2.85	−2.19	−0.81	−0.11	0.90	1.83	54.39	0.3834	46.98	0.6713	52

ID	Item	Slope[1]	Category Threshold[1]					Item Fit Statistics[2]				
		a	b1	b2	b3	b4	b5	S-G2	Prob_G2	S-X2	Prob_X2	df
PAINBE34	When I was in pain I asked for help when walking or changing positions	2.28	−2.19	0.71	1.35	2.07	3.21	27.53	0.9753	26.73	0.9815	44
PAINBE35	When I was in pain I groaned	2.42	−2.25	−0.20	0.53	1.40	2.33	42.02	0.7496	37.16	0.8926	49
PAINBE37	When I was in pain I isolated myself from others	2.71	−2.26	−0.24	0.26	0.92	1.85	76.82	0.0039	64.82	0.0433	47
PAINBE38	When I was in pain I drew my knees up	2.17	−2.24	0.34	0.86	1.78	2.79	37.80	0.8775	37.94	0.8739	49
PAINBE39	When I was in pain I moaned, whined or whimpered	2.86	−2.14	0.03	0.65	1.37	2.21	40.91	0.7218	36.23	0.8727	47
PAINBE40	When I was in pain I flung my arms or limbs around	2.66	−2.06	1.01	1.51	2.20		21.83	0.9835	18.33	0.9971	38
PAINBE41	When I was in pain I screamed	3.36	−2.03	0.96	1.46	2.12		25.39	0.9256	27.02	0.8860	37
PAINBE42	When I was in pain my upper body would tense up	2.32	−2.21	−0.39	0.11	0.85	1.67	53.79	0.4056	51.09	0.5098	52
PAINBE43	When I was in pain I walked carefully	2.14	−2.23	−0.94	−0.40	0.44	1.19	59.85	0.2123	51.78	0.4824	52
PAINBE44	When I was in pain I bit or pursed my lips	2.32	−2.26	0.04	0.59	1.39	2.41	54.78	0.3696	47.73	0.6426	52
PAINBE45	When I was in pain I thrashed	2.40	−2.17	0.73	1.28	2.00	2.88	29.83	0.9364	26.99	0.9732	43
PAINBE46	When I was in pain I protected the part of my body that hurt	2.12	−2.28	−0.92	−0.29	0.62	1.51	69.74	0.0613	63.70	0.1491	53
PAINBE47	When I was in pain my body became stiff	2.15	−2.25	−0.44	0.06	0.87	1.84	64.58	0.1323	50.39	0.5763	53
PAINBE48	When I was in pain I clenched my jaw or gritted my teeth	2.86	−2.09	−0.17	0.36	1.07	1.87	59.97	0.1578	48.47	0.5349	50
PAINBE49 PAINBE50	When I was in pain I winced When I was in pain I moved my limbs protectively	2.57 2.16	−2.20 −2.22	−0.78 −0.73	−0.10 −0.15	0.93 0.79	1.94 1.59	53.72 64.62	0.3338 0.1124	47.94 55.65	0.5563 0.3390	50 52
PAINBE51	When I was in pain I avoided physical contact with others	2.78	−2.07	−0.23	0.30	0.92	1.71	40.74	0.7932	34.52	0.9417	49

[1] Using IRT software MULTILOG [25]

[2] Using SAS macros IRTFIT [1]

* Two highest response categories are collapsed for items 40, and 41

Table 4

IRT Item Parameters and Fit Statistics for PROMIS Pain Behavior Items, Excluding Subjects with 0 or 1 Global Pain Responses: Pooled WAVE-1 and ACPA Data (N=9,589)

ID	Item	Slope[1]						Category Threshold[1]						Item Fit Statistics[2]				
		a1	a2	a3	a4	a5	a6	b1	b2	b3	b4	b5	b6	S-G2	Prob_G2	S-X2	Prob_X2	df
PAINBE2	When I was in pain I became irritable	0.00	6.24	9.22	12.20	15.19	18.17	0.00	5.48	6.37	6.06	3.72	0.08	122.3	0.0008	113.1	0.0046	77
PAINBE3	When I was in pain I grimaced	0.00	6.70	9.11	11.52	13.92	16.33	0.00	5.91	6.58	6.36	4.02	0.17	97.18	0.0512	87.03	0.1820	76
PAINBE6	When I was in pain I would lie down	0.00	8.44	10.32	12.20	14.08	15.97	0.00	7.43	7.58	7.48	5.87	2.87	78.86	0.4514	70.70	0.7089	78
PAINBE8	When I was in pain I moved extremely slowly	0.00	8.76	11.79	14.82	17.85	20.87	0.00	7.40	8.25	8.17	6.09	2.87	95.13	0.0583	89.76	0.1174	75
PAINBE9	When I was in pain I became angry	0.00	7.07	8.95	10.83	12.72	14.60	0.00	5.99	5.20	4.29	1.99	-1.35	138.1	0.0000	112.0	0.0023	73
PAINBE11	When I was in pain I clenched my teeth	0.00	5.37	6.72	8.08	9.43	10.79	0.00	5.35	4.36	4.04	2.53	0.19	134.2	0.0000	116.3	0.0016	75
PAINBE13	When I was in pain I tried to stay very still	0.00	5.77	7.23	8.68	10.14	11.59	0.00	5.45	5.22	5.17	3.70	0.89	179.8	0.0000	151.1	0.0000	80
PAINBE16	When I was in pain I appeared upset or sad	0.00	6.24	9.77	13.30	16.83	20.35	0.00	6.17	6.34	5.35	2.38	-1.82	78.44	0.2821	74.10	0.4095	72
PAINBE17	When I was in pain I gasped	0.00	7.77	10.01	12.26	14.51	16.75	0.00	7.43	6.61	5.32	2.13	-2.74	65.11	0.7045	62.99	0.7667	72
PAINBE18	When I was in pain I asked for help doing things that needed to be done	0.00	7.39	8.83	10.28	11.73	13.17	0.00	6.30	5.79	5.28	3.63	0.97	121.7	0.0009	96.56	0.0653	77
PAINBE21	When I was in pain it showed on my face (squinching eyes, opening eyes wide, fro	0.00	5.99	9.84	13.70	17.55	21.40	0.00	5.87	6.97	6.51	3.81	-0.32	117.4	0.0013	111.1	0.0043	75
PAINBE22	Pain caused me to bend over while walking	0.00	5.07	6.36	7.64	8.93	10.21	0.00	4.64	3.32	2.60	0.81	-1.78	136.3	0.0000	102.2	0.0137	73
PAINBE23	When I was in pain I asked one or more people to leave me alone	0.00	5.66	7.15	8.64	10.13	11.62	0.00	5.41	4.10	3.32	1.53	-1.59	101.7	0.0150	81.67	0.2279	73
PAINBE24	When I was in pain I moved stiffly	0.00	5.91	7.71	9.51	11.32	13.12	0.00	5.54	6.17	6.60	5.41	3.39	97.31	0.0913	82.70	0.3960	80
PAINBE25	When I was in pain I called out for someone to help me	0.00	6.37	7.96	9.54	11.12	12.70	0.00	5.77	4.17	2.64	-0.22	-3.53	103.0	0.0097	87.24	0.1066	72
PAINBE26	Pain caused me to curl up in a ball	0.00	7.46	9.32	11.17	13.02	14.87	0.00	7.17	5.41	4.16	1.54	-3.13	62.38	0.5691	49.05	0.9297	65
PAINBE27	I had pain so bad it made me cry	0.00	7.98	10.86	13.75	16.64	19.52	0.00	7.50	5.72	4.20	1.20	-4.30	44.75	0.9678	33.97	0.9993	64
PAINBE28	When I was in pain I squirmed	0.00	7.47	9.33	11.18	13.04	14.89	0.00	6.85	6.17	5.72	3.75	0.56	85.80	0.1850	79.95	0.3263	75
PAINBE29	When I was in pain I used a cane or something else for support	0.00	7.03	8.06	9.10	10.13	11.17	0.00	6.17	4.20	3.59	2.27	1.21	99.45	0.0077	79.75	0.1560	68
PAINBE31	I limped because of pain	0.00	6.11	7.25	8.39	9.52	10.66	0.00	5.66	5.08	5.12	3.91	2.61	109.0	0.0144	99.89	0.0563	79
PAINBE32	When I was in pain I became quiet and withdrawn	0.00	6.51	8.22	9.92	11.63	13.34	0.00	6.08	5.88	5.66	3.85	0.80	94.30	0.0878	86.55	0.2138	77
PAINBE33	When I was in pain I frowned	0.00	6.53	9.05	11.58	14.10	16.63	0.00	6.10	6.91	7.06	4.79	1.00	98.22	0.0441	80.82	0.3310	76
PAINBE34	When I was in pain I asked for help when walking or changing positions	0.00	8.38	10.01	11.64	13.27	14.90	0.00	7.57	5.51	3.80	0.90	-8.52	58.40	0.7640	49.68	0.9441	67
PAINBE35	When I was in pain I groaned	0.00	6.78	8.38	9.97	11.56	13.16	0.00	6.56	6.15	5.41	3.32	0.17	83.52	0.2102	69.00	0.6427	74
PAINBE37	When I was in pain I isolated myself from others	0.00	7.93	10.21	12.50	14.78	17.06	0.00	7.72	7.01	6.55	4.91	1.42	113.7	0.0010	96.24	0.0248	71
PAINBE38	When I was in pain I drew my knees up	0.00	6.64	7.74	8.85	9.95	11.05	0.00	6.02	4.45	3.99	1.70	-1.26	98.46	0.0172	77.04	0.2915	71
PAINBE39	When I was in pain I moaned, whined or whimpered	0.00	7.21	9.45	11.69	13.93	16.17	0.00	6.71	5.86	4.50	1.84	-2.27	71.06	0.5092	56.33	0.9127	72
PAINBE40	When I was in pain I flung my arms or limbs around	0.00	8.18	9.44	10.71	11.97		0.00	7.26	4.58	3.16	0.44		42.91	0.9619	41.89	0.9707	61
PAINBE41	When I was in pain I screamed	0.00	10.52	12.43	14.33	16.24		0.00	9.58	6.63	4.34	0.06		29.55	0.9986	30.85	0.9975	56
PAINBE42	When I was in pain my upper body would tense up	0.00	8.13	9.83	11.54	13.25	14.95	0.00	7.33	6.79	6.74	5.37	3.16	96.54	0.0404	89.68	0.1035	74
PAINBE43	When I was in pain I walked carefully	0.00	5.88	7.50	9.11	10.73	12.35	0.00	5.13	5.32	5.78	4.85	3.54	114.2	0.0109	104.5	0.0477	82
PAINBE44	When I was in pain I bit or pursed my lips	0.00	5.63	7.00	8.37	9.74	11.11	0.00	5.32	4.37	3.83	2.19	-0.79	126.1	0.0002	102.7	0.0186	75
PAINBE45	When I was in pain I thrashed	0.00	6.16	7.56	8.97	10.37	11.78	0.00	5.85	3.98	2.70	0.25	-3.44	76.13	0.2084	63.48	0.5995	67
PAINBE46	When I was in pain I protected the part of my body that hurt	0.00	7.68	9.12	10.57	12.01	13.46	0.00	6.76	6.93	7.37	6.37	4.69	120.9	0.0017	106.3	0.0220	79
PAINBE47	When I was in pain my body became stiff	0.00	8.68	10.03	11.38	12.73	14.08	0.00	7.93	7.36	7.48	6.25	4.13	71.21	0.6645	61.98	0.8933	77
PAINBE48	When I was in pain I clenched my jaw or my body	0.00	9.46	11.62	13.79	15.95	18.11	0.00	8.85	8.17	7.61	5.56	2.44	101.7	0.0121	87.38	0.1047	72

ID	Item	Slope[1]						Category Threshold[1]						Item Fit Statistics[2]				
		a1	a2	a3	a4	a5	a6	b1	b2	b3	b4	b5	b6	S-G2	Prob_G2	S-X2	Prob_X2	df
	gritted my teeth																	
PAINBE49	When I was in pain I winced	0.00	6.36	8.69	11.02	13.36	15.69	0.00	5.89	6.55	6.63	4.57	0.87	106.6	0.0144	87.96	0.1846	77
PAINBE50	When I was in pain I moved my limbs protectively	0.00	7.21	8.80	10.39	11.97	13.56	0.00	6.22	6.23	6.49	5.11	3.26	103.7	0.0388	91.12	0.1858	80
PAINBE51	When I was in pain I avoided physical contact with others	0.00	7.61	9.83	12.06	14.28	16.50	0.00	6.92	6.28	5.60	3.94	0.96	89.19	0.1258	78.08	0.3812	75

[1] Using IRT software MULTILOG [25]

[2] Using SAS macros IRTFIT [1]

* Two highest response categories are collapsed for items 40, and 41