HHS Public Access

Author manuscript

J Pain. Author manuscript; available in PMC 2018 September 01.

Published in final edited form as:

J Pain. 2017 September ; 18(9): 1096–1110. doi:10.1016/j.jpain.2017.05.001.

Responsiveness and Minimally Important Differences for Four Patient-Reported Outcomes Measurement Information System (PROMIS) Short Forms: Physical Function, Pain Interference, Depression, and Anxiety in Knee Osteoarthritis

Augustine C. Lee, MD[1], **Jeffrey B. Driban, PhD, ATC, CSCS**[1], **Lori Lyn Price, MAS**[2,3], **William F. Harvey, MD, MSc**[1], **Angie Mae Rodday, PhD, MS**[2], and **Chenchen Wang, MD, MSc**[1]

[1]Center for Complementary and Integrative Medicine, Division of Rheumatology, Tufts Medical Center, Boston, USA

[2]Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, USA

[3]Tufts Clinical and Translational Science Institute, Tufts University, Boston, USA

Abstract

Patient-Reported Outcomes Measurement Information System (PROMIS) instruments can provide valid, interpretable measures of health status among adults with osteoarthritis (OA). However, their ability to detect meaningful change over time is unknown. We evaluated the responsiveness and minimally important differences (MID) for four PROMIS Short Forms: Physical Function, Pain Interference, Depression, and Anxiety. We analyzed adults with symptomatic knee OA from our randomized trial comparing Tai Chi and physical therapy. Using baseline and 12-week scores, responsiveness was evaluated according to consensus standards by testing 6 *a priori* hypotheses of the correlations between PROMIS and legacy change scores. Responsiveness was considered high if ≥5 hypotheses were confirmed, and moderate if 3 or 4 were confirmed. MIDs were evaluated according to prospective change for people achieving previously-established MID on legacy comparators. The lowest and highest MIDs meeting *a priori* quality criteria formed an MID range for each PROMIS Short Form. Among 165 predominantly female (70%) and white (57%)

Corresponding Author: Chenchen Wang, MD, MSc, Center for Complementary and Integrative Medicine, Division of Rheumatology, Tufts Medical Center, 800 Washington Street, Box 406, Boston, MA, 02111, Phone: 617-636-3251, Fax: 617-636-1542, CWang2@Tuftsmedicalcenter.org.

Disclosures:
Conflicts of Interest: The authors declare no conflicts of interest.

Author contributions
All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Lee had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study conception and design: Lee, Driban, Price, Harvey, Rodday, Wang
Acquisition of data: Lee, Driban, Price, Harvey, Rodday, Wang
Analysis and interpretation of data: Lee, Driban, Price, Harvey, Rodday, Wang

Trial Registration: Clinicaltrials.gov NCT01258985. Registered 10 December 2010

participants, mean age was 61 years and body mass index was 33kg/m^2. PROMIS Physical Function had 5 confirmed hypotheses and Pain Interference, Depression, and Anxiety had 3 or 4. MID ranges were: Depression=3.0–3.1; Anxiety=2.3–3.4; Physical Function=1.9–2.2; and Pain Interference=2.35–2.4. PROMIS Physical Function has high responsiveness, and Depression, Anxiety, and Pain Interference have moderate responsiveness among adults with knee OA. We established the first MIDs for PROMIS in this population, and provided an important standard of reference to better apply or interpret PROMIS in future trials or clinical practice.

Keywords

Osteoarthritis; Patient-Reported Outcomes; Responsiveness; Minimally Important Difference PROMIS

Introduction

Patient-reported outcomes (PROs) are increasingly used in clinical research of osteoarthritis (OA), and may serve as tools for monitoring disease status or treatment effect in patient-centered clinical care.[67] However, many challenges associated with currently-available PROs, including the need for better accessibility, comparability, and interpretability,[5,6,34,56] have impeded their application in research or clinical-care settings.[11,23,25,56] As a result, the European League Against Rheumatism and other scientific societies have called for the use of structured methodological frameworks, such as the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN), to identify superior PRO instruments for consideration in hierarchical lists of recommended instruments.[8,11,25] PROs are meaningful to OA clinical practice and research if they are able to measure change over time (i.e. responsiveness); particularly, a minimal amount of change that can be considered clinically meaningful (i.e. minimally important difference [MID]). In particular, PROs that are succinct, easy to interpret, and applicable across heterogeneous groups of patients are most practical for clinical use because OA is commonly treated in both rheumatology and primary-care. However, no standard PROs are currently used to monitor outcomes in the routine clinical setting for OA.[25]

In the past decade, Patient-Reported Outcomes Measurement Information System (PROMIS) was created as a freely-accessible, national resource of universal PRO instruments by the National Institutes of Health (NIH), and recently completed initial calibration. PROMIS is uniquely intended to be comparable and interpretable among both the general population and across various patient groups.[10,32] Thus, as an ideal candidate for use in both research and clinical care, PROMIS represents a novel solution to the limitations of available PROs used in arthritis. PROMIS Physical Functioning has been well-characterized in rheumatoid arthritis, and was found to have superior responsiveness over existing measures.[29] Recent studies have also demonstrated clinical validity of PROMIS instruments across various chronic conditions, including rheumatoid arthritis[13]. In a cross-sectional study, we found that four PROMIS short forms had construct validity among people with knee OA.[16] Furthermore, cross-sectional known group validity (comparison with general population) and ecological validity (comparison of aggregated daily measures

with a 7-day recall score) were found among participants with OA[7]. However, as identified in 2 recent systematic reviews, the responsiveness or MIDs of PROMIS instruments are unknown in OA.[25,77] Therefore, a longitudinal examination of how PROMIS instruments perform in OA will fill an important research gap.

The purpose of this study is to evaluate the responsiveness and MIDs of four PROMIS Short Forms: Physical Function, Pain Interference, Depression, and Anxiety; which are important health outcomes among persons with knee OA.[38] We hypothesize that all four PROMIS Short Forms will demonstrate at least moderate responsiveness to treatment. These PROMIS measures are relevant for evaluating treatments in OA because both pain interference and physical function are widely recognized as primary symptoms of interest in OA clinical research[38,40,65]. Moreover, because they are able to both affect treatment efficacy and be improved [by] treatments in this patient group, psychological health outcomes such as anxiety and depression are also important outcomes of OA research[38,40,65]. Following state-of-the-art methodology from the COSMIN and PROMIS Standards,[42,43,54] this work can help researchers and clinicians meaningfully use PROMIS in future research and patient-care[33], and help health systems make informed decisions on whether to integrate PROMIS for routine electronic medical record collection.[25]

METHODS

Study participants

This study was conducted as part of an NIH-funded, randomized trial comparing Tai Chi vs. Physical Therapy among adults with symptomatic knee OA (Trial Registry #NCT01258985). There is a substantial body of evidence supporting the beneficial effects of both interventions on pain interference, physical function, depression, and anxiety[20,24,35,46,62,71,73,74]. Eligible participants had to meet the following criteria: 1) age ≥ 40 years; 2) Western Ontario and McMasters Arthritis Index (WOMAC) pain subscale > 40mm on at least 1 out of 5 questions; 3) fulfillment of the American College of Rheumatology criteria for knee OA; 4) radiographic evidence of knee OA defined as the presence of osteophytes in the tibiofemoral and/or the patellofemoral compartment, as assessed on standing anterior-posterior and lateral views; and 5) confirmation of knee pain, discomfort, or disability by clinical examination. We excluded individuals who had experience in the past year with physical therapy, Tai Chi, or similar types of alternative medicine (e.g., Qi Gong or Yoga); serious medical conditions limiting their ability to fully participate as determined by a primary care physician; intra-articular steroid injections or replacement surgery on the affected knee in the previous three months; or a Mini-Mental examination score<24.

Data were collected at Tufts Medical Center, an urban tertiary care academic hospital in Boston, USA. Both interventions were administered for 12 weeks and resulted in similar improvements in pain, function, and other health-related outcomes.[73] Therefore, we used the pooled data from participants receiving either intervention. Further details of the trial protocol[72] and primary results are published elsewhere.[73] PROMIS instruments were administered for the intended purpose of this investigation. Because establishing responsiveness among PROs requires at least 2 time points, only participants who completed

PROMIS Short Forms at baseline and follow-up were included. Ethics approval was given by Tufts Medical Center/Tufts University Human Institutional Review Board. All enrolled participants provided informed consent.

PROMIS Short Forms

PROMIS instruments are calibrated on Item Response Theory (IRT) models[58], which allows for the creation of short forms from large item banks without sacrificing precision.[10,29] The instruments use a 7-day recall period (except physical function, which measures the present time) and a 5-point Likert scale. Raw scores are converted to an interval-standardized T-score, which is calibrated on large samples of the general population to facilitate normative comparisons. T-scores are centered on a mean of 50 and a standard deviation (SD) of 10. Higher scores indicate a greater amount of the target domain.[3,51,53,57] Missing items from PROMIS instruments were imputed according to recommended algorithms from the PROMIS manuals (Supplementary Text).

Pain Interference short form (v.6b) measures self-reported consequences of pain on relevant aspects of life, including engagement with social, cognitive, emotional, physical, and recreational activities.[3] Physical Function short form (v.10a) measures the self-reported capabilities of physical activities, including dexterity, mobility, neck and back functioning, and instrumental activities of daily living.[3] Anxiety short form (v.7a) measures self-reported fear (fearfulness, panic), anxious misery (worry, dread), hyperarousal (tension, nervousness, restlessness), and somatic symptoms related to arousal (racing heart, dizziness).[51] Depression short form (v.8b) measures self-reported negative mood (sadness, guilt), views of self (self-criticism, worthlessness), social cognition (loneliness, interpersonal alienation), and reduced positive affect and engagement (loss of interest, meaning, and purpose).[51] Somatic symptoms of depression (change in appetite or sleeping patterns) are not included to prevent potential confounding when assessing patients with comorbid physical conditions.

Legacy instruments

The *SF-36* is the most widely used PRO in the world, has a 28-day recall, contains 36 Likert-scale items, and measures 8 physical and psychosocial health domains: Physical Functioning, Social Role Functioning, Energy and Vitality, Bodily Pain, Mental Health, Physical Role Functioning, Emotional Role Functioning, and General Health Perceptions. All subscale scores are converted to a standardized 0–100 score. Higher scores indicate a greater amount of the target domain.[76,78] The *WOMAC* subscales measure pain (5 items) and function (17 items), have a 2–day recall period, and uses a 100mm visual analog scale per item. Higher scores respectively indicate greater pain or worse physical functioning[50]. *Beck Depression, 2nd edition,* contains twenty-one 4-point Likert-scale items, total score of 0–63, and is extensively used to measure the psychological and somatic manifestations of 2-week major depressive episodes, as operationalized in the Data Statistical Manual-IV. Higher scores indicating greater level of depressive symptoms.[75] The *Perceived Stress Scale* has a 30-day recall, ten 5-point Likert-scale items, measures an individual's perception of psychological stress, and a total score of 0–40 with higher scores indicating greater level of stress.[36] The *Patient Global Assessment* consists of a single, 0–10cm visual analog scale item, with higher numbers indicating greater perception of disease activity.[34] The *Six-*

minute and *20-Meter Walk* are performance tests that respectively assess functional capacity (measured in meters) and gait speed (measured in seconds).[19,45]

Responsiveness analyses

Both the COSMIN and PROMIS Instrument Validation Standards recommend assessing responsiveness according to *a priori* hypotheses concerning the expected relationships between changes on PROMIS and changes on legacies; or expected differences between groups of change correlations (Supplementary Figure 1).[41,43,54,69] Using this method, the evaluation of responsiveness is less dependent on the magnitude of change and more dependent on the change of the intended construct.[41,43,54,69] After comparing each instrument, we came to consensus on 6 independently formulated hypotheses for each PROMIS Short Form (Table 1). Each pre-specified hypothesis was created by estimating degree of relatedness after an item-by-item analysis of each instrument and its relevant properties, including specific target construct, recall period, scale type, scale range, total score range, and total number of items (Supplementary Table 1). In each hypothesis, we clearly defined the anticipated (Spearman) correlation direction and magnitude (or difference in group magnitudes), according to Cohen's estimate of correlation strength,[12] and a rationale for each hypothesis.[41] We determined an overall responsiveness rating for each PROMIS instrument based on its total number of confirmed hypotheses: High, 5 or 6 (≥75%); Moderate, 3 or 4 (50%≤ x<75%); or Poor, 0 to 2 (<50%).[41] In addition to descriptive information and correlation coefficients, we used t-tests and chi-square tests to check for demographic (i.e. age, body mass index, pain duration, radiographic severity, sex, race, and education) or clinical (i.e. WOMAC pain and function, Patient Global Assessment, SF-Mental Health, Beck Depression, and Perceived Stress) differences between the participants from the parent trial and the participant group included in this analysis. All analyses were calculated using SAS statistical software (Version 9.4). In accordance with COSMIN guidelines, statistical significance was not used to formulate or interpret the pre-defined hypotheses of anticipated correlations because low correlations can become statistically significant in large populations and because it is not relevant whether the correlation deviates from zero, but whether there is some predefined magnitude of correlation[2,41,42,68,69].

In psychometric methodology, assessing responsiveness according to *a priori* hypotheses is a continuous process of accumulating evidence. In general, the more specific the hypotheses or the more hypotheses tested, the better the evidence.[68,69] However, concrete methodological guidelines do not yet exist for: 1) cut-off scores of correlation magnitude; 2) total number of hypotheses to include; or 3) adjusting for floor/ceiling effects. Floor/ceiling effects were considered significant if more than 15% of respondents achieved the lowest/highest possible score.[43] Accounting for floor/ceiling effects are important because, inherently, they can interfere with instrument responsiveness. Some define the ceiling of a measure as the score that indicates the best possible state or performance, and the floor as the score that indicates the worst possible[81]. Therefore, the lowest value on PROMIS Physical Function represents the floor but the lowest value on Pain Interference, Anxiety and Depression represents the ceiling.

Accordingly, we "stress-tested" the methodology by performing 3 sensitivity analyses: 1) thresholds of correlation strength was increased by 0.1; 2) total number of hypotheses per instrument was changed from 6 to 8; and 3) participants with floor/ceiling effects were excluded, and change correlations were re-calculated. Each sensitivity analysis was employed to isolate the effects of each methodological aspect (Tables 2A&B and Supplementary Tables 2A&B; Note: for brevity, the hypotheses for PROMIS Depression and Anxiety were moved to the Supplement).

Minimally important differences

It is well documented that the MID is best estimated using multiple anchors that are selected based on valid criteria of relevance for the disease indication, clinical acceptance, and evidence that the anchors have some relationship with the PRO measure[55]. This approach is especially useful so long as a portion of participants experience important, but minimal, change[55]. Accordingly, we followed highlighted methodology from the PROMIS Instrument Validation Standards wherein longitudinal anchor variables (e.g. SF-36 subscales) are used as an external reference to classify participants (who have completed both anchor instruments and PROMIS instruments) into a distinct group that experiences important, but *minimal,* improvement (Supplementary Figure 2).[54,79] It is conceptually necessary to identify the smallest difference or change that is important to the patient when estimating an MID because the change in PROMIS scores reported by this subgroup is used as the PROMIS MID estimate[55]. In this way, the final MID estimates will reflect minimally important differences rather than the type of intervention used. A more comprehensive description of this methodology is found elsewhere.[54,79] In brief, legacy anchors were paired with each PROMIS instrument based on similarity of their target constructs or clinical relevance[55]. Minimal important change in each multi-item anchor (e.g. SF-Physical Function) was defined based on an MID that was previously published for the legacy anchor (range=MID to 2x MID).[79] Minimal important change in a single-item anchor (i.e. Patient Global Assessment), was defined as: (range=MID to 1+MID).[79] Hence, by establishing an upper limit for what constitutes minimal important change, we exclude participants who experienced important change beyond the minimal amount[55]. When available, we used legacy anchor MIDs that were previously published from among participants with OA. However, legacy anchor MIDs estimated from among other patient groups or from using universal benchmarks (e.g. one half of the standard deviation)[47] are also acceptable[18,54,79,80]. We calculated the absolute mean change (and percent change) of the PROMIS T-scores from among participants classified with minimal important change based on a given legacy anchor to determine a single potential MID estimate.

Next, both anchor-based and distribution-based methods were used to evaluate the quality of each potential MID estimate, and only MIDs that met 3 *a priori* criteria were selected.[4,79] These criteria were designed to ensure that the: 1) Legacy anchor is associated with the PROMIS measure (Spearman correlation≥0.3 between change scores[55]); 2) Important change group has ≥10 participants (i.e. legacy change score greater than or equal to the legacy MID, but no greater than twice the legacy MID); and 3) Effect size (Cohen's *d*) of each MID falls within plausible range ($0.2 \leq x \leq 0.8$)[55,79]. The lowest- and highest-selected MIDs respectively formed the lower- and upper-bounds of an initial MID range for each

PROMIS instrument. If the absolute value of the lower bound estimate was smaller than the mean Standard Error of Measurement (SEM, i.e. smallest change score exceeding measurement error), then the lower bound would be set to the SEM.[66,79] SEM for PROMIS Short Forms correspond to the mean T-score at a given time point, and, rather than through the traditionally-used formula for SEM, its value can be found in the respective PROMIS scoring manual.[53,79] We averaged the SEM at baseline and follow-up to create a mean total SEM. By using anchor-based methods to estimate multiple potential MID estimates for each PROMIS instrument and distribution-based methods to evaluate the quality of each potential MID estimate, this approach triangulates a range of MIDs across multiple methods[18,55,79,80].

The COSMIN Standard recommends to report distribution-based MIDS as secondary estimates because they are solely based on statistical distribution and do not take into account information directly from the patient.[33,70] Accordingly, we determined a T-score range that corresponds to a widely-used range of 1/5 to ½ of our sample's Standard Deviation (SD) at baseline [Effect Size] and of our sample's SD of change, from baseline to follow-up, [Standardized Response Mean] as secondary estimates of MIDs[33]. (See Supplementary Text for further description of study procedure, PROs, and handling of missing items)

RESULTS

Among 165 predominantly female (70%) and white (57%) participants, mean age was 61 years, pain duration was 8.7 years, body mass index was 33kg/m^2, and most (92%) had a Kellgren/Lawrence grade≥2 (Supplementary Table 3). 81% of participants from the parent trial were included in this analysis. When comparing participants included in this analysis and participants from the parent trial, no statistically significant or clinically meaningful differences were found (data not shown).

Descriptive statistics and correlations for PROMIS

All PROMIS measures reflected improved health outcomes after intervention (Table 3). Change in physical outcomes was larger than change in psychological outcomes. Significant ceiling effects were detected in PROMIS Anxiety at baseline (17%) and 12 weeks (20%); and in Pain Interference at 12 weeks (22%). Compared to the general population, our participants reported having worse baseline physical health (PROMIS Pain Interference scores and Physical Function T-scores at 78th and 18th percentiles respectively), but similar psychological health (PROMIS Depression and Anxiety T-scores at 48th and 50th percentiles respectively). All change associations were in the anticipated direction; that is, PROMIS change scores indicating better health were positively associated with legacy change scores indicating better health (Table 4). For each PROMIS Short Form, no more than 3 individuals had missing items. Among these individuals, the total number of missing items was 2 or less at either baseline or 12 weeks (Supplementary Text).

Responsiveness

We confirmed 5 of 6 *a priori* hypotheses (83%) for PROMIS Physical Function. For Pain Interference, Depression, and Anxiety, respectively 4 of 6 (67%), 3 of 6 (50%) and 4 of 6

(67%) hypotheses were confirmed (Tables 2A&B and Supplementary Tables 2A&B). Therefore, PROMIS Physical Function has high responsiveness, and Pain Interference, Depression, and Anxiety have moderate responsiveness.

Increasing the correlation cut-points (Sensitivity Analysis 1) resulted in a lower overall responsiveness rating for each PROMIS measure (Table 5). Using 8 total hypotheses instead of 6 (Sensitivity Analysis 2) did not result in different responsiveness ratings. Excluding people with floor/ceiling effects (Sensitivity Analysis 3) did not change the responsiveness rating for PROMIS Pain Interference, but Anxiety changed from Moderate to High.

MIDs

Of 14 estimated MIDs, 8 met all criteria (Supplementary Table 5). Four MIDs did not meet criterion #3 (MID effect size); 3 MIDs did not meet criterion #1 (anchor correlation); and no MIDs failed to meet criterion #2 (sample size; Table 6). The lower bounds of the respective MID range for PROMIS Anxiety (1.6) and Physical Function (1.8) did not exceed the SEM (2.3 and 1.9 respectively), and were, therefore, replaced with their respective SEMs to ensure that the final MID ranges exceeded the amount of change that cannot be distinguished from statistical noise. Final anchor-based MID ranges (% change) were: PROMIS Depression= 3.0 to 3.1 (6.0 to 6.2%); Anxiety= 2.3 to 3.4 (4.5 to 6.6%); Physical Function= 1.9 to 2.2 (4.5 to 5.3%); and Pain Interference= 2.35 to 2.4 (4.2%; Figure 1). These MIDs correspond to 26–39% of the normalized SD units of our study sample per scale. (See Supplementary Text for descriptive statistics for legacy instruments, distribution-based MID estimates, and detailed missing PROMIS items)

DISCUSSION

This is the first study to evaluate responsiveness and establish a range of MIDs for PROMIS Short Forms among adults with knee OA. We specifically found that PROMIS Physical Function has high responsiveness, while Pain Interference, Depression, and Anxiety have moderate responsiveness. Our results support that PROMIS Physical Function, Pain Interference, Depression, and Anxiety Short Forms are able to detect change over time and convey the minimal amount of change that can be considered clinically meaningful.

Our study provides important information to support the utility of PROMIS Short Forms in both research and clinical care of OA. For the research setting, our findings imply that PROMIS Short Forms, which complement recommended outcome domains for clinical trials in pain and knee OA,[38,63] can detect response from novel treatments; and our MIDs can guide power analyses for novel clinical studies. While response criteria for use in OA trials have been established by Outcome Measures in Rheumatology/Osteoarthritis Research Society International (OMERACT/OARSI)[49], these thresholds predominantly involve composite improvements in physical domains: pain and physical function. Our MID results for Depression and Anxiety contribute thresholds of minimally important response for psychological health domains to the OA literature. However, because our MID estimates are intended, we clarify that this threshold may differ from the OMERACT/OARSI response criteria to indicate *minimally* important change rather than change beyond the minimally important amount (i.e. ≥50% improvement in pain or function)[49]. For the clinical setting,

our findings imply that PROMIS Short Forms could be used to monitor effects of treatment policy at the group level over time. For example, by indicating a minimally important improvement, the PROMIS Pain Interference MID can provide an early indication that referral to a new exercise program helped reduce the interference of knee OA pain on activity participation.[25] Indeed, PROMIS is a landmark achievement that can revolutionize how PROs are used in OA.[10] While instrument validation is a continual process of accumulating evidence, our results support the consideration of these PROMIS Short Forms in both research and clinical settings.

As a universal metric, PROMIS is uniquely suited to facilitate comparison of MID estimates across various chronic conditions[13]. Our PROMIS MID findings are consistent with those of Yost and colleagues, whose estimated longitudinal MIDs ranged from 2.4 to 3.5 among cancer patients[79]. Notably, Yost and colleagues observed that their longitudinal MID estimates were smaller in magnitude than their cross-sectional estimates, which ranged from 4 to 5.7. They theorized that this discordance may be due to the methodological approach used for cross-sectional MIDS, which relies on established cross-sectional cut-points of legacy measures to categorize patients into clinically distinct groups. In fact, it may be that the clinical difference between subgroups categorized by cross-sectional cut-points is larger than the minimal amount[79]. This may also explain the slightly larger MID estimates (3.5 to 5.5) recently found among patients with low back pain and depression[4]. Our findings are also concordant with those from our cross-sectional study, which detected ceiling effects in PROMIS Anxiety and supported the cross-sectional construct validity of each PROMIS instrument.[16] Importantly, our combined results establish the principal evidence of overall construct validity (i.e. construct validity at both a single point and over time) for these PROMIS Short Forms in an OA clinical sample. Overall construct validity is arguably the most crucial measurement property of PROs because it requires confirmatory testing among specific disease groups. Therefore, our results fulfill a crucial step of the instrument validation process for this patient population,[21,55,68] and can serve as a timely point of reference for researchers and clinicians developing PROMIS cross-cultural validation studies throughout Europe.[1] A noticeable finding in our results is that the physical performance-based measures were not correlated with PROMIS Physical Function or Pain Interference, which is discordant with what we found at baseline[16]. This may be partially due to the relatively larger standard errors found among change scores compared to single scores. In addition, there is a substantial body of literature showing that patient-reported and physical performance-based measures do not always correlate[15,30,61,64]. Thus, the accumulating evidence supports that patient-reported and physical performance-based measures may provide complementary, rather than identical information[15,30,61,64]. A more comprehensive discussion of this topic, including the potential role of common method variance, can be found elsewhere[39,59–61,64]. Further study may be needed to further disentangle any discordance between PROMIS measures and physical performance-based measures. Nevertheless, the responsiveness of these PROMIS instruments are supported by the strength and number of confirmed hypotheses concerning their anticipated correlations with legacy measures.

Our study's strengths have implications for the interpretation and application of PROMIS Short Forms. For responsiveness, our findings advance the conversation on refining the

methodological approach; and for MIDs, our ranges communicate nuanced information to assist the interpretation of PRO change scores, and our MID analyses highlight methodological elements that can confound the quality of a given MID estimate. We used sensitivity analyses to "stress-test" 3 different aspects of the responsiveness methodological model. *First*, regarding correlation thresholds, our findings confirm that higher thresholds can reduce responsiveness of PROMIS Short Forms. Despite the wide application of Cohen's estimated thresholds, opposing views exist regarding interpretation.[17,52] Moreover, because more lenient cut-points would likely yield higher responsiveness ratings of PRO instruments, additional examination using sensitivity analyses with more lenient cut-points are warranted. Our findings underscore the need for consensus guidelines on threshold-levels and can contribute toward their data-driven development. In addition, further refinement of hypotheses testing methodology may be needed, including whether to interpret correlation results using exact differences or standard errors. *Second*, floor/ceiling effects can also reduce responsiveness. As PROMIS Short Forms were developed from large item banks, they are less likely to detect floor/ceiling effects than other PRO instruments.[22] Additionally, different versions of PROMIS Short Forms can target specific areas of the domain spectrum.[22] Other precautions in study design, including requiring a minimal symptom severity for study inclusion or excluding those from extreme ends of the spectrum, can mitigate the influence of floor/ceiling effects. *Third*, after increasing the number of hypotheses, we found no difference in responsiveness. Therefore, the relevance of the assumption, in which adding hypotheses generally improves precision when testing responsiveness,[44,69] may depend on the number or quality of hypotheses added. From a practical perspective, advancing the methodology for responsiveness is important because PROs can often best quantify the life-changing effects of novel treatments. Our simplified methodological paradigm for responsiveness contributes a practical point of reference for further study (Supplementary Figure 1).

Our approach to estimating MIDs has notable strengths. *First*, by reporting a range of MIDs, we provide more robust interpretive information compared to a single fixed-value. PROMIS change scores *above* a MID range (i.e. 2 or more single point-estimates) inherently indicate meaningful change with greater certainty than a change score above a single point. Accordingly, a PROMIS change score falling *within* an MID range indicates meaningful change with lesser certainty than one falling *above* the range. Traditionally, the MID was operationalized as a single fixed value.[33] However, recent evidence shows that MID can be influenced by several factors, including clinical population or variability of baseline values.[55] Therefore, these findings underscore the fundamental fluidity of MIDs, and can promote a more accurate interpretation of PRO change scores. However, it is important to note that the ranges presented for PROMIS Depression and Pain Interference are narrow. Therefore, additional research is needed to further elucidate the range of MID for these instruments. *Second,* each of our final MID estimates met 3 *a priori* quality criteria. These criteria accentuate methodological elements, including anchor correlation, sample size, or effect size, from both anchor-based and distribution-based approaches that can confound MID quality. In other words, our final MID estimates: 1) were derived from legacy anchors that were at least moderately (rho: ≥0.3) associated with the respective PROMIS instrument[14,18]; 2) were derived from a participant size classified as having 'minimally

important change' that was large enough (n≥10) to estimate PROMIS MIDs with adequate stability[79]; and 3) represent an effect size of change that was both large enough to be considered important (≥0.2) and small enough to be considered minimal (0.8≥). Although a consensus on the ideal thresholds for these criteria is not yet established, this methodology uniquely employs both anchor and distribution-based approaches to triangulate a range of MID estimates, which is a widely recommended approach[9,18,27,28,33,55,79,80]. In addition, this approach promotes a better understanding of the underlying quality of MIDs and can inform the decision-making process of clinicians and researchers on how to ideally apply them (Supplementary Figure 2). Importantly, we emphasize that rather than a static property of an instrument, the MID is best conceptualized as a value that may vary by context and setting.

This study has limitations. Although PROMIS instruments were administered for the purpose of this study, the parent clinical trial had a distinct primary purpose, and therefore has effects which limit this study. First, the generalizability of our findings may not include participants with severe psychological distress, recent surgery, or without active pain. Second, the types of legacy instruments available for use were limited to the instruments from the parent trial. Accordingly, PROMIS Anxiety was not compared with an anxiety-specific legacy measure, and the most commonly-used anchor for MIDs, the Global Rating of Change, was not applied. Third, this study lacked a control group receiving no treatment. Thus, confirmation of discriminant validity of these PROMIS instruments is still required. Despite these limitations, our results generalize to a large portion of patients with symptomatic knee OA that are likely to be encountered in a clinical setting or recruited for clinical trials. Furthermore, anxiety-related legacy measures, including mental health and perceived stress, were appropriately utilized in the responsiveness and MID analyses according to accepted methodological standards.[44,54,55] Although commonly used, Global Rating of Change instruments have well-documented limitations with recall bias, and recently were ineffective at establishing high-quality MIDs in PROMIS.[26,33,79] In fact, prospective, longitudinal approaches have been recommended as a superior alternative[37,79]. The criterion developed by Yost and colleagues requiring ≥10 participants with minimally important change may require further refinement because the optimal number required may be larger[79]. Yet in our study, all but one MIDs estimated in our study were based on sub-sample sizes no smaller than 20 participants, which is twice the required amount. Importantly, this limitation does not detract from the overall value of the methodological approach, which is one of the few existing approaches that triangulate both anchor-based and distribution-based principles to identify higher quality MID estimates. Finally, some legacy anchor MIDs that were used to determine our MID estimates for PROMIS Depression and Anxiety were not derived from participants with OA (Supplementary Table 5). Because instrument validation studies in OA more commonly involve measures of physical health rather than psychosocial health, this was not unexpected. While inherently optimal to utilize legacy MIDs estimated from among OA participants, the methodological use of 3 distinct criteria to appraise the quality of each potential MID estimate supports the utility and applicability of our final range of MIDs among groups with OA.

Future studies should confirm responsiveness of these PROMIS Short Forms among groups that are expected to change versus those that are not. In addition, our MID estimates should

be compared with those that are derived from legacy anchor MIDs that were reported from among participants with OA. Because MID thresholds for symptom improvement may differ from those for deterioration, MIDs for symptom deterioration should be examined in future studies. Furthermore, the various cross-cultural validation studies developing throughout Europe may confirm the versatility of PROMIS instruments among different languages and cultures.[1,31]

In conclusion, we found that PROMIS Depression, Anxiety, Physical Function, and Pain Interference are sensitive to change in adults with knee OA, and established a range of MIDs for each instrument. This standard of reference can assist investigators and clinicians in the application or interpretation of these novel instruments in the future. PROMIS Short Forms are less daunting options for older individuals and easily deployed in community-based settings.[48] Therefore, our results provide important evidence for healthcare systems considering PROMIS as a standard tool for patient-centered care,[25] and for scientific groups considering PROMIS for hierarchical lists of recommended instruments.[5,11,32]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: This study is supported by the National Institutes of Health (R01AT005521, K24AT007323, and UL1TR001064). The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. The investigators are solely responsible for the content of the manuscript and the decision to submit for publication. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

PROMIS® was funded with cooperative agreements from the National Institutes of Health Common Fund Initiative (U54AR057951, U01AR052177, U54AR057943, U54AR057926, U01AR057948, U01AR052170, U01AR057954, U01AR052171, U01AR052181, U01AR057956, U01AR052158, U01AR057929, U01AR057936, U01AR052155, U01AR057971, U01AR057940, U01AR057967, U01AR052186). 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.

The Authors gratefully appreciate Dr. Mei Chung for suggestions on the presentation of data, and Sun Yu and Stephanie Hyon for feedback on the written manuscript. We especially thank Dr. Kathleen Yost for her invaluable knowledge and expert feedback. Finally, we thank our study staff who tirelessly collected the data; and the study participants who took the time to complete each outcome measure.

References

1. Alonso J, Bartlett SJ, Rose M, Aaronson NK, Chaplin JE, Efficace F, Leplège A, Lu A, Tulsky DS, Raat H, Ravens-Sieberer U, Revicki D, Terwee CB, Valderas JM, Cella D, Forrest CB. The case for an international patient-reported outcomes measurement information system (PROMIS®) initiative. Health Qual Life Outcomes [Internet]. 2013; 11:210. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3879205&tool=pmcentrez&rendertype=abstract.
2. Altman, DG. Practical statistics for medical research. 1. London; New York: Chapman and Hall; 1991.
3. Amtmann D, Cook KF, Jensen MP, Chen W-H, Choi S, Revicki D, Cella D, Rothrock N, Keefe F, Callahan L, Lai J-S. Development of a PROMIS item bank to measure pain interference. Pain [Internet]. 2010; 150:173–82. [cited 2016 May 4]. Available from: http://

www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2916053&tool=pmcentrez&rendertype=abstract.

4. Amtmann D, Kim J, Chung H, Askew RL, Park R, Cook KF. Minimally important differences for Patient Reported Outcomes Measurement Information System pain interference for individuals with back pain. J Pain Res [Internet]. 2016; 9:251–5. [cited 2016 Jun 3]. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4854267&tool=pmcentrez&rendertype=abstract.
5. Beaton DE, Terwee CB, Singh JA, Hawker GA, Patrick DL, Burke LB, Toupin-April K, Tugwell PS. A Call for Evidence-based Decision Making When Selecting Outcome Measurement Instruments for Summary of Findings Tables in Systematic Reviews: Results from an OMERACT Working Group. J Rheumatol [Internet]. 2015; 42:1954–61. Available from: http://www.jrheum.org/cgi/doi/10.3899/jrheum.141446.
6. Boers M, Kirwan JR, Gossec L, Conaghan PG, D'Agostino M-A, Bingham CO, Brooks PM, Landewé R, March L, Simon L, Singh JA, Strand V, Wells GA, Tugwell P. How to choose core outcome measurement sets for clinical trials: OMERACT 11 approves filter 2.0. J Rheumatol [Internet]. 2014; 41:1025–30. [cited 2016 May 16]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24584913.
7. 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 (Hoboken) [Internet]. 2013; 65:1625–33. [cited 2016 Feb 24]. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3779528&tool=pmcentrez&rendertype=abstract.
8. Castrejón I, Gossec L, Carmona L. The EULAR Outcome Measures Library: an evolutional database of validated patient-reported instruments. Ann Rheum Dis [Internet]. 2015; 74:475–6. [cited 2016 May 16]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25452307.
9. Cella D, Eton DT, Lai J-S, Peterman AH, Merkel DE. Combining Anchor and Distribution-Based Methods to Derive Minimal Clinically Important Differences on the Functional Assessment of Cancer Therapy (FACT) Anemia and Fatigue Scales. J Pain Symptom Manage. 2002; 24:547–61. [PubMed: 12551804]
10. Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., Devellis, R., DeWalt, D., Fries, JF., Gershon, R., Hahn, EA., Lai, J-SS., Pilkonis, P., Revicki, D., Rose, M., Weinfurt, K., Hays, R. J Clin Epidemiol [Internet]. Vol. 63. Elsevier Inc; 2010. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008; p. 1179-94.[cited 2015 Nov 9]. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2965562&tool=pmcentrez&rendertype=abstract
11. Christensen R, Maxwell LJ, Jüni P, Tovey D, Williamson PR, Boers M, Goel N, Buchbinder R, March L, Terwee CB, Singh JA, Tugwell P. Consensus on the need for a hierarchical list of patient-reported pain outcomes for metaanalyses of knee osteoarthritis trials: An OMERACT objective. J Rheumatol [Internet]. 2015; 42:1971–5. [cited 2016 May 16]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25934823%5Cnhttp://www.jrheum.org/content/early/2015/04/27/jrheum.141384.full.pdf.
12. Cohen J. Quantitative Methods in Psychology. Psychol Bull [Internet]. 1992; 112:155–9. Available from: http://psycnet.apa.org/doi/10.1037/0033-2909.112.1.155.
13. Cook, KF., Jensen, SE., Schalet, BD., Beaumont, JL., Amtmann, D., Czajkowski, S., Dewalt, DA., Fries, JF., Pilkonis, PA., Reeve, BB., Stone, AA., Weinfurt, K., Cella, D. J Clin Epidemiol [Internet]. Elsevier Ltd; 2016. PROMIS® Measures of Pain, Fatigue, Negative Affect, Physical Function and Social Function Demonstrate Clinical Validity across a Range of Chronic Conditions. [cited 2016 Mar 12]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26952842
14. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol. 2003; 56:395–407. [PubMed: 12812812]
15. Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, Buchbinder R, Snyder-Mackler L, Henrotin Y, Thumboo J, Hansen P, Bennell KL. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthr Cartil. 2013; 21:1042–52. [PubMed: 23680877]

16. Driban JB, Morgan N, Price LL, Cook KF, Wang C. Patient-Reported Outcomes Measurement Information System (PROMIS) instruments among individuals with symptomatic knee osteoarthritis: a cross-sectional study of floor/ceiling effects and construct validity. BMC Musculoskelet Disord [Internet] BMC Musculoskeletal Disorders. 2015; 16:253. Available from: http://dx.doi.org/10.1186/s12891-015-0715-y. [PubMed: 26369412]
17. Ellis DP. Effect sizes and the interpretation of research results in international business. J Int Bus Stud [Internet]. 2010; 41:1581–8. Available from: http://dx.doi.org/10.1057/jibs.2010.39.
18. Eton DT, Cella D, Yost KJ, Yount SE, Peterman AH, Neuberg DS, Sledge GW, Wood WC. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. J Clin Epidemiol [Internet]. 2004; 57:898–910. [cited 2016 May 17]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15504633.
19. Focht BC, Rejeski WJ, Ambrosius WT, Katula JA, Messier SP. Exercise, self-efficacy, and mobility performance in overweight and obese older adults with knee osteoarthritis. Arthritis Care Res [Internet]. 2005; 53:659–65. [cited 2016 Feb 9]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16208674.
20. Fransen, M., Mcconnell, S., Harmer, AR., Van der Esch, M., Simic, M., Bennell, KL. Cochrane Database Syst Rev [Internet]. England: 2015. Exercise for osteoarthritis of the knee; p. CD004376[cited 2016 Jan 27]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25569281
21. Fries, JF., Krishnan, E., Rose, M., Lingala, B., Bruce, B. Arthritis Res Ther [Internet]. Vol. 13. BioMed Central Ltd; 2011. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory; p. R147Available from: http://arthritis-research.com/content/13/5/R147
22. Fries JF, Lingala B, Siemons L, Glas CAW, Cella D, Hussain YN, Bruce B, Krishnan E. Extending the floor and the ceiling for assessment of physical function. Arthritis Rheumatol (Hoboken, NJ) [Internet]. 2014; 66:1378–87. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4012831&tool=pmcentrez&rendertype=abstract.
23. Garratt AM, Brealey S, Gillespie WJ. Patient-assessed health instruments for the knee: a structured review. Rheumatology [Internet]. 2004; 43:1414–23. [cited 2016 Apr 27]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15316121.
24. Geneen, LJ., Moore, RA., Clarke, C., Martin, D., Colvin, LA., Smith, BH. Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. In: Geneen, LJ., editor. Cochrane Database Syst Rev [Internet]. Chichester, UK: John Wiley & Sons, Ltd; p. CD0112792017[cited 2017 Mar 3]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28087891
25. Golightly, YM., Allen, KD., Nyrop, KA., Nelson, AE., Callahan, LF., Jordan, JM. Semin Arthritis Rheum [Internet]. Vol. 45. Elsevier; 2015. Patient-reported outcomes to initiate a provider-patient dialog for the management of hip and knee osteoarthritis; p. 123-31.[cited 2016 Jul 2]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25979178%5Cnhttp://ac.els-cdn.com/S0049017215000700/1-s2.0-S0049017215000700-main.pdf?_tid=3341bd54-13a2-11e5-b5cb-00000aacb35d&acdnat=1434402487_7a5e6820ac7aa2b47f615c7513703eef
26. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. J Clin Epidemiol. 2002; 55:900–8. [PubMed: 12393078]
27. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to Explain the Clinical Significance of Health Status Measures. Mayo Clin Proc. 2002; 77:371–83. [PubMed: 11936935]
28. Haley SM, Fragala-Pinkham MA. Interpreting Change Scores of Tests and Measures Used in Physical Therapy. Phys Ther. 2006; 86
29. Hays RD, Spritzer KL, Fries JF, Krishnan E. Responsiveness and minimally important difference for the Patient-Reported Outcomes Measurement Information System (PROMIS) 20-item physical functioning short form in a prospective observational study of rheumatoid arthritis. Ann Rheum Dis [Internet]. 2013; 74:1–4. [cited 2016 Jan 29]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24095937.
30. Jordan KP, Wilkie R, Muller S, Myers H, Nicholls E. Arthritis Research Campaign National Primary Care Centre. Measurement of change in function and disability in osteoarthritis: current approaches and future challenges. Curr Opin Rheumatol [Internet]. 2009; 21:525–30. [cited 2017

Mar 3]. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00002281-200909000-00015.

31. Keller S, Correia H. The Patient-Reported Outcome Measurement Information System (PROMIS ®): International Update. Patient Rep Outcomes Newsl [Internet]. 2012; 47:9–10. Available from: http://www.pro-newsletter.com/images/PDF/pronl47_final.pdf.
32. Khanna D, Krishnan E, Dewitt EM, Khanna PP, Spiegel B, Hays RD. The future of measuring patient-reported outcomes in rheumatology: Patient-Reported Outcomes Measurement Information System (PROMIS). Arthritis Care Res (Hoboken) [Internet]. 2011; 63(Suppl 1):S486–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22588770.
33. King, MT. Expert Rev Pharmacoecon Outcomes Res [Internet]. Vol. 11. Taylor & Francis; 2011. A point of minimal important difference (MID): a critique of terminology and methods; p. 171-84. [cited 2016 Feb 22]. Available from: http://www.tandfonline.com.ezproxy.library.tufts.edu/doi/abs/10.1586/erp.11.9?journalCode=ierp20#.VstC2X0rKUk
34. Kirwan JR, Fries JF, Hewlett SE, Osborne RH, Newman S, Ciciriello S, van de Laar MA, Dures E, Minnock P, Heiberg T, Sanderson TC, Flurey CA, Leong AL, Montie P, Richards PAM. Patient perspective workshop: Moving towards OMERACT guidelines for choosing or developing instruments to measure patient-reported outcomes. J Rheumatol [Internet]. 2011; 38:1711–5. [cited 2016 Jan 29]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21807790.
35. Lauche, R., Langhorst, J., Dobos, G., Cramer, H. Complement Ther Med [Internet]. Vol. 21. Elsevier Ltd; 2013. A systematic review and meta-analysis of Tai Chi for osteoarthritis of the knee; p. 396-406.Available from: http://dx.doi.org/10.1016/j.ctim.2013.06.001
36. Lee E-H. Review of the psychometric evidence of the perceived stress scale. Asian Nurs Res (Korean Soc Nurs Sci) [Internet]. 2012; 6:121–7. Available from: http://www.sciencedirect.com/science/article/pii/S1976131712000527.
37. Liang MH, Lew RA, Stucki G, Fortin PR, Daltroy L. Measuring clinically important changes with patient-oriented questionnaires. Med Care [Internet]. 2002; 40:II45–51. [cited 2016 Dec 8]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12064581.
38. McAlindon, TE., Driban, JB., Henrotin, Y., Hunter, DJ., Jiang, G-L., Skou, ST., Wang, S., Schnitzer, T. Osteoarthr Cartil [Internet]. Vol. 23. Elsevier Ltd; 2015. OARSI Clinical Trials Recommendations: Design, conduct and reporting of clinical trials for knee osteoarthritis; p. 747-60.[cited 2016 May 16]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25952346
39. Meade, AW., Watson, AM., Kroustalis, CM. Assessing Common Methods Bias in Organizational Research. 22nd Annu Meet Soc Ind Organ Psychol; 2007. p. 1-10.
40. Messier SP, Callahan LF, Golightly YM, Keefe FJ. OARSI Clinical Trials Recommendations: Design and conduct of clinical trials of lifestyle diet and exercise interventions for osteoarthritis. Osteoarthr Cartil. 2015; 23:787–97. [PubMed: 25952349]
41. Mokkink, LB., Terwee, CB., Knol, DL., Stratford, PW., Alonso, J., Patrick, DL., Bouter, LM., de Vet, HC. BMC Med Res Methodol [Internet]. Vol. 10. England: 2010. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content; p. 22Available from: http://www.biomedcentral.com/1471-2288/10/22
42. Mokkink, LB., Terwee, CB., Patrick, DL., Alonso, J., Stratford, PW., Knol, DL., Bouter, LM., de Vet, HC. COSMIN Checklist Manual [Internet]. p. 562012www.cosmin.nl[cited 2016 Jun 5]. Available from: http://www.cosmin.nl/images/upload/files/COSMINchecklist manual v9.pdf
43. Mokkink, LB., Terwee, CB., Patrick, DL., Alonso, J., Stratford, PW., Knol, DL., Bouter, LM., de Vet, HCW. J Clin Epidemiol [Internet]. Vol. 63. Elsevier Inc; 2010. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes; p. 737-45.Available from: http://linkinghub.elsevier.com/retrieve/pii/S0895435610000909
44. Mokkink, LB., Terwee, CB., Patrick, DL., Alonso, J., Stratford, PW., Knol, DL., Bouter, LM., WHC. Qual Life Res [Internet]. Vol. 19. Netherlands: 2010. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study; p. 539-49.Available from: http://link.springer.com/10.1007/s11136-010-9606-8
45. Motyl JM, Driban JB, McAdams E, Price LL, McAlindon TE. Test-retest reliability and sensitivity of the 20-meter walk test among patients with knee osteoarthritis. BMC Musculoskelet Disord

[Internet] BMC Musculoskeletal Disorders. 2013; 14:166. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3661341&tool=pmcentrez&rendertype=abstract. [PubMed: 23663561]

46. Nahin RL, Boineau R, Khalsa PS, Stussman BJ, Weber WJ. Evidence-Based Evaluation of Complementary Health Approaches for Pain Management in the United States. Mayo Clin Proc [Internet]. 2016; 91:1292–306. [cited 2017 Mar 3]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27594189.
47. Norman GR, Sloan JA, Wyrwich KW. The truly remarkable universality of half a standard deviation: confirmation through another look. Expert Rev Pharmacoecon Outcomes Res England. 2004; 4:581–5.
48. Palmen LN, Schrier JCM, Scholten R, Jansen JHW, Koëter S. Is it too early to move to full electronic PROM data collection? Foot Ankle Surg [Internet]. 2016; 22:46–9. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1268773115000740.
49. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, Simon L, Strand V, Woodworth T, Dougados M. OMERACT-OARSI initiative: Osteoarthritis research society international set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthr Cartil. 2004; 12:389–99. [PubMed: 15094138]
50. Pham T, Van Der Heijde DVD, Lassere M, Altman RD, Anderson JJ, Bellamy N, Hochberg M, Simon L, Strand V, Woodworth T, Dougados M. Outcome variables for osteoarthritis clinical trials: The OMERACT-OARSI set of responder criteria. J Rheumatol. 2003; 30:1648–54. [PubMed: 12858473]
51. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. Assessment [Internet]. 2011; 18:263–83. [cited 2016 May 4]. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3153635&tool=pmcentrez&rendertype=abstract.
52. Post, MW. What to do with “moderate” reliability and validity coefficients?. Arch Phys Med Rehabil [Internet]. 2016. [cited 2016 Apr 18]. Available from: http://www.sciencedirect.com/science/article/pii/S0003999316300508
53. PROMIS Health Organization and PROMIS Cooperative Group. Health Measures Scoring [Internet]. 2016. [cited 2016 Jul 15]. Available from: http://www.healthmeasures.net/score-and-interpret/calculate-scores
54. PROMIS Health Organization PROMIS Cooperative Group. PROMIS ® Instrument Development and Validation Scientific Standards Version 2.0 (revised May 2013) [Internet]. 2013. [cited 2016 May 7]. Available from: www.nihpromis.org/Documents/PROMISStandards_Vers2.0_Final.pdf
55. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol [Internet]. 2008; 61:102–9. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0895435607001199.
56. Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: A systematic review. Arthritis Care Res. 2008; 59:876–83.
57. Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). J Clin Epidemiol [Internet]. 2008; 61:17–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18083459.
58. Siemons L, Krishnan E. A short tutorial on item response theory in rheumatology. Clin Exp Rheumatol. 2014; 32:581–6. [PubMed: 25065775]
59. Singh-Manoux A, Marmot MG, Adler NE. Does subjective social status predict health and change in health status better than objective status? Psychosom Med. 2005; 67:855–61. [PubMed: 16314589]
60. Spector PE. Method Variance in Organizational Research. Organ Res Methods [Internet]. 2006; 9:221–32. Available from: http://resolver.scholarsportal.info/resolve/10944281/v09i0002/221_mvior%5Cnhttp://orm.sagepub.com/content/9/2/221.abstract.

61. Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. J Clin Epidemiol. 2006; 59:160–7. [PubMed: 16426951]
62. Stubbs B, Vancampfort D, Rosenbaum S, Firth J, Cosco T, Veronese N, Salum GA, Schuch FB. An examination of the anxiolytic effects of exercise for people with anxiety and stress-related disorders: A meta-analysis. Psychiatry Res [Internet]. 2017; 249:102–8. [cited 2017 Mar 3]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28088704.
63. Taylor AM, Phillips K, Patel KV, Turk DC, Dworkin RH, Beaton D, Clauw DJ, Gignac MAM, Markman JD, Williams DA, Bujanover S, Burke LB, Carr DB, Choy EH, Conaghan PG, Cowan P, Farrar JT, Freeman R, Gewandter J, Gilron I, Goli V, Gover TD, Haddox JD, Kerns RD, Kopecky EA, Lee DA, Malamut R, Mease P, Rappaport BA, Simon LS, Singh JA, Smith SM, Strand V, Tugwell P, Vanhove GF, Veasley C, Walco GA, Wasan AD, Witter J. Assessment of physical function and participation in chronic pain clinical trials: IMMPACT/OMERACT recommendations. Pain [Internet]. 2016; 157:1836– 50. [cited 2016 Sep 6]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27058676.
64. Terwee, CB., Mokkink, LB., Steultjens, MPM., Dekker, J. Rheumatology [Internet]. Vol. 45. England: 2006. Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: a systematic review of measurement properties; p. 890-902.Available from: http://www.rheumatology.oxfordjournals.org/cgi/doi/10.1093/rheumatology/kei267
65. Turk DC, Dworkin RH, Revicki D, 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 [Internet]. 2008; 137:276–85. [cited 2016 Oct 27]. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00006396-200807150-00007.
66. Turner D, Schünemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, Guyatt GH. The minimal detectable change cannot reliably replace the minimal important difference. J Clin Epidemiol [Internet]. 2010; 63:28–36. [cited 2016 May 4]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19800198.
67. van Tuyl LHD, Boers M. Patient-reported outcomes in core domain sets for rheumatic diseases. Nat Rev Rheumatol [Internet]. 2015; 11:705–12. [cited 2016 May 27]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26324860.
68. de Vet, HCW., Terwee, CB., Mokkink, LB., Knol, DL. Meas Med A Pract Guid. 1. Cambridge: Cambridge University Press; 2011. Chapter 6: Validity; p. 150-201.
69. de Vet, HCW., Terwee, CB., Mokkink, LB., Knol, DL. Meas Med A Pract Guid. 1. Cambridge: Cambridge University Press; 2011. Chapter 7: Responsiveness; p. 202-26.
70. de Vet, HCW., Terwee, CB., Mokkink, LB., Knol, DL. Meas Med A Pract Guid. 1. Cambridge: Cambridge University Press; 2011. Chapter 8: Interpretability; p. 227-74.
71. Wang, C., Bannuru, R., Ramel, J., Kupelnick, B., Scott, T., Schmid, CH. BMC Complement Altern Med [Internet]. Vol. 10. England: 2010. Tai Chi on psychological well-being: systematic review and meta-analysis; p. 23[cited 2016 Aug 25]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20492638
72. Wang, C., Iversen, MD., McAlindon, T., Harvey, WF., Wong, JB., Fielding, RA., Driban, JB., Price, LL., Rones, R., Gamache, T., Schmid, CH. BMC Complement Altern Med [Internet]. Vol. 14. England: 2014. Assessing the comparative effectiveness of Tai Chi versus physical therapy for knee osteoarthritis: design and rationale for a randomized trial; p. 333[cited 2016 May 23]. Available from: http://www.biomedcentral.com/1472-6882/14/333
73. Wang, C., Schmid, CH., Iversen, MD., Harvey, WF., Fielding, RA., Driban, JB., Price, LL., Wong, JB., Reid, KF., Rones, R., McAlindon, T. Comparative Effectiveness of Tai Chi Versus Physical Therapy for Knee Osteoarthritis: A Randomized Trial. Ann Intern Med [Internet]. 2016. [cited 2016 May 18]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27183035
74. Wang F, Lee E-KO, Wu T, Benson H, Fricchione G, Wang W, Yeung AS. The Effects of Tai Chi on Depression, Anxiety, and Psychological Well-Being: A Systematic Review and Meta-Analysis. Int J Behav Med [Internet]. 2014; 21:605–17. Available from: http://link.springer.com/10.1007/s12529-013-9351-9.

75. Wang YP, Gorenstein C. Psychometric properties of the Beck Depression Inventory-II: A comprehensive review. Rev Bras Psiquiatr. 2013; 35:416–31. [PubMed: 24402217]

76. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Med Care Health Institute, Boston, MA 02111. 1992; 30:473–83.

77. White DK, Master H. Patient-Reported Measures of Physical Function in Knee Osteoarthritis. Rheum Dis Clin North Am [Internet]. 2016; 42:239–52. [cited 2016 May 3]. Available from: http://www.sciencedirect.com/science/article/pii/S0889857X16000065.

78. Wyrwich KW, Tierney WM, Babu AN, Kroenke K, Wolinsky FD. A comparison of clinically important differences in health-related quality of life for patients with chronic lung disease, asthma, or heart disease. Health Serv Res. 2005; 40:577–91. [PubMed: 15762908]

79. Yost, KJ., Eton, DT., Garcia, SF., Cella, D. J Clin Epidemiol [Internet]. Vol. 64. Elsevier Inc; 2011. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients; p. 507-16.Available from: http://dx.doi.org/10.1016/j.jclinepi.2010.11.018

80. Yost, KJ., Sorensen, MV., Hahn, EA., Glendenning, GA., Gnanasakthy, A., Cella, D. Value Heal [Internet]. Vol. 8. International Society for Pharmacoeconomics and Outcomes Research (ISPOR); 2005. Using multiple anchor- and distribution-based estimates to evaluate clinically meaningful change on the Functional Assessment of Cancer Therapy-Biologic Response Modifiers (FACT-BRM) instrument; p. 117-27.[cited 2016 May 17]. Available from: http://dx.doi.org/10.1111/j.1524-4733.2005.08202.x

81. How Rehabilitation Can Help People Living with HIV in Sub-Saharan Africa. An Evidence-Informed Tool for Rehab Providers [Internet]. 2015. [cited 2017 Mar 8]. Available from: http://ssa.hivandrehab.ca/pdf/emodule-ssa-all-sections.pdf

Perspective

This study examined whether PROMIS short form instruments (Physical Function, Pain Interference, Depression, and Anxiety) were able to detect change over time among adults with knee osteoarthritis, and provided minimally important change estimates for each measure. This standard of reference can help apply or interpret these instruments in the future.

Highlights

- Four PROMIS short forms were responsive to treatment in knee osteoarthritis.
- Physical Function had high responsiveness.
- Pain Interference, Depression, and Anxiety had moderate responsiveness.
- Minimally important difference estimates ranged from 1.9 to 3.4.
- This standard of reference can help apply or interpret these instruments in the future.

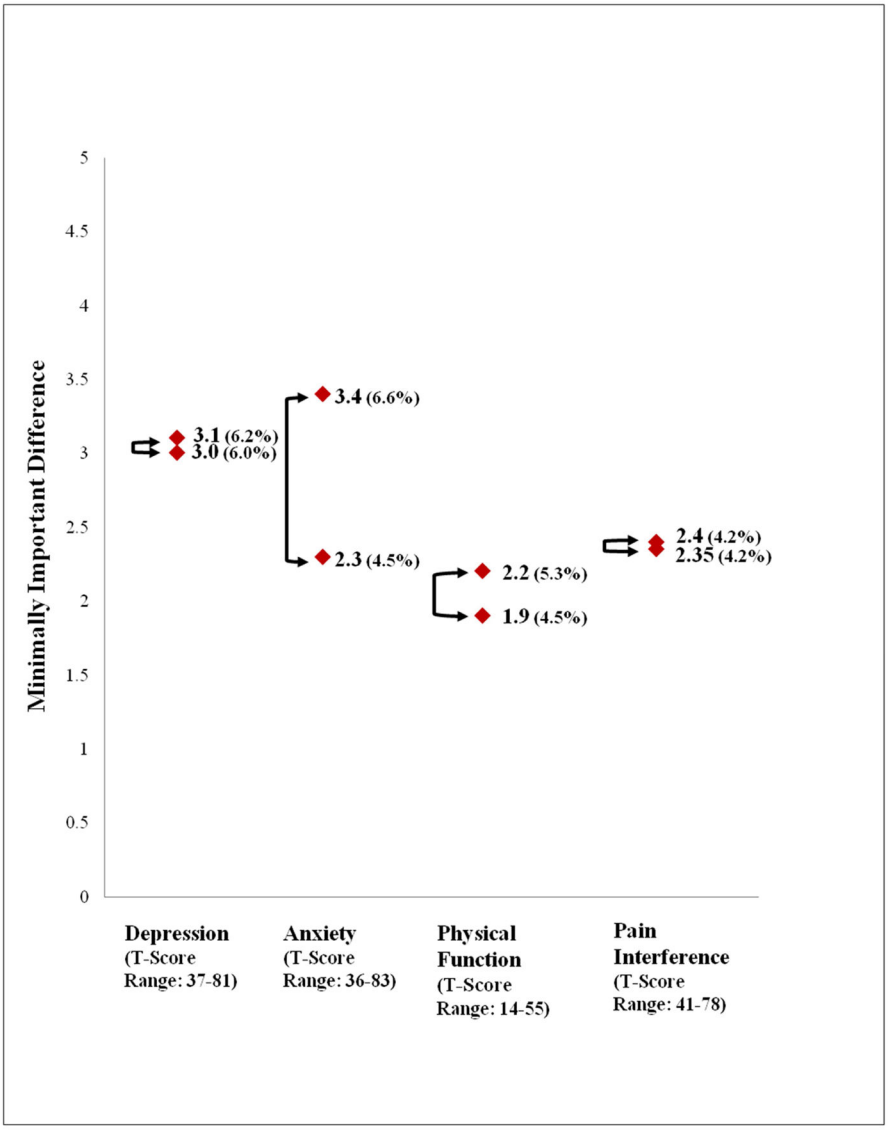

Figure 1. Range of Minimally Important Differences for PROMIS Short Forms in Knee Osteoarthritis

All Minimally Important Difference (MID) estimates was calculated using prospective, longitudinal anchors. Anchor-based MIDs were only included if all 3 *a priori* quality criteria were met.

Table 1

PROMIS Short Forms with Corresponding Legacy Measures Used for Hypotheses Testing of Responsiveness

PROMIS Short Forms			
Depression	**Anxiety**	**Physical Function**	**Pain Interference**
Beck Depression Inventory	Beck Depression Inventory	SF-Physical Function	SF-Bodily Pain
SF-Mental Health	Perceived Stress Scale	WOMAC Function	WOMAC Pain
Perceived Stress Scale	SF-Mental Health	WOMAC Pain	SF-Physical Function
SF-Role Emotional	SF-Role Emotional	SF-Role Physical	WOMAC Function
SF-General Health *	SF-Physical Function *	6 Minute Walk	SF-Mental Health *
WOMAC Pain *	6 Minute Walk *	SF-Mental Health *	20 Meter Walk *

SF=Short Form-36; WOMAC= Western Ontario and McMasters Osteoarthritis Index.

* Used as discriminant variable.

Table 2A

Hypotheses for the Responsiveness of PROMIS Pain Interference 6b

Hypotheses	Rationale	Correlation Result(s)		Confirmed?	Hypothesis Type
1. There will be a strong (0.5≤) negative correlation between change of PROMIS Pain Interference and change of SF-Bodily Pain after intervention.	PROMIS Pain Interference and SF-Bodily Pain attempt to measure similar constructs	Expect 0.50≤	Actual 0.53	Yes	Without Sensitivity Analysis *
		Expect 0.50≤	***Actual 0.49***	***No***	***Sensitivity Analysis 3*** ¶
2. The negative correlation of change on the PROMIS Pain Interference with change on the SF- Bodily Pain will be at least 0.1 lower than the negative correlation of change on the PROMIS Pain Interference with the SF-Mental Health	PROMIS Pain Interference and SF-Bodily Pain attempt to measure similar constructs, but PROMIS Pain Interference and SF-Mental Health attempt to measure loosely related constructs	Expect 0.10≤	Actual 0.16	Yes	Without Sensitivity Analysis
		Expect 0.10≤	***Actual 0.30***	***Yes***	***Sensitivity Analysis 3***
3. There will be at least a moderate- strong (0.4≤) positive correlation between change of PROMIS Pain Interference and change of WOMAC Pain after intervention.	PROMIS Pain Interference and WOMAC Pain attempt to measure similar constructs, but WOMAC Pain is lower limb- activity specific and pain intensity rather than interference	Expect 0.40≤	Actual 0.37	No	Without Sensitivity Analysis
		Expect 0.40≤	***Actual 0.37***	***No***	***Sensitivity Analysis 3***
4. There will be at least a moderate (0.3≤) positive correlation between change of PROMIS Pain Interference and change of WOMAC Function after intervention.	PROMIS Pain Interference and WOMAC Function attempt to measure partially related constructs	Expect 0.30≤	Actual 0.43	Yes	Without Sensitivity Analysis
		Expect 0.30≤	***Actual 0.47***	***Yes***	***Sensitivity Analysis 3***
5. There will be at least a weak- moderate (0.2≤) negative correlation between change of PROMIS Pain Interference and change of SF- Physical Function after intervention.	PROMIS Pain Interference and SF-Physical Function attempt to measure loosely related constructs	Expect 0.20≤	Actual 0.50	Yes	Without Sensitivity Analysis
		Expect 0.20≤	***Actual 0.47***	***Yes***	***Sensitivity Analysis 3***
6. There will be at least a weak (0.1≤) positive correlation between change of PROMIS Pain Interference and change of 20 Meter Walk after intervention.	PROMIS Pain Interference and 20 Meter Walk attempt to measure loosely- related constructs, and 20 Meter Walk is not a patient- reported outcome	Expect 0.10≤	Actual 0.07	No	Without Sensitivity Analysis
		Expect 0.10≤	***Actual 0.13***	***Yes***	***Sensitivity Analysis 3***

SF=Short Form-36; WOMAC= Western Ontario and McMasters Osteoarthritis Index. Note: All correlations are shown as absolute values to facilitate easier interpretation.

*Correlation Scale (Spearman's rho) Without Sensitivity Analysis: 0.5≤ r, Strong; 0.4≤ r< 0.5, Moderate-Strong; 0.3≤ r <0.4, Moderate; 0.2≤ r <0.3, Weak-Moderate; 0.1≤ r < 0.2, Weak; r < 0.1, Negligible.

†Sensitivity Analysis 1: Increasing Correlation Cut-Off Points by 0.1, Correlation Scale (Spearman's rho): 0.6= r, Strong; 0.5= r< 0.6, Moderate-Strong; 0.4= r <0.5, Moderate; 0.3= r <0.4, Weak-Moderate; 0.2= r < 0.3, Weak; r < 0.2, Negligible.

‡Sensitivity Analysis 2: Added Hypothesis, Correlation Scale same as for the "Without Sensitivity Analysis", but total number of hypotheses is 8 instead of 6.

¶***Sensitivity Analysis 3: Based on change score correlations after removing those found with floor/ceiling effects for the given PROMIS measure.***

Table 2B

Hypotheses for the Responsiveness of PROMIS Physical Function 10a

Hypotheses	Rationale	Correlation Result(s)		Confirmed?	Hypothesis Type
1. There will be a strong (0.5≤)positive correlation between change of PROMIS Physical Function and change of SF-Physical Function after intervention	PROMIS Physical Function and SF-Physical Function attempt to measure similar constructs	Expect 0.50≤	Actual 0.59	Yes	Without Sensitivity Analysis *
					Sensitivity Analysis 3¶: Not applicable, Floor/ceiling effects were not found in Physical Function
2. The positive correlation of change on the PROMIS Physical Function with change on the SF-Physical Function will be at least 0.1 higher than the positive correlation of change on the PROMIS Physical Function with the SF-Mental Health	PROMIS Physical Function and SF-Physical Function attempt to measure similar constructs, but PROMIS Physical Function and SF-Mental Health attempt to measure unrelated constructs.	Expect 0.10≤	Actual 0.19	Yes	Without Sensitivity Analysis
					Sensitivity Analysis 3¶: Not applicable, Floor/ceiling effects were not found in Physical Function
3. There will be at least a moderate- strong (0.4≤) negative correlation between change of PROMIS Physical Function and change of WOMAC- Function after intervention.	PROMIS Physical Function and WOMAC-Function attempt to measure similar constructs, but the WOMAC Function is lower limb-specific	Expect 0.40≤	Actual 0.51	Yes	Without Sensitivity Analysis
					Sensitivity Analysis 3¶: Not applicable, Floor/ceiling effects were not found in Physical Function
4. There will be at least a weak- moderate (0.2 ≤) negative correlation between change of PROMIS Physical Function and change of WOMAC- Pain after intervention.	PROMIS Physical Function and WOMAC-Pain attempt to measure loosely-related constructs	Expect 0.20≤	Actual 0.49	Yes	Without Sensitivity Analysis
					Sensitivity Analysis 3¶: Not applicable, Floor/ceiling effects were not found in Physical Function
5. There will be at least a weak- moderate (0.2≤) positive correlation between change of PROMIS Physical Function and change of SF-Role Physical after intervention.	PROMIS Physical Function and SF-Role Physical attempt to measure loosely-related constructs	Expect 0.20≤	Actual 0.49	Yes	Without Sensitivity Analysis
					Sensitivity Analysis 3¶: Not applicable, Floor/ceiling effects were not found in Physical Function
6. There will be at least a weak- moderate (0.2≤) positive correlation between change of PROMIS Physical Function and change of 6 Minute Walk after intervention.	PROMIS Physical Function and 6 Minute Walk attempt to measure partially related constructs, but 6 Minute Walk is not a patient-reported outcome	Expect 0.20≤	Actual 0.08	No	Without Sensitivity Analysis
					Sensitivity Analysis 3¶: Not applicable, Floor/ceiling effects were not found in Physical Function

SF=Short Form-36; WOMAC= Western Ontario and McMasters Osteoarthritis Index. Note: All correlations are shown as absolute values to facilitate easier interpretation.

[*] Correlation Scale (Spearman's rho) Without Sensitivity Analysis: 0.5≤ r, Strong; 0.4≤ r< 0.5, Moderate-Strong; 0.3≤ r <0.4, Moderate; 0.2≤ r <0.3, Weak-Moderate; 0.1≤ r < 0.2, Wzeak; r < 0.1, Negligible.

[†] Sensitivity Analysis 1: Increasing Correlation Cut-Off Points by 0.1, Correlation Scale (Spearman's rho): 0.6= r, Strong; 0.5= r< 0.6, Moderate-Strong; 0.4= r <0.5, Moderate; 0.3= r <0.4, Weak-Moderate; 0.2= r < 0.3, Weak; r < 0.2, Negligible.

[‡] Sensitivity Analysis 2: Added Hypothesis, Correlation Scale same as for the "Without Sensitivity Analysis", but total number of hypotheses is 8 instead of 6.

[¶] ***Sensitivity Analysis 3: Based on change score correlations after removing those found with floor/ceiling effects for the given PROMIS measure.***

Table 3

Descriptive Statistics and Floor/Ceiling Effects of PROMIS Instruments

PROMIS Short Form	Baseline			Week 12			Mean Change (SD)	Mean Total SEM†
	Mean (SD) [SEM‡]	Floor* N (%)	Ceiling* N (%)	Mean (SD) [SEM‡]	Floor* N (%)	Ceiling* N (%)		
Depression	49.5 (8.8) [2.2]	0 (0)	0 (0)	48.5 (8.7) [2.4]	0 (0)	0 (0)	−1.0 (6.9)	2.3
Anxiety	50.1 (9.0) [2.3]	0 (0)	**28 (17.0)** *	49.4 (9.2) [2.3]	0 (0)	**33 (20.1)** *	−0.6 (7.3)	2.3
Physical Function	41.0 (5.6) [1.8]	0 (0)	1 (0.6)	44.8 (7.5) [2.0]	0 (0)	10 (6.1)	3.8 (6.9)	1.9
Pain Interference	57.7 (7.1) [1.7]	0 (0)	10 (6.1)	53.4 (8.3) [1.8]	0 (0)	**36 (21.8)** *	−4.3 (8.1)	1.8

Mean values reported as T-Scores, unless otherwise specified.

*Significant Floor or Ceiling Effect defined as ≥ 15% participants with ‘least healthy’ or ‘most healthy’ possible score, respectively.

†Mean Total Standard Error of Measurement (SEM) is calculated as the mean of the SEM at baseline and 12 weeks.

‡SEM for IRT-based instruments are not calculated using the traditional formula for SEM; rather, the SEM corresponds to the mean PROMIS T-score at a given time point, and its value can be found in the respective PROMIS scoring manual.

Table 4

Coefficients between Δ PROMIS and Δ Legacy Measures

gacy Health Outcome Measures								
General Health * −0.32	SF-Role Emotional * −0.26	SF-Mental Health * −0.36	SF-Role Physical * −0.36	SF- Vitality * −0.29	WOMAC Pain 0.24	Beck Depression 0.38	Perceived Stress 0.51	6-Minute Walk * −0.19
001	<0.001	<0.001	<0.001	<0.001	0.002	<0.001	<0.001	0.02
45, −0.18	−0.40, −0.11	−0.48, −0.22	−0.48, −0.21	−0.42, −0.14	0.09, 0.38	0.23, 0.50	0.39, 0.62	−0.34, −0.03
General Health * −0.34	SF-Role Emotional * −0.20	SF-Mental Health * −0.39	SF- Physical Function * −0.26	WOMAC Pain 0.18	Beck Depression 0.40	Perceived Stress 0.43	6-Minute Walk * −0.19	
001	0.01	<0.001	<0.001	0.02	<0.001	<0.001	0.02	
47, −0.20	−0.34, −0.04	−0.51, −0.25	−0.40, −0.11	0.03, 0.33	0.25, 0.52	0.30, 0.55	−0.34, −0.03	
General Health * 0.37	SF- Physical Function * 0.59	SF-Mental Health * 0.40	SF-Role Physical * 0.49	SF- Vitality * 0.48	WOMAC Pain −0.49	WOMAC Function −0.51	6-Minute Walk * 0.08	
001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.32	
3, 0.49	0.48, 0.68	0.26, 0.52	0.36, 0.60	0.35, 0.59	−0.60, −0.36	−0.62, −0.39	−0.08, 0.24	
General Health * −0.25	SF- Physical Function * −0.50	SF-Mental Health * −0.16	SF-Bodily Pain −0.53	WOMAC Pain 0.37	WOMAC Function 0.43	Perceived Stress 0.22	20-Meter Walk 0.07	
01	<0.001	0.04	<0.001	<0.001	<0.001	0.004	0.39	
39, −0.10	−0.60, −0.37	−0.31, −0.01	−0.63, −0.41	0.22, 0.49	0.30, 0.55	0.07, 0.36	−0.09, 0.23	

e to 12 weeks; CI= Confidence Intervals; SF=Short Form-36; WOMAC= Western Ontario and McMasters Osteoarthritis Index.

ealth.

Table 5

Responsiveness of PROMIS Short Forms in Knee Osteoarthritis

PROMIS Short Form	Sensitivity Analysis Type	Number of Hypotheses Confirmed, %	Responsiveness Rating*
Depression	**Without Sensitivity Analysis†**	**3 of 6; 50%**	**Moderate**
Sensitivity Analysis 1	Increasing Correlation Cut-Off Points‡	2 of 6; 33%	Poor
Sensitivity Analysis 2	Using 2 Additional Hypotheses¶	4 of 8; 50%	Moderate
Sensitivity Analysis 3	Controlling for Floor/Ceiling Effects#	Not applicable: Floor/ceiling effects were not found in Depression	
Anxiety	**Without Sensitivity Analysis†**	**4 of 6; 67%**	**Moderate**
Sensitivity Analysis 1	Increasing Correlation Cut-Off Points‡	1 of 6; 17%	Poor
Sensitivity Analysis 2	Using 2 Additional Hypotheses¶	6 of 8; 75%	Moderate
Sensitivity Analysis 3	Controlling for Floor/Ceiling Effects#	5 of 6; 83%	High
Physical Function	**Without Sensitivity Analysis†**	**5 of 6; 83%**	**High**
Sensitivity Analysis 1	Increasing Correlation Cut-Off Points‡	3 of 6; 50%	Moderate
Sensitivity Analysis 2	Using 2 Additional Hypotheses¶	7 of 8; 83%	High
Sensitivity Analysis 3	Controlling for Floor/Ceiling Effects#	Not applicable: Floor/ceiling effects were not found in Physical Function	
Pain Interference	**Without Sensitivity Analysis†**	**4 of 6; 67%**	**Moderate**
Sensitivity Analysis 1	Increasing Correlation Cut-Off Points‡	2 of 6; 33%	Poor
Sensitivity Analysis 2	Using 2 Additional Hypotheses¶	6 of 8; 75%	Moderate
Sensitivity Analysis 3	Controlling for Floor/Ceiling Effects#	4 of 6; 67%	Moderate

Note: for further details and individual hypotheses, see Tables 2A&B and Supplementary Tables 2A&B.

*Responsiveness Rating based on % of hypotheses confirmed: 75%< High, 50–75% Moderate, <50% Poor;

†Based on the results of the 6 original *a priori* hypotheses;

‡Original correlation cut-off points were increased by 0.1;

¶Based on the original 6 hypotheses plus 2 additional *a priori* hypotheses;

#Based on the 6 original *a priori* hypotheses after removing participants with lowest/highest possible scores.

Table 6

Quality Criteria for Minimally Important Difference Estimates

PROMIS Depression			
MID Legacy Anchor	**Spearman correlation† (Criterion #1)**	**Sub-sample Size‡ (Criterion #2)**	**MID Effect Size¶ (Criterion #3)**
Patient global *		X	X
SF-36 Mental Health	X	X	X
Beck Depression	X	X	
Perceived Stress	X	X	X
PROMIS Anxiety			
MID Legacy Anchor	**Spearman correlation (Criterion #1)**	**Sub-sample Size (Criterion #2)**	**MID Effect Size (Criterion #3)**
Patient global *		X	
SF-36 Mental Health	X	X	X
Beck Depression	X	X	
Perceived Stress	X	X	X
PROMIS Physical Function			
MID Legacy Anchor	**Spearman correlation (Criterion #1)**	**Sub-sample Size (Criterion #2)**	**MID Effect Size (Criterion #3)**
Patient global *	X	X	X
SF-36 Physical Function	X	X	X
WOMAC Function	X	X	
PROMIS Pain Interference			
MID Legacy Anchor	**Spearman correlation (Criterion #1)**	**Sub-sample Size (Criterion #2)**	**MID Effect Size (Criterion #3)**
Patient global *		X	X
SF-36 Bodily Pain	X	X	X
WOMAC Pain	X	X	X

SF=Short Form-36; WOMAC= Western Ontario and McMasters Osteoarthritis Index.

*Because the Patient Global is a single-item scale, important change was defined as having an improvement at least as large as its previously published MID, but no more than (1 + legacy MID).

†MID Quality Criteria 1: Anchor estimates must have had a Spearman correlation of (0.3≤) between its anchor and PROMIS change scores.

‡MID Quality Criteria 2: At least 10 participants must have been classified with minimal important change.

¶MID Quality Criteria 3: MID estimate must reflect an Effect size (Cohen's *d*) that is no smaller than 0.2 (plausibly important), but no larger than 0.8 (plausibly minimal).