Responsiveness of the PROMIS and FAAM Instruments in Foot and Ankle Orthopedic Population

Man Hung, PhD[1,2,3,4], **Judith F. Baumhauer, MD, MPH**[5], **Frank W. Licari, DDS**[6], **Jerry Bounsanga, BS**[1], **Maren W. Voss, ScD**[1], **Charles L. Saltzman, MD**[1]

[1]Department of Orthopaedic Surgery Operations, University of Utah, Salt Lake City, UT, USA

[2]Department of Family & Preventive Medicine, University of Utah, Salt Lake City, UT, USA

[3]Study Design and Biostatistics Center, University of Utah, Salt Lake City, UT, USA

[4]Huntsman Cancer Institute, Salt Lake City, UT, USA

[5]Department of Orthopaedic Surgery, University of Rochester, Rochester, NY, USA

[6]Roseman University of Health Sciences, South Jordan, UT, USA

Abstract

Background: Investigating the responsiveness of an instrument is important in order to provide meaningful interpretation of clinical outcomes. This study examined the responsiveness of the Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function (PF), the PROMIS Pain Interference (PI), and the Foot and Ankle Ability Measure (FAAM) Sports subscale in an orthopedic sample with foot and ankle ailments.

Methods: Patients presenting to an orthopedic foot and ankle clinic during the years 2014–2017 responded to the PROMIS and FAAM instruments prior to their clinical appointments. The responsiveness of the PROMIS PF v1.2, PROMIS PI v1.1, and FAAM Sports were assessed using paired samples *t* test, effect size (ES), and standardized response mean (SRM) at 4 different follow-up points. A total of 785 patients with an average age of 52 years (SD = 17) were included.

Results: The PROMIS PF had ESs of 0.95 to 1.22 across the 4 time points (3, >3, 6, and <6 months) and SRMs of 1.04 to 1.43. The PROMIS PI had ESs of 1.04 to 1.63 and SRMs of 1.17 to 1.23. For the FAAM Sports, the ESs were 1.25 to 1.31 and SRMs were 1.07 to 1.20. The ability to detect changes via paired samples *t* test provided mixed results. But in general, the patients with improvement had statistically significant improved scores, and the worsening patients had statistically significant worse scores.

Corresponding Author: Man Hung, PhD, Department of Orthopaedic Surgery Operations, University of Utah, 590 Wakara Way, Salt Lake City, UT 84108, USA., Man.Hung@hsc.utah.edu.

Conclusion: The PROMIS PF, PROMIS PI, and FAAM Sports were sensitive and responsive to changes in patient-reported health.

Level of Evidence: Level II, prospective comparative study.

While physical examination has been and always will be an important part of clinical care, the development and use of quality patient-reported outcome (PRO) instruments provides important information from the patient's perspective. Well-developed PRO instruments with sufficient reliability, validity, and interpretability of scores, particularly if they do not impose excessive respondent burden, can be an important adjunct to care.[10] The Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function (PF) and PROMIS Pain Interference (PI) are instruments increasingly utilized in clinical practice. They were developed to improve upon prior instruments, pulling validated and time-proven questions from existing questionnaires into a large pool of potential items, which were then categorized, reviewed, revised, and selected.[12] The systematic development process and ease of administration with computerized adaptive testing (CAT) make the PROMIS instruments an important contribution to clinical and research practice.[5,16]

Responsiveness refers to the ability of an instrument to detect differences in outcomes over time,[25] which is necessary to assess treatment effects. The PROMIS instruments were developed with individually calibrated items, providing a sensitivity to change 4 times greater than possible through other test development methods.[15] The PROMIS instruments were validated during the development process on a general population, but like all PRO instruments, they can be improved by testing longitudinally on clinical populations and in specific diagnostic or treatment groups. Responsiveness to a general population does not automatically ensure responsiveness in a specific clinical population. Longitudinal studies with data collected from multiple time points on the same individuals using the same PROs are necessary to determine responsiveness to treatment-related change.[44] This data collection takes time and the PROMIS instruments are new, so clinical studies assessing whether the PROMIS instruments are able to detect change are still needed.

Responsiveness can be assessed with both internal and external methods. Internal measures of responsiveness assess change from a baseline score to a follow-up score, providing valuable information about whether an instrument has detected a change and the magnitude of the change.[25] Because internal responsiveness uses the measure itself as the measuring stick for change, it does not automatically provide information on whether the change might be considered meaningful. External responsiveness addresses this limitation with a comparison of the instrument to a standard that anchors the amount of change to some other health measure.[25] The comparison might be made to a gold standard in treatment, a long-used clinical tool, physical examination, or any method that links the current treatment response to an alternative indicator of the patient's condition.[51] PROMIS measures are available for a number of domains, with pain and function as 2 domains specifically relevant to orthopedic practice. Prior research has demonstrated that function and pain, while related,

are only moderately correlated.[11] This difference indicates a uniqueness between the 2 constructs and emphasizes the importance of assessing changes in pain and function individually. For this reason, both the PROMIS PF and PROMIS PI were selected for evaluation. It is also useful to compare newer measures with existing instruments; thus, the Foot and Ankle Ability Measure (FAAM) Sports was selected for evaluation as well.

Given the recent development of PROMIS instruments, there has been limited investigation into the PF and PI instrument responsiveness in orthopedics. Initial investigation of responsiveness of PROMIS measures has shown that the PROMIS PF-20 outperformed older instruments (i.e., the Health Assessment Questionnaire [HAQ] Disability Index and the PF-10) in terms of responsiveness in a sample of 1000 individuals with rheumatoid arthritis. The PROMIS PF-20 detected a 1.2% difference at 80% power compared with the best legacy test with a 2.3% difference at 80% power.[15] In that analysis, the authors concluded that the sensitivity of the PROMIS instruments allows for smaller sample sizes in research without a loss of power due to their superior ability to detect change. Another study of 451 individuals with rheumatoid arthritis found the PROMIS-PF 20 to be more responsive than 2 other legacy measures of physical function (SF-36 and the HAQ).[19]

The FAAM Sports subscale is a targeted instrument developed to assess foot and ankle pathologies in clinical practice.[34] It has shown good correlation with PROMIS instruments. [38,39,47] Initial validation of the FAAM Sports showed that its subscale was more responsive to change at 4 weeks than a general measure of physical function,[35] and generally responsive to change at 6 months.[21] Further validation studies are needed to assess the responsiveness of the PROMIS instruments and FAAM Sports within specific clinical populations and contexts.[1]

The purpose of this study was to evaluate the responsiveness of the FAAM Sports and the 2 PROMIS item banks (PI and PF) administered by CAT to patients with foot and ankle problems. This study asked, "How responsive are the PROMIS PF and the FAAM Sports to changes in patients' report of function?" and "How responsive is the PROMIS PI to changes in patients' report of pain?"

Methods

Study Design

Meaningful change was defined as any patient-perceived improvement or worsening in condition, so only patients reporting improvement or deterioration (little/some/great relief/improvement or little/some/great deterioration/worsening) were included in each analysis. Individuals reporting worsening symptoms have shown evidence of lower levels of change from baseline,[13,50] and many measures have been found to be less responsive to deterioration than improvement.[18,20,33,36,49] One approach to handling the difference in positive or negative change is to evaluate improvement separately from deterioration. [2,3,9,14,37] The main focus of this study was to assess positive change in order to examine responsiveness to improvement. There was a small sample who reported worsening symptoms, examined as a secondary analysis for the purpose of verifying whether the scores

were dropping when symptoms were worsening. Institutional review board approval was obtained prior to the start of this study.

Four sequential follow-up periods were examined in this study: (1) 3-month follow-up (80 to 100 days after initial assessment); (2) >3-month follow-up (90 days or more after initial assessment); (3) 6-month follow-up (170 to 190 days after initial assessment); and (4) >6-month follow-up (180 days or more after initial assessment). Three and 6-month follow-up periods are common in orthopedic practices,[4,8,29–32,41,48] at times measured in terms of periods extending past the 3- and 6-month follow-ups.[17,27,46] These time points were included in this study to correspond with prior literature and clinical practice.

Participants

This study consisted of 785 consecutive patients presenting to an academic foot and ankle clinic with a minimum of 2 visits greater than 3 months apart between 2014 and 2017. All patients sought treatment for foot and ankle musculoskeletal conditions and were aged 18 years or older at the time of baseline assessment. Demographic questions and PRO instruments were administered on handheld tablet computers at the first clinic visit immediately prior to their appointment. Because patients reported their level of change in pain or function (which were used here as the external anchors) at each visit, the anchor responses might change or differ for each visit. Different patient samples might also be included in the analysis for each follow-up period as not all patients returned for follow-up visits at identical intervals of time.

Inclusion of patients in the sample for each analysis, for each measure at each time point, was dependent on whether there were nonmissing responses to an anchor question about either function or pain. Change levels for the PROMIS PF and FAAM Sports analyses were anchored by patient responses to the question: "Compared to your FIRST EVALUATION at the xx Center: how would you describe your PHYSICAL FUNCTION LEVEL now?" (much worse, worse, slightly worse, no change, slightly improved, improved, much improved). Similarly, the PROMIS PI consisted of the following anchor question: "Compared to your FIRST EVALUATION at the xx Center: how would you describe your episodes of PAIN now?" (much worse, worse, slightly worse, no change, slightly improved, improved, much improved). This type of anchor question is referred to as a global rating of change (GRC) scale, is a recommended anchor for evaluating PROs,[43] and has been commonly used in orthopedics.[28] The GRC was only used as an anchor, relating levels of change to patient-perceived improvement, ensuring that responsiveness could be assessed in terms of meaningful change from a patient perspective.

This study included a total of 785 patients: 471 women and 314 men with ages ranging from 18 to 91 years old (mean age = 52, SD = 17). The majority of patients were white ($n = 707$, 90.1%). Full demographic information is displayed in Table 1. There were 43 different diagnostic conditions in this foot and ankle population that precluded stratification by procedure groupings. Procedures included a range of services, such as amputations, fusions, reconstructions, repairs, arthroplasty, bunionectomy, and tendon transfers, among other treatments.

Instruments

The PROMIS PF v1.2 CAT and the PROMIS PI v1.1 CAT were administered to consecutive patients seeking treatment between 2013 and 2017. The FAAM Sports was administered, in addition to the PROMIS instruments, to all patients seeking treatment between 2016 and 2017, resulting in a smaller sample size of FAAM Sports administrations. The PROMIS PF v1.2 contains both upper extremity and lower extremity items and draws from a 121-item test bank. The PROMIS PI v1.1 has a 40-item test bank. The FAAM is constructed with 2 subscales, the 21-item AD subscale and the 8-item Sports subscale. Only the Sports subscale was administered as relevant to the orthopedic population being evaluated in our clinics.

The PROMIS instruments have been developed with a standardized scoring mechanism with a mean of 50 and a standard deviation (SD) of 10 in the *t* score scale.[45] All PROMIS items and measures were calibrated among the general population, with some oversampling among specific groups.[6] Higher levels of function are represented by higher scores on the PROMIS PF, and higher levels of pain interference with daily activity are represented by higher scores on the PROMIS PI. FAAM Sports items are scored on a 5-point scale from 0 (unable to do) to 4 (no difficulty at all), for a total score range from 0 to 32, with raw scores transformed into percentage scores[34] so that higher percentage scores reflect higher levels of function. Administration of these instruments was through the mEVAL—a web-based application maintained by the University of Utah. The PROMIS CAT algorithms were set at default (see http://www.healthmeasures.net). Administration occurred at baseline (either within 7 days prior to the clinic visit of a new foot and ankle condition or on the day of the first clinic visit) and at each follow-up visit patients attended.

Statistical Analyses

Analyses of patient demographic characteristics were performed using descriptive statistics. Change scores represented the difference between the baseline score and the follow-up score compared at 3 months (90 days ± 10 days, and 90 days and beyond) and 6 months (180 days ± 10 days, and 180 days and beyond) on the PROMIS PF, PROMIS PI, and FAAM Sports. One-way analysis of variance (ANOVA) examined the discriminative ability of the anchor question and the outcome scores at each time point, anchoring the PROMIS and FAAM Sports change scores to the patients' own reports of change during the treatment period. Paired sample *t* tests were used to test the hypothesis that there was no change in the outcomes between time points, tested both for change groups and for measures at each follow-up point. The significance level was set at .05, 2-sided. Paired sample *t* tests were appropriate for evaluation of relationships between 2 continuous variables, such as pre- and post-PROMIS scores.

Cohen's *d* was used as a standardized measure of effect size (ES), addressing variability that exists in scores, which is not included in the mean difference comparisons of the *t* test.[25] Cohen's *d* removes the dependence on sample size and is normalized based on the cross-sectional SD of scores, calculated as the difference between the baseline score and the follow-up score, divided by the baseline score's SD. In interpreting Cohen's *d*, a small ES is approximately *d* = 0.20, whereas *d* = 0.50 is considered a medium effect, and a large ES can be considered *d* = 0.80. An ES of 0.80 represents a large change where the difference is as

least as great as four-fifths of an SD in scores.[9] The standardized response mean (SRM) is similar to the paired *t* test, but this calculation removes the dependence on sample size from the equation[25] and is normalized based on the SD of the change score. SRM is calculated as the mean difference between baseline and follow-up scores divided by the SD of difference scores, reflecting individual changes in scores. Recommended guidelines for interpreting SRM are similar to Cohen's *d*, with 0.20, 0.50, and 0.80 for small, medium, and large ESs, respectively.[25]

All analyses were performed using SPSS 24.0 (IBM SPSS Statistics for Windows, IBM Corp., Armonk, NY)[26] and R 3.30 (R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria).[42]

Results

Descriptive statistics of the PROMIS PF, PROMIS PI, and FAAM Sports are presented in Table 2 for the improved and deteriorated groups. Inclusion criteria required a follow-up instrument score and anchor question response within the specified time window of the period. For the PROMIS PF, there were between 6491 and 8681 clinic patients who were not eligible for inclusion in each analysis. The non–follow-up group had baseline PROMIS PF scores ranging from 2 to 4 points higher than those in the follow-up window, suggesting higher function within this group. For the PROMIS PI, the 3545 to 4795 patients not having follow-up within the analyzed time windows had baseline PI scores that were an average of 2 to 3 points lower than those within the follow-up periods, indicating less pain interference. The 1795 to 2156 nonparticipants for the FAAM Sports had baseline scores that were 6 to 14 points higher on average, indicating higher function.

One-way ANOVA between the anchor question and the PROMIS PF at the 3-month follow-up was significant ($F(6, 67) = 2.42$, $P = .035$), providing evidence of discrimination between groups and showing that the improving group had higher function than the worsening group. The PROMIS PF showed significant Spearman correlations with the anchor questions at follow-up, though the 6-month correlations were lower. Spearman correlations were used due to the ordinal nature of the anchor questions. The associations between the PROMIS PF and the anchor questions were $r = 0.423$ ($P < .001$) at the 3-month, $r = 0.233$ ($P = .005$) at the >3-month, $r = 0.363$ ($P = .068$) at the 6-month, and $r = 0.111$ ($P = .287$) at the >6-month follow-ups. For those who improved, the PROMIS PF had significant change scores measured by paired sample *t* tests at the 3-month follow-up ($P = .044$), but they were not significant at the other time points. For the deteriorated group, there was no significant difference at the 3-month, 6-month, and >6-month follow-ups, and only the >3-month follow-up time point was significant ($P = .017$). Full results are presented in Table 3. The ES was large for all time points for the PROMIS PF, with values of 0.99, 0.96, 1.22, and 0.95 (3, >3, 6, and >6 months), respectively. The deteriorated group showed medium to large ESs for the PROMIS PF with 0.85, 0.74, 1.11, and 0.70 at the respective time points. The PROMIS PF had SRM values of 1.43, 1.17, 1.05, and 1.04 across the 4 time points (3, >3, 6, and >6 months), respectively. Similarly, the deteriorated group also showed high SRM values of 1.28, 1.08, 1.43, and 1.03, respectively.

The results for the one-way ANOVA between the anchor question and the PROMIS PI at the 3-month follow-up were significant ($F(6, 98) = 4.95$, $P < .001$). One-way ANOVA between the anchor questions and the other time points (i.e., >3-month and >6-month follow-ups) showed statistically significant differences with the exception of the 6-month follow-up period due to low sample size in the deteriorated group. The PROMIS PI showed significant Spearman correlations with the anchor questions at the 3-month ($r = -0.552$, $P < .001$), >3-month ($r = -0.362$, $P < .001$), 6-month ($r = -0.400$, $P = .026$), and >6-month ($r = -0.216$, $P = .019$) follow-ups. For those who improved, the PROMIS PI had significant change scores as measured by paired sample t tests at the 3-month follow-up ($P < .001$), >3-month follow-up ($P < .001$), and >6-month follow-up ($P < .001$). For the 6-month follow-up, the change scores were not significant, likely due to low sample size. For the deteriorated group, t tests of change scores for the PROMIS PI were not significant at all time points. The ES was large for the PROMIS PI, with values of 1.12, 1.04, 1.63, and 1.16 across the 4 time points, respectively. The deteriorated group's ESs for the PROMIS PI were 1.83, 0.76, 0.66, and 0.85 at the 4 time points. The PROMIS PI showed similarly large SRMs of 1.19, 1.22, 1.23, and 1.17, respectively, across the 4 time points. For the deteriorated group, the SRMs were 1.36, 1.24, 3.13, and 1.33, respectively.

The relationship between the anchor questions and the FAAM Sports at the 3-month follow-up as demonstrated by the one-way ANOVA was significant ($F(6, 74) = 2.99$, $P = .011$). The >3-month follow-up and the >6-month follow-up were also statistically significant, though the 6-month follow-up period was not, likely due to low sample size in the deteriorated group. The FAAM Sports showed significant Spearman correlations with the anchor questions at the 3-month ($r = 0.353$, $P < .001$) and >3-month ($r = 0.292$, $P < .001$) follow-ups, though not at the 6-month ($r = 0.026$, $P = .905$) and >6-month ($r = 0.104$, $P = .622$) follow-ups. For the improved group, the FAAM Sports had significant change scores at the 3-month ($P = .001$), >3-month ($P < .001$), and >6-month ($P < .001$) follow-ups, though not for the 6-month follow-up, likely due to a low sample size of 16. The deteriorated group showed a significant change score at the 3-month follow-up ($P = .040$) and at the >3-month follow-up ($P = .001$), whereas no significant change was demonstrated at the 6-month and >6-month follow-ups. The ES was large for the FAAM Sports, with values of 1.31, 1.25, 1.30, and 1.28, respectively, across the 4 time points. The deteriorated group showed medium to large ESs of 1.01, 1.24, 0.75, and 0.62, respectively. The FAAM Sports showed similarly high SRM values of 1.16, 1.20, 1.07, and 1.11, respectively, for the improved group. Similarly, the deteriorated group also showed high SRMs of 1.62, 1.28, 1.24, and 0.88. Detailed information is presented in Table 3.

Discussion

This study examined the responsiveness of the PROMIS PF, PROMIS PI, and FAAM Sports in patients with foot and ankle musculoskeletal conditions. Prior literature has stated that the PROMIS instruments, such as the PROMIS PF, have shown high responsiveness and strong psychometric properties, and have outperformed other legacy measures of physical function.[19,22–24] Findings from this current study suggest that the PROMIS PF exhibited high responsiveness to change with a large ES and SRM among a lower extremity population. Though the PROMIS PF was more responsive than the FAAM Sports for the improved

group, it was less responsive than the FAAM Sports for the deteriorated group. The PROMIS PI has had limited investigation into responsiveness in prior research. This study was able to establish that the PROMIS PI was highly responsiveness to change, where large ESs and SRMs were observed that were larger for the improved group than for the deteriorated group. This finding is consistent with prior research regarding the measurement of responsiveness to deterioration, that instruments tend to be less responsive to deterioration overall.[18,20,33,36,49] In terms of paired sample t test significance, the PROMIS PI demonstrated significant change for the improved group, but not for the deteriorated group. The FAAM Sports also showed high responsiveness for the improved group as well as medium to high responsiveness for the deteriorated group. The FAAM Sports was the only measure evaluated that showed significant responsiveness to change among the deteriorated group, though only at the 3-month and >3-month follow-up points.

The significant one-way ANOVA at the 3-month follow-up provides empirical evidence that the deteriorated group of patients had the lowest PROMIS PF and FAAM Sports scores, as would be expected, and the improved group had the highest PROMIS PF and FAAM Sports scores. The significant one-way ANOVA between the pain anchor question and 3-month follow-up PROMIS PI also showed evidence of validity of the anchor question used in this study. The PROMIS PF, PROMIS PI, and FAAM Sports all demonstrated meaningful changes among individuals who sought follow-up treatment at the clinic.

Utilizing 3 different methods, we found that the ES method identified medium to high responsiveness with values greater than 0.62 for all 3 instruments at 4 different follow-up periods for both the improved and deteriorated groups. Similarly, the SRM method indicated a large effect, with values greater than 0.88 for all instruments and groups at all time points evaluated. On the other hand, the paired sample t tests provided nonsignificant results at times when the sample sizes were low. This implies that the results from the ES and SRM methods are more trustworthy than the paired sample t test in the evaluation of responsiveness, since the ES and SRM methods are not sample size dependent. This study also included varying follow-up time points, which is important for cross-validation and understanding of the robustness of the findings.

Limitations

A potential limitation to this study is that the clinic sample had different patients with varying diseases and conditions, but we were unable to stratify by procedure code due to limited sample size for each procedure. Future studies may consider stratification if sufficient sample size can be obtained. Future research that is able to stratify the data by specific procedure codes might enhance our understanding of the sensitivity of the PROMIS PF and PROMIS PI to specific orthopedic clinical populations. Additionally, the demographics of the clinic sample may not be representative of the US population and may not be considered generalizable beyond the sample characteristics.

In this study, change was anchored with the use of global reports that rely on retrospective recall of prior function and may be subject to recall bias. A potential problem with GRC is that it relies on retrospective reflection, which may be only weakly correlated with treatment effect.[40] Reports of change may be more related to the current health status than the baseline

change value.[7] However, in this study the repeat measures of the PROs at follow-up were used to document changes in condition and were not subject to recall bias as measures of both baseline and current functioning, limiting the role of recall bias in the analysis. Another concern is that patients who return to the clinic for the longer-term visits may be more likely to come for continued pain or functional limitations and may not reflect the clinic population as a whole. This might result in a follow-up sample that does not capture the full range of outcomes, but is skewed in severity. Average PROMIS PF, PROMIS PI, and FAAM Sports scores for those clinic patients not included within each follow-up time period analyzed showed higher levels of function and less pain than the analyzed group. However, our sample is typical in orthopedic clinics, relevant to our clinical practice, and may be generalizable to other similar clinical practices. Future studies examining responsiveness for specific conditions and across different populations are recommended.

Conclusion

This study was able to determine the PROMIS PF, PROMIS PI, and FAAM Sports responsiveness to change. These instruments showed medium to high responsiveness to change regardless of the different indices across 4 follow-up periods. Results from this study can be used alongside other peer-reviewed studies for clinicians and researchers wanting to further confirm their decisions in selecting the most efficient and effective instruments to measure patient outcomes. The minimized respondent burden with CAT administration of the PROMIS measures makes them a preferred or desirable measurement tool where available. Our findings of responsiveness are important in advancing treatment protocols and give clinicians and researchers the knowledge to most effectively and efficiently interpret outcome measures.

Acknowledgments

References

1. Alonso J, Bartlett SJ, Rose M, et al. The case for an international Patient-Reported Outcomes Measurement Information System (PROMIS(R)) initiative. Health Qual Life Outcomes 2013;11:210. [PubMed: 24359143]

2. Beaton D, Bombardier C, Katz J, Wright J. A taxonomy for responsiveness. J Clin Epidemiol 2001;54(12):1204–1217. [PubMed: 11750189]

3. Beaton DE, Bombardier C, Katz JN, et al. Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome measures in rheumatology. Minimal clinically important difference. J Rheumatol 2001;28(2):400–405. [PubMed: 11246687]

4. Carmont MR, Silbernagel KG, Nilsson-Helander K, Mei-Dan O, Karlsson J, Maffulli N. Cross cultural adaptation of the Achilles tendon Total Rupture Score with reliability, validity and

responsiveness evaluation. Knee Surg Sports Traumatol Arthrosc 2013;21(6):1356–1360. [PubMed: 22847247]

5. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol 2010;63(11):1179–1194. [PubMed: 20685078]

6. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007;45(5 suppl 1):S3.

7. Cook CE. Clinimetrics corner: the minimal clinically important change score (MCID): a necessary pretense. J Man Manip Ther 2008;16(4):E82–E83. [PubMed: 19771185]

8. Cornell CN, Levine D, O'Doherty J, Lyden J. Unipolar versus bipolar hemiarthroplasty for the treatment of femoral neck fractures in the elderly. Clin Orthop Relat Res 1998;348:67–71.

9. de Groot V, Beckerman H, Uitdehaag BM, et al. The usefulness of evaluative outcome measures in patients with multiple sclerosis. Brain 2006;129(10):2648–2659. [PubMed: 16956906]

10. Deutsch L, Smith L, Gage B, Kelleher C, Garfinkel D. Patient-Reported Outcomes in Performance Measurement: Commissioned Paper on PRO-Based Performance Measures for Healthcare Accountable Entities Washington, DC: National Quality Forum; 2012.

11. DeVine J, Norvell DC, Ecker E, et al. Evaluating the correlation and responsiveness of patient-reported pain with function and quality-of-life outcomes after spine surgery. Spine (Phila Pa 1976) 2011;36(21 suppl):S69–S74. [PubMed: 21897347]

12. DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: the PROMIS qualitative item review. Med Care 2007;45(5 suppl 1):S12–S21. [PubMed: 17443114]

13. Eurich DT, Johnson JA, Reid KJ, Spertus JA. Assessing responsiveness of generic and specific health related quality of life measures in heart failure. Health Qual Life Outcomes 2006;4:89. [PubMed: 17125512]

14. Eyssen I, Steultjens M, Oud T, Bolt EM, Maasdam A, Dekker J. Responsiveness of the Canadian occupational performance measure. J Rehabil Res Dev 2011;48(5):517–528. [PubMed: 21674402]

15. Fries J, Rose M, Krishnan E. The PROMIS of better outcome assessment: responsiveness, floor and ceiling effects, and Internet administration. J Rheumatol 2011;38(8):1759–1764. [PubMed: 21807798]

16. Fries JF, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. J Rheumatol 2014;41(1):153–158. [PubMed: 24241485]

17. Gregory J, Harwood D, Gochanour E, Sherman S, Romeo A. Clinical outcomes of revision biceps tenodesis. Int J Shoulder Surg 2012;6(2):45–50. [PubMed: 22787333]

18. Harrison MJ, Davies LM, Bansback NJ, et al. The comparative responsiveness of the EQ-5D and SF-6D to change in patients with inflammatory arthritis. Qual Life Res 2009;18(9):1195–1205. [PubMed: 19777373]

19. 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 2015;74(1):104–107. [PubMed: 24095937]

20. Haywood K, Garratt A, Jordan K, Dziedzic K, Dawes P. Spinal mobility in ankylosing spondylitis: reliability, validity and responsiveness. Rheumatology 2004;43(6):750–757. [PubMed: 15163832]

21. Hung M, Baumhauer JF, Brodsky JW, et al. Psychometric comparison of the PROMIS Physical Function CAT with the FAAM and FFI for measuring patient-reported outcomes. Foot Ankle Int 2014;35(6):592–599. [PubMed: 24677217]

22. Hung M, Clegg DO, Greene T, Saltzman CL. Evaluation of the PROMIS Physical Function item bank in orthopaedic patients. J Orthop Res 2011;29(6):947–953. [PubMed: 21437962]

23. Hung M, Hon SD, Cheng C, et al. Psychometric evaluation of the lower extremity computerized adaptive test, the modified Harris hip score, and the hip outcome score. Orthop J Sports Med 2014;2(12):2325967114562191. [PubMed: 26535291]

24. Hung M, Stuart AR, Higgins TF, Saltzman CL, Kubiak EN. Computerized adaptive testing using the PROMIS Physical Function item bank reduces test burden with less ceiling effects compared

with the Short Musculoskeletal Function Assessment in orthopaedic trauma patients. J Orthop Trauma 2014;28(8):439–443. [PubMed: 24378399]

25. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000;53(5):459–468. [PubMed: 10812317]

26. IBM Corp. SPSS Statistics for Windows [computer program] Armonk, NY: IBM Corp.; 2015.

27. Ibrahim T, Beiri A, Azzabi M, Best AJ, Taylor GJ, Menon DK. Reliability and validity of the subjective component of the American Orthopaedic Foot & Ankle Society clinical rating scales. J Foot Ankle Surg 2007;46(2):65–74. [PubMed: 17331864]

28. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther 2009;17(3):163–170. [PubMed: 20046623]

29. Kotsis SV, Chung KC. Responsiveness of the Michigan Hand Outcomes Questionnaire and the Disabilities of the Arm, Shoulder and Hand Questionnaire in carpal tunnel surgery. J Hand Surg Am 2005;30(1):81–86. [PubMed: 15680560]

30. Landauer F, Wimmer C, Behensky H. Estimating the final outcome of brace treatment for idiopathic thoracic scoliosis at 6-month follow-up. Pediatr Rehabil 2003;6(3–4):201–207. [PubMed: 14713586]

31. Little DG, MacDonald D. The use of the percentage change in Oswestry Disability Index score as an outcome measure in lumbar spinal surgery. Spine (Phila Pa 1976) 1994;19(19):2139–2143. [PubMed: 7809744]

32. MacDermid JC, Richards RS, Donner A, Bellamy N, Roth JH. Responsiveness of the Short Form-36, Disability of the Arm, Shoulder, and Hand Questionnaire, patient-rated wrist evaluation, and physical impairment measurements in evaluating recovery after a distal radius fracture. J Hand Surg Am 2000;25(2):330–340. [PubMed: 10722826]

33. Mannion A, Porchet F, Kleinstück F, et al. The quality of spine surgery from the patient's perspective: part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. Eur Spine J 2009;18(3):374–379. [PubMed: 19296136]

34. Martin RL, Irrgang JJ. A survey of self-reported outcome instruments for the foot and ankle. J Orthop Sports Phys Ther 2007;37(2):72–84. [PubMed: 17366962]

35. Martin RL, Irrgang JJ, Burdett RG, Conti SF, Van Swearingen JM. Evidence of validity for the Foot and Ankle Ability Measure (FAAM). Foot Ankle Int 2005;26(11):968–983. [PubMed: 16309613]

36. Mehta T, Subramaniam AV, Chetter I, McCollum P. Assessing the validity and responsiveness of disease-specific quality of life instruments in intermittent claudication. Eur J Vasc Endovasc Surg 2006;31(1):46–52. [PubMed: 16226903]

37. Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: reliability, validity, and responsiveness. J Shoulder Elbow Surg 2002;11(6):587–594. [PubMed: 12469084]

38. Nixon D, McCormick J, Johnson J, Klein S. FAAM ADL scores correlate with PROMIS Physical Function, Pain Interference, and Depression outcomes. Foot Ankle Orthop 2017;2(2): 2473011417S000011.

39. Nixon DC, McCormick JJ, Johnson JE, Klein SE. PROMIS Pain Interference and Physical Function scores correlate with the Foot and Ankle Ability Measure (FAAM) in patients with hallux valgus. Clin Orthop Relat Res 2017;475(11): 2775–2780. [PubMed: 28836098]

40. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol 1997;50(8):869–879. [PubMed: 9291871]

41. Paatelma M, Kilpikoski S, Simonen R, Heinonen A, Alen M, Videman T. Orthopaedic manual therapy, McKenzie method or advice only for low back pain in working adults: a randomized controlled trial with one year follow-up. J Rehabil Med 2008;40(10):858–863. [PubMed: 19242624]

42. R Development Core Team. R: A Language and Environment for Statistical Computing [computer program] Vienna: R Foundation for Statistical Computing; 2010.

43. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61(2):102–109. [PubMed: 18177782]

44. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. Health Qual Life Outcomes 2006;4:70. [PubMed: 17005038]

45. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. J Clin Epidemiol 2014;67(5):516–526. [PubMed: 24698295]

46. Segal NA, Glass NA, Teran-Yengle P, Singh B, Wallace RB, Yack HJ. Intensive gait training for older adults with symptomatic knee osteoarthritis. Am J Phys Med Rehab 2015;94(10 0 1):848–858.

47. Slullitel GA. CORRInsights®: PROMIS Pain Interference and Physical Function scores correlate with the Foot and Ankle Ability Measure (FAAM) in patients with hallux valgus. Clin Orthop Relat Res 2017;475(11):2781–2782. [PubMed: 28940035]

48. Uchiyama S, Imaeda T, Toh S, et al. Comparison of responsiveness of the Japanese Society for Surgery of the Hand version of the carpal tunnel syndrome instrument to surgical treatment with DASH, SF-36, and physical findings. J Orthop Sci 2007;12(3):249–253. [PubMed: 17530376]

49. van der Heijden GJ, Leffers P, Bouter LM. Shoulder disability questionnaire design and responsiveness of a functional status measure. J Clin Epidemiol 2000;53(1):29–38. [PubMed: 10693901]

50. van der Maas NA. Patient-reported questionnaires in MS rehabilitation: responsiveness and minimal important difference of the multiple sclerosis questionnaire for physiotherapists (MSQPT). BMC Neurol 2017;17:50. [PubMed: 28302081]

51. Wyrwich K, Norquist J, Lenderking W, Acaster S. Methods for interpreting change over time in patient-reported outcome measures. Qual Life Res 2013;22(3):475–483. [PubMed: 22528240]

Table 1.

Demographic Characteristics of the Patients ($n = 785$).

Patient Characteristics	n	Percent	Mean (SD)	Range
Age, y			51.8 (16.9)	18–91
Gender				
Male	314	60.0		
Female	471	40.0		
Race				
White or Caucasian	707	90.1		
Asian	5	0.6		
American Indian and Alaska Native	6	0.9		
Black or African American	10	1.3		
Native Hawaiian and Other Pacific Islander	2	0.3		
Other	46	5.8		
Unknown/missing	8	1.0		
Ethnicity				
Non-Hispanic	736	93.8		
Hispanic	41	5.2		
Missing	8	1.0		

Table 2.

Descriptive Statistics of PROMIS and FAAM Sports Scores.

Time Points		Baseline	Follow-Up	Change from Baseline
Instrument	N	Mean (SD)	Mean (SD)	Mean (SD)
Improved Group				
3-month (90 ± 10 days)				
PROMIS PF	49	33.7 (9.6)	37.1 (8.0)	3.3 (8.4)
PROMIS PI	128	63.2 (6.8)	58.6 (8.2)	−4.6 (8.7)
FAAM Sports	63	18.8 (20.2)	31.9 (27.5)	13.1 (31.9)
3-month (90 days)				
PROMIS PF	321	36.8 (8.8)	37.5 (8.4)	0.8 (11.0)
PROMIS PI	327	62.6 (7.0)	58.9 (7.9)	−3.7 (8.5)
FAAM Sports	208	21.4 (22.2)	33.0 (25.9)	11.6 (33.4)
6-month (180 ± 10 days)				
PROMIS PF	20	36.5 (8.6)	32.2 (7.7)	−4.3 (14.0)
PROMIS PI	23	63.8 (5.7)	59.5 (8.2)	−4.3 (11.2)
FAAM Sports	16	22.9 (20.5)	25.1 (26.8)	3.7 (37.0)
6-month (180 days)				
PROMIS PF	170	37.1 (8.7)	38.6 (9.5)	1.5 (11.4)
PROMIS PI	169	62.9 (6.8)	57.9 (8.7)	−5.0 (8.6)
FAAM Sports	30	20.3 (20.3)	42.9 (24.1)	22.6 (26.4)
Deteriorated Group				
3-month (90 ± 10 days)				
PROMIS PF	13	35.8 (7.5)	31.8 (5.8)	−4.0 (7.2)
PROMIS PI	32	62.2 (7.4)	63.0 (8.6)	0.8 (7.7)
FAAM Sports	10	28.8 (18.0)	15.7 (12.2)	−13.1 (17.3)
3-month (90 days)				
PROMIS PF	143	38.4 (8.5)	36.7 (9.0)	−1.7 (8.4)
PROMIS PI	122	62.7 (7.5)	62.3 (8.1)	−0.4 (7.3)
FAAM Sports	49	34.0 (17.3)	26.0 (21.9)	−8.0 (24.5)
6-month (180 ± 10 days)				

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

| Time Points | | Baseline | Follow-Up | Change from Baseline |
Instrument	N	Mean (SD)	Mean (SD)	Mean (SD)
PROMIS PF	4	35.8 (3.7)	33.9 (5.0)	–2.0 (5.0)
PROMIS PI	7	64.1 (6.4)	64.1 (5.9)	0.0 (4.7)
FAAM Sports	6	27.6 (24.2)	21.9 (23.8)	–5.7 (24.3)
6-month (180 days)				
PROMIS PF	79	37.7 (8.2)	37.9 (8.1)	0.2 (7.9)
PROMIS PI	70	62.7 (6.3)	61.7 (7.0)	–1.0 (6.7)
FAAM Sports	11	38.7 (21.9)	30.8 (21.2)	–8.0 (19.9)

Abbreviations: FAAM, Foot and Ankle Ability Measure; PROMIS, Patient-Reported Outcomes Measurement Information System; PROMIS PF, PROMIS Physical Function; PROMIS PI, Pain Interference.

Table 3.

Responsiveness of the PROMIS and FAAM Sports Instruments.

Follow-Up Period	Instrument	n	SRM	ES	Paired *t* Test *P* Value
Improved Group					
3-month (90 ± 10 days)					
	PROMIS PF	49	1.43	0.99	.044
	PROMIS PI	128	1.19	1.12	.000
	FAAM Sports	63	1.16	1.31	.001
3-month (90 days)					
	PROMIS PF	321	1.17	0.96	.215
	PROMIS PI	327	1.22	1.04	.000
	FAAM Sports	208	1.20	1.25	.000
6-month (180 ± 10 days)					
	PROMIS PF	20	1.05	1.22	.179
	PROMIS PI	23	1.23	1.63	.080
	FAAM Sports	16	1.07	1.30	.693
6-month (180 days)					
	PROMIS PF	170	1.04	0.95	.064
	PROMIS PI	169	1.17	1.16	.000
	FAAM Sports	30	1.11	1.28	.000
Deteriorated Group					
3-month (90 ± 10 days)					
	PROMIS PF	13	1.28	0.85	.071
	PROMIS PI	32	1.36	1.83	.626
	FAAM Sports	10	1.62	1.01	.040
3-month (90 days)					
	PROMIS PF	143	1.08	0.74	.017
	PROMIS PI	122	1.24	0.76	.409
	FAAM Sports	49	1.28	1.24	.001
6-month (180 ± 10 days)					
	PROMIS PF	4	1.43	1.11	.485

Follow-Up Period	Instrument	*n*	SRM	ES	Paired *t* Test *P* Value
	PROMIS PI	7	3.13	0.66	.988
	FAAM Sports	6	1.24	0.75	.656
6-month (180 days)					
	PROMIS PF	79	1.03	0.70	.844
	PROMIS PI	70	1.33	0.85	.958
	FAAM Sports	11	0.88	0.62	.317

Abbreviations: FAAM, Foot and Ankle Ability Measure; PROMIS, Patient–Reported Outcomes Measurement Information System; PROMIS PF, PROMIS Physical Function; PROMIS PI, Pain Interference.