6-15-2022

The Current Utilization of Patient-reported Outcome Measurement Information System in Shoulder, Elbow, and Sports Medicine

David P. Trofa

Sohil S. Desai

Xinning Li

Eric C. Makhni

Review Article

The Current Utilization of Patient-reported Outcome Measurement Information System in Shoulder, Elbow, and Sports Medicine

David P. Trofa, MD
Sohil S. Desai, MD
Xinning Li, MD
Eric C. Makhni, MD, MBA

ABSTRACT

Clinical research using patient-reported outcome measures has been critical within the field of shoulder, elbow, and sports medicine in helping clinicians deliver evidence-based and value-based medicine. Recently, however, clinicians have advocated for improving the process of obtaining clinically meaningful information from patients while decreasing survey fatigue and increasing compliance. To that end, the National Institutes of Health created the Patient-Reported Outcome Measures Information System (PROMIS) in which a number of institutions and research investigations have adopted for reporting outcomes. A special focus has also been placed on PROMIS Computer Adaptive Testing forms, which tailor questioning through item response theory. The purpose of this study was to provide insight into the utilization, advantages, and disadvantages of PROMIS within the field of shoulder, elbow, and sports medicine and provide a comparison with legacy patient-reported outcome measure measurements.

From the Department of Orthopaedic Surgery, Columbia University Medical Center, NY, NY (Trofa and Desai), the Department of Orthopaedic Surgery, Boston University Medical Center, Boston, MA (Li), and the Department of Orthopaedic Surgery, Henry Ford Health System, Detroit, MI (Makhni).

Li or an immediate family member serves as a paid consultant to DePuy and A Johnson & Johnson Company and serves as a paid consultant and receives IP royalties from FH Ortho. Makhni or an immediate family member serves as a paid consultant to Smith & Nephew and serves as a paid presenter for Xodus Medical. Neither of the following authors nor any immediate family member has received anything of value from or has stock or stock options held in a commercial company or institution related directly or indirectly to the subject of this article: Trofa and Desai.

Supplemental digital content is available for this article. Direct URL citation appears in the printed text and is provided in the HTML and PDF versions of this article on the journal's Web site (www.jaaos.org).

J Am Acad Orthop Surg 2022;30:554-562

DOI: 10.5435/JAAOS-D-22-00030

Clinical research based on patient-reported outcome measures (PROMs) has allowed practitioners to critically assess treatment efficacies regarding patients' symptoms, function, and quality of life. Evaluating and understanding patient-reported outcomes (PROs) is increasingly important given today's ever-increasing focus on delivery of evidence-based medicine. In particular, these outcome measurements have been critical within the field of shoulder, elbow, and sports medicine (SESM) in the development and advancement of the treatment algorithms currently used within the field. Furthermore, PROs are critical for measuring the improvement of patient's health, which is an essential component for value-based care.

Currently, there are dozens of instruments used to assess PROs in the field of SESM. Regarding the upper extremity, the American Shoulder and Elbow Surgeons (ASES) scores and Disabilities of the Arm, Shoulder, and Hand (DASH) are commonly used. Regarding the lower extremity, the Knee Injury and Osteoarthritis Outcomes Score (KOOS), International Knee Documentation Committee (IKDC) scores, and Hip Outcome Score (HOS) are commonly used. Short Form-36 (SF-36) and EuroQol 5-dimensions (EQ-5D)

Questionnaire are popular validated general health measures used in orthopaedics.[1] However, there exist potential pitfalls when using PROMs for clinical outcomes evaluation and research, and these are evident in many of the "legacy" instruments.

To address these potential concerns, the National Institutes of Health sought to create a benchmark outcome metric that is psychometrically validated, standardized, and easy to administer. The result was the creation of the Patient-Reported Outcome Measures Information System (PROMIS), which is defined specifically as "a set of person-centered measures that evaluates and monitors physical, mental, and social health in adults and children."[2] A number of institutions and research investigations have adopted PROMIS for reporting outcomes related to the management of SESM injuries. The purpose of this study was to provide insight into the utilization, advantages, and disadvantages of PROMIS within the field of SESM and to provide a comparison with legacy PROMs.

PROMIS Overview

PROMIS was developed by the National Institutes of Health to serve as a population health and research tool. Despite the relative infancy of this instrument, over 2,000 peer-reviewed studies using PROMIS as a primary measure have been published.[2] The items in PROMIS place an emphasis on general health and health domains such as pain, physical function (PF), and depression, rather than on any specific diagnosis. There are currently 102 PROMIS adult domains and 1,903 total items, which encompass the critical tenets of physical and mental health.[2] Within SESM, four specific domains of PROMIS are commonly used: Pain Interference (PI), PF, PF-Upper Extremity (PF-UE), and Depression (Table 1). In addition, there are pediatric versions of these domains tailored to individuals younger than 18 years.

PROMIS can be administered in multiple ways. Within each domain, there are item banks that have been created from existing literature. From these, short forms (SF) with roughly four to eight items can be generated and administered electronically or on paper. The computer adaptive test (CAT) format, however, was the primarily intended form of administration. CATs have been designed to adapt to patient responses from previous items and pursue additional detail where relevant while avoiding excessive questions in redundant or irrelevant areas. This allows the potential for high

Table 1. PROMIS Domains in Shoulder, Elbow, and Sports Medicine

Domain	Description[a]	Scoring
Physical function	"Self-reported capability rather than actual performance of physical activities. This includes the functioning of one's upper extremities (dexterity), lower extremities (walking or mobility), and central regions (neck, back), as well as instrumental activities of daily living, such as running errands."	Higher score indicates better clinical status
Physical function—upper extremity	A subdomain of physical function. "Activities that require use of the upper extremity including shoulder, arm, and hand activities. Examples include writing, using buttons, or opening containers."	Higher score indicates better clinical status
Pain interference	"Consequences of pain on relevant aspects of one's life. This includes the extent to which pain hinders engagement with social, cognitive, emotional, physical, and recreational activities."	Lower scores indicate better clinical status
Depression	"Negative mood (sadness, guilt), views of self (self-criticism, worthlessness), and social cognition (loneliness, interpersonal alienation), as well as decreased positive affect and engagement (loss of interest, meaning, and purpose)."	Lower scores indicate better clinical status

[a]Descriptions from https://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/list-of-adult-measures.

psychometric validity without the need for excessively lengthy surveys.

All PROMIS items use 5-point scales—the most common scale in SESM-related PROMIS forms is as follows: 1 = "unable to do," 2 = "with much difficulty," 3 = "with some difficulty," 4 = "with a little difficulty," and 5 = "without any difficulty." Additional scales are presentedin Figure 1. The total PROMIS score from each assessment is generated using the principles of item response theory (IRT). With this methodology, each individual item is given a weight based on the responses and performance of the population, and thus, the respondent receives a scaled score calculated from the total number of weighted points from each item.[3] PROMIS scores follow a t-distribution with a mean of 50 and a standard deviation of 10 points. This allows clinicians and researchers to easily quantify and compare PROs, regardless of the specific item content or number of items used in the SF or CAT. Of note, a high score on PF suggests better physical function than the mean while a high score on PI suggests more pain interference than the mean (ie, poorer clinical status).

PROMIS Versus Legacy Instruments in Shoulder, Elbow, and Sports Medicine

Concepts of Validation and Overall Advantages of PROMIS Versus Legacy Instruments

PROMIS measures are widely regarded as validated PRO instruments.[4,5] It is important to understand, however, that there are several distinct aspects of psychometric validation that must be achieved for the instrument to be deemed "valid." Critical aspects of validation include responsiveness (ie, improvement in scores as a patient recovers from injury), criterion validity (ie, correlation between new instrument and existing validated instruments), content validity (ie, ability of an instrument to evaluate all facets of a given domain or syndrome), and test-retest reliability (ie, same result achieved on multiple administrations).[6] Construct validity is generally regarded as the qualitative determination of whether the instrument effectively measures its target construct. Validation of a PROM for one specific population does not guarantee its validation for a different population. Thus, subsequent sections of this review will examine joint-specific and condition-specific validity for PROMIS instruments. All key findings have been summarized in Appendices 1 through 4 (http://links.lww.com/JAAOS/A795).

An important advantage of PROMIS is the standardized nature of this instrument. Different legacy instruments have been used to evaluate and research the same sports medicine pathologies. This creates difficulty when comparing the outcome literature surrounding each sports medicine condition. The use of PROMIS may facilitate larger systematic reviews and overall improved translatability of data from one clinical research study to the next by using a standardized measure for each specific condition.

Another important advantage of PROMIS is the overall decreased rate of ceiling and floor effects compared with legacy instruments.[7] A ceiling effect occurs when high proportions of respondents achieve the

Figure 1

PROMIS Item Scoring Scales					
Scale	1	2	3	4	5
	——————— Physical Function (PF) Scales ———————>				
PF Style 1	Unable to do	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty
PF Style 2	Cannot do	Quite a lot	Somewhat	Very little	Not at all
PF Style 3	Cannot do because of hea	A lot of difficulty	Some difficulty	A little bit of difficulty	No difficulty at all
PF Style 4	Not at all	A little	Moderately	Mostly	Completely
	<——————— Pain Interference (PI) Scales ———————				
PI Style 1	Not at all	A little bit	Somewhat	Quite a bit	Very much
PI Style 2*	Never	Rarely	Sometimes	Often	Always
PI Style 3	Never	Once a week or less	Once every few days	Once a day	Every few hours

*Represents the scale used in all Depression items as well
Arrows and green color represent improving clinical status

Diagram demonstrating various PROMIS item scoring scales. It also highlights that high scores (both gross and when scaled to a mean of 50) represent good clinical status for domains such as Physical Function but poor clinical status for domains such as Pain Interference and Depression. *Represents the scale used in all Depression items as well. Arrows and green color represent improving clinical status.

maximum score on an instrument, thus allowing poor discrimination between the highest scorers. Similarly, a floor effect occurs when a high proportion of respondents achieve the minimum possible score, providing poor discrimination between the lowest scorers. These effects can lower content validity by failing to capture the upper or lower extreme of a given syndrome and lower responsiveness. The t-distribution of PROMIS scores and IRT scoring approach greatly reduce the occurrence of these effects. However, ceiling and floor effects with PROMIS have been identified in young and healthy populations, as well as with some specific assessments in certain sports medicine conditions as will be discussed later.

PROMIS has the potential to improve overall efficiency when collecting PROs. Both the SF and CAT methods of administration have been shown to require decreased overall time for completion when compared with many legacy instruments.[8] The PF CAT has shown median completion times as low as 40 to 45 seconds while the PI CAT has shown median times as low as 25 to 30 seconds.[5] This is valuable in light of the reports of "survey fatigue" that have hindered current PRO-collection efforts.[4]

PROMIS also allows for improved interpretability of PROM results. Each SF or CAT addresses a single domain, and a poor score represents a specific deficit in that domain. Legacy instruments often involve multiple domains (eg, pain, PF, and range of motion), and as a result, poor scores on such instruments may not identify the primary source of the poor score. Finally, PROMIS and IRT-based PROMs in general are dynamic constructs that have the ability for continued improvements over time. Specific items can be interchanged as their relative utility becomes more evident with additional individuals completing the assessment. This represents a unique advantage over legacy measures, which are fixed once published.

Overall Disadvantages of PROMIS Versus Legacy Instruments

A limitation of PROMIS is the increased up-front cost associated with computerized testing. In addition, there are some electronic health records that do not support PROMIS CAT integration. The paper SF version of PROMIS can be used in these settings, but this results in submaximal efficiency. Furthermore, although PROMIS has been translated to include over 40 languages, cultural and language competence can affect the objective measurements.

Multiple SFs or CATs are necessary to encompass multiple domains. Legacy instruments often cover multiple domains, such as the DASH score evaluating physical function and pain. To evaluate shoulder pain and functionality with PROMIS, at least two separate questionnaires would be required.

Because PROMIS domains generally are not specific to one anatomic region, persistent deficits in one body part may mask posttreatment improvements in a different anatomic region.[9] For example, improvements in clinical status after hip arthroscopy that are identified with more anatomically specific legacy measures (ie, modified Harris Hip Score [mHHS]) may not be identified with PROMIS PF because of persistent deficits caused by concomitant knee or ankle pathology in the cohort. This may limit the overall responsiveness of PROMIS measures in patients with multiple musculoskeletal comorbidities.

Finally, PROMIS item banks are a dynamic entity and are subject to change over time. This poses theoretical challenges when integrating results from older PROMIS item banks or related research studies with newer assessments of each PROMIS domain. Although the principles of IRT ultimately provide a scaled PROMIS score for each domain that is valid over time, regardless of the specific items used, this is an important consideration nonetheless as PROMIS continues to evolve.

PROMIS in Upper Extremity and Shoulder Conditions

PROMIS Validation and Comparisons With Legacy Instruments: General Upper Extremity and Shoulder

The most commonly used PROMIS instruments for shoulder pathology are PF, PF-UE, PI, and Depression. Commonly used legacy instruments include the ASES score, DASH, and Simple Shoulder Test (SST). Many studies have sought to validate and compare PROMIS scores with legacy instruments for general upper extremity pathology and specific shoulder conditions.[4,8,10-18] The primary goals of many of these comparisons have been to establish criterion validity between PROMIS and legacy instruments, assess for floor and ceiling effects, and to determine whether PROMIS reduces overall question burden. Regarding correlation values presented in this review, weak correlation is defined as $r = 0.30$ to 0.49, moderate as $r = 0.50$ to 0.69, and strong as $r \geq 0.70$.

One study involving 84 patients from an upper extremity clinic identified a strong correlation between

PROMIS PF-UE CAT and QuickDASH scores. In addition, the authors found a markedly shorter mean administration time with PROMIS versus the 11-question QuickDASH instrument (70 versus 116 seconds, $P < 0.001$) and identified no floor or ceiling effects with either measurement.[19] A separate study noted a moderate correlation between PROMIS PF CAT and QuickDASH scores and also found correlations between PROMIS PI (r = 0.74) and Depression (r = 0.34) scores with QuickDASH scores in these patients.[20] This may represent the ability of PROMIS to capture the overall burden of injury, especially mental health and general well-being, regardless of the specific diagnosis. Finally, a cross-sectional study involving 134 patients with non-shoulder upper extremity injuries administered both PF CAT and 30-question DASH instrument.[10] A strong correlation between the two measures was found, but there was a shorter mean time for completion of the PF CAT (57 versus 262 seconds, $P < 0.001$).

PROMIS Validation and Comparisons in Specific Shoulder Conditions

Among 187 patients with rotator cuff pathology, Beckmann et al[8] compared the PROMIS PF CAT with the SST and ASES Shoulder Assessment. The mean number of questions required in the PF CAT was 4.3, lower than the 11-question and 12-question SST and ASES Shoulder Assessment forms, respectively. There was a moderate correlation identified between PF CAT and both legacy assessments. The PF CAT demonstrated low floor (3.20%) and ceiling (0.53%) effects, whereas the SST demonstrated notable floor effects (21%). All three instruments showed excellent item reliability (ie, respondents with similar shoulder function answered items similarly), but only the PF CAT showed excellent person reliability (ie, respondents with different shoulder function were reliably ranked).[8] Fisk et al[12] demonstrated that the PF-UE, PI, and Depression CATs all demonstrated responsiveness in patients undergoing arthroscopic rotator cuff repair. Patterson et al[21] also showed that the PROMIS PF-UE scores had a stronger correlation with legacy shoulder scores compared with the PROMIS PF scores in patients with symptomatic rotator cuff tears.

Hajewski et al[15] evaluated PROMIS PF CAT, PF-UE SF, and legacy PROMs in 71 patients with surgically managed glenohumeral instability. There was moderate-to-strong correlation between each of the PROMIS measures and the ASES and SF-36 PF scores at all time points except between the PF CAT and ASES at 6 weeks, which demonstrated a weak correlation. The authors identified responsiveness in both PF CAT and PF-UE SF, with both scores declining in the initial postoperative assessment secondary to postoperative restrictions, but increasing above preoperative levels by 6 months postoperatively. Of note, notable ceiling effects were identified at 6 months postoperatively with the PF-UE (68.1%) but not with PF CAT (0%). Ceiling effects at this time point were also noted with the ASES score (23.6%) and SF-36 PF (41.7%). Thus, the relatively high function of this young population (mean age 22.1 years) with glenohumeral instability invalidates many of these measures due to inability to distinguish high scorers, with only the PROMIS PF CAT showing both responsiveness and overall construct validity.

Kolade et al evaluated PROMIS PF-UE CAT scores with both ASES scores and Constant Shoulder Scores (CSS) in 100 patients with adhesive capsulitis. The PF-UE CAT required fewer questions and less time to complete compared with both ASES scores and CSS. In addition, the PF-UE CAT displayed a strong correlation with both legacy measures and displayed no ceiling or floor effects in this patient population. The authors concluded that PROMIS PF-UE CAT displayed comparable efficacy with the commonly used shoulder legacy outcome scores and with less survey burden to the patients.[22]

Morgan et al[17] analyzed a cohort of 47 patients older than 60 years with displaced proximal humerus fractures after treatment. They found moderate-to-strong correlations between PROMIS PF CAT scores and three general orthopaedic questionnaires (DASH, Short Musculoskeletal Functional Assessment [SMFA] Bothersome Index, and SMFA Functional Index) and the CSS. This study also found a statistically significant decrease in the median time to completion between the PROMIS PF CAT (98 seconds) and both the DASH (336 seconds, $P < 0.001$) and SMFA (482 seconds, $P < 0.001$) instruments, with no statistically significant difference from the CSS (91 seconds).

PROMIS in Elbow Conditions

Studies evaluating PROMIS have also been conducted in patient populations with elbow injuries, with PROMIS PF CAT and PF-UE CAT as the primary instruments evaluated. Hung et al administered the PROMIS PF CAT and DASH to 1,759 patients seeking care for elbow conditions. They found high internal reliability (ie, the ability for an item to distinguish between groups of patients) for both measures and also found that both instruments displayed unidimensionality (ie, assessed one

primary domain or factor rather than multiple) with an unexplained variance of 3.3% for PF CAT and 4.5% for DASH.[23] It is likely that both instruments primarily evaluated PF in this cohort. There was a gender bias identified in 15% of PF CAT and 40% of DASH items through the use of differential item functioning, demonstrating the important concept that scores on PROMs may not always be comparable across sex and other demographic categories.[24] Finally, they did not identify notable floor or ceiling effects for either instrument in their general elbow complaint cohort; however, the PF CAT ceiling of 9.04% was higher than in other orthopaedic cohorts. The minimal clinically important difference (MCID) for PROMIS PF and PF-UE has been evaluated by Randall et al in a group of 146 isolated elbow trauma patients, which included both surgical and nonsurgical patients. The authors found that the MCID ranged from 4.3 to 5.7 for the PROMIS PF CAT and 4.6 to 4.8 for the PROMIS PF-UE CAT.[25]

Rojas et al[26] analyzed PROMIS PF CAT, PF-UE, and several legacy measures in 76 patients undergoing one of three elbow procedures: ulnar collateral ligament reconstruction, distal biceps tendon repair, or elbow arthroscopic surgery. The authors found strong correlations between PROMIS PF CAT with SF-36 PF and DASH, as well as between PROMIS UE and both SF-36 and DASH. Notable ceiling effects were observed with both the PROMIS UE (33%) and the SF-36 (20%), perhaps again related to the relatively high baseline function of this young cohort (mean age 35 years).

PROMIS in Knee Conditions

PROMIS Validation and Comparisons With Legacy Instruments: General Knee

Regarding the knee, the primary PROMIS instruments that have been evaluated thus far are PROMIS PF, PI, and Depression.[27-38] The most commonly used legacy PROMs in knee pathology include the KOOS and IKDC.[31] Miles et al[27] administered the PROMIS PF CAT and the IKDC Subjective Knee Form to 412 patients undergoing knee surgery identifying a strong correlation between the two measures without floor or ceiling effects. Another study from the same patient cohort noted that the Numeric Pain Scale resulted in floor and ceiling effects, whereas the PROMIS PI CAT did not. Tenan et al[28] interestingly identified that PROMIS PF CAT and PI CAT can be used, along with patient demographic identifiers, to accurately predict IKDC scores using a multivariate model.

PROMIS Validation and Comparisons in Specific Knee Conditions

Hancock et al[29] administered the PROMIS PF CAT, SF-36 PF, KOOS, EQ-5D, and Marx Knee Activity Rating Scale (MARS) to 100 patients undergoing anterior cruciate ligament (ACL) surgery. Strong correlations were identified between the PF CAT and all legacy measures. No floor or ceiling effects were identified with the PROMIS PF CAT, but there were notable ceiling effects for SF-36 General Health (15.0%) and MARS (42.0%). A low question burden (mean 4.2) was identified with the PF CAT. Gulledge et al[38] evaluated the responsiveness of PROMIS PF, PI, and Depression CATs in 100 patients undergoing ACL reconstruction demonstrating notable responsiveness in all three measures, with mean preoperative PF, PI, and Depression scores improving markedly.

Lu et al administered the PROMIS PF, PI, and Depression CATs as well as the IKDC, KOOS, and MARS instruments to 152 patients undergoing meniscal surgery. A strong correlation between PF CAT and KOOS-Sports, KOOS-PF, and IKDC was identified, as well as between PI CAT and IKDC and KOOS-ADL. No ceiling or floor effects were observed for PROMIS PF or PI CAT; however, notable floor effects were observed with the Depression CAT (25%). The MARS instrument demonstrated notable ceiling (18.8%) and floor (17.6%) effects.[31] Bernholt et al[32] evaluated the responsiveness of PROMIS PF, PI, and Depression CATs in 75 patients undergoing partial meniscectomy identifying substantial baseline morbidity with a mean preoperative PF of 38.5 and PI of 63.5. Notable responsiveness was identified in PF CAT (mean 43.4, $P < 0.001$) and PI CAT (mean 55.5, $P < 0.001$) 6 weeks postoperatively. Patients with high-grade cartilage lesions had less improvement in PF ($P = 0.014$) and PI ($P = 0.010$) at the 6-week follow-up compared with those with low-grade or no chondromalacia.

In a study of 319 knee articular cartilage procedures, Shamrock et al[34] reported strong correlation of the PROMIS PF CAT as a preoperative outcome assessment tool when compared with the knee legacy PROM measures with no ceiling or floor effect and a minimal time burden for completion. In another large cohort of 250 patients undergoing knee articular cartilage procedures, a strong correlation was identified at 6 months postoperatively between both PROMIS PF and PI CATs and several legacy measures.[33] There were no ceiling or floor effects identified for PROMIS PF, PI, and Depression CATs at the preoperative or postoperative assessment. Regarding responsiveness, only small

effect sizes on Cohen *d* analysis were observed between preoperative and postoperative administrations of the three PROMIS CATs, suggesting relatively weak PROMIS score responsiveness in patients undergoing cartilage procedures.

PROMIS in Hip Conditions

PROMIS PF, PI, and Depression CATs along with three legacy PROMs (Hip Outcome Score [HOS]-ADL, HOS-Sport Subscale, and the International Hip Outcome Tool) were compared in 96 patients undergoing arthroscopic surgery for femoroacetabular impingement syndrome (FAI).[9] At the 6-month follow-up, strong correlations were found between both PF and PI CATs and all three legacy measures, with the exception of a moderate correlation between PF CAT and International Hip Outcome Tool-12. All scores were significantly improved at the 6-month postoperative assessment compared with preoperative levels ($P < 0.001$), except for PROMIS Depression CAT. Notable ceiling effects were identified with PF CAT (15.22%) and HOS-ADL (24.4%).

In a retrospective analysis of 124 patients undergoing primary hip arthroscopy for FAI, Bodendorfer et al sought to determine the MCID, patient acceptable symptom state (PASS), and substantial clinical benefit (SCB) for PROMIS PF and PI CATs. The authors determined the MCID, PASS, and SCB thresholds at 1 year postoperatively to be 3.3, 47.0, and 49.9, respectively, for PF CAT and −3.1, 53.7, and 51.9, respectively, for PI CAT. They found that patients with younger age and lower BMI preoperatively were more likely to attain these thresholds.[39]

Patients with symptomatic acetabular dysplasia were evaluated by Li et al in a prospective study.[40] A total of 57 patients underwent periacetabular osteotomy and were assessed preoperatively and at a mean of 1.5 years postoperatively with the PROMIS PF and PI CATs along with several legacy PROMs (including HOS-Pain, HOS-ADL, and mHHS). PROMIS PF CAT showed weak-to-moderate correlations with HOS-ADL ($r = 0.51$, $P < 0.01$) and mHHS ($r = 0.49$, $P < 0.01$) preoperatively but stronger correlations with HOS-ADL ($r = 0.56$, $P < 0.01$) and mHHS ($r = 0.56$, $P < 0.01$) postoperatively. PROMIS PI CAT showed moderate-to-strong correlations with HOS-Pain and mHHS both preoperatively and postoperatively. None of the PROMs showed notable floor or ceiling effects preoperatively, but the legacy instruments showed notable ceiling effects postoperatively.[40]

Summary

PROMIS has the potential to reshape the way PROs are measured in SESM, and the literature has demonstrated several key findings regarding its use thus far. PROMIS measures have been shown to reduce overall question burden, which is critical in combating "survey fatigue" that may be increasing administrative costs and decreasing scientific validity because of unanswered survey items and less thoughtful responses per item. PROMIS has shown criterion validity and responsiveness in a range of SESM populations, including general upper extremity, rotator cuff, shoulder instability, proximal humerus fractures, ACL reconstruction, meniscal injury, FAI, and symptomatic acetabular dysplasia patients. Appropriate criterion validity has been demonstrated for PROMIS instruments in knee articular cartilage disorders, but poor responsiveness of these instruments has been observed thus far. There remain ceiling and floor effects (at the "better" clinical status extreme, in all cases) with some PROMIS instruments in young populations. Although the overall occurrence of these effects has been lower with PROMIS than with legacy PROMs in most populations, this suggests that additional improvements can be made to distinguish the highest or lowest performers from each other. Finally, room for additional research exists regarding PROMIS in SESM. This includes more prospective studies to evaluate clinical significance (ie, MCID, PASS, and SCB) for each specific SESM condition and to continue establishing responsiveness of PROMIS scores in specific SESM pathologies and patient populations.

References

References printed in bold type are those published within the past 5 years.

1. Hancock KJ, Glass N, Anthony CA, et al: Performance of PROMIS for healthy patients undergoing meniscal surgery. *J Bone Joint Surg Am Volume* 2017;99:954-958.

2. NIH: *PROMIS Health Organization*. 2021. https://www.healthmeasures.net/explore-measurement-systems/promis. Accessed November 2021.

3. Cella D, Chang CH: A discussion of item response theory and its applications in health status assessment. *Med Care* 2000;38(9 suppl):II66-72.

4. Makhni EC, Meadows M, Hamamoto JT, Higgins JD, Romeo AA, Verma NN: Patient Reported Outcomes Measurement Information System (PROMIS) in the upper extremity: The future of outcomes reporting? *J Shoulder Elbow Surg* 2017;26:352-357.

5. Kadri O, Jildeh TR, Meldau JE, et al: How long does it take for patients to complete PROMIS scores?: An assessment of PROMIS CAT questionnaires administered at an ambulatory sports medicine clinic. *Orthop J Sports Med* 2018;6:2325967118791180.

6. Nunnally JC: *Psychometric Theory*. New York, NY, McGraw-Hill, 1978.

7. Gulledge CM, Smith DG, Ziedas A, Muh SJ, Moutzouros V, Makhni EC: Floor and ceiling effects, time to completion, and question burden of PROMIS CAT domains Among shoulder and knee patients undergoing nonoperative and operative treatment. *JB JS Open Access* 2019;4:e0015.1-7.

8. Beckmann JT, Hung M, Bounsanga J, Wylie JD, Granger EK, Tashjian RZ: Psychometric evaluation of the PROMIS Physical Function Computerized Adaptive Test in comparison to the American Shoulder and Elbow Surgeons score and Simple Shoulder Test in patients with rotator cuff disease. *J Shoulder Elbow Surg* 2015;24:1961-1967.

9. Nwachukwu BU, Rasio J, Beck EC, et al: Patient-reported outcomes measurement information system physical function has a lower effect size and is less responsive than legacy hip specific patient reported outcome measures following arthroscopic hip surgery. *Arthroscopy* 2020;36:2992-2997.

10. Tyser AR, Beckmann J, Franklin JD, et al: Evaluation of the PROMIS physical function computer adaptive test in the upper extremity. *J Hand Surg* 2014;39:2047-2051.e4.

11. Matar RN, Shah NS, Grawe BM: Patient-Reported Outcomes Measurement Information System scores are inconsistently correlated with legacy patient-reported outcome measures in shoulder pathology: A systematic review. *Arthroscopy* 2021;37: 1301-1309.e1.

12. Fisk F, Franovic S, Tramer JS, et al: PROMIS CAT forms demonstrate responsiveness in patients following arthroscopic rotator cuff repair across numerous health domains. *J Shoulder Elbow Surg* 2019;28:2427-2432.

13. Minoughan CE, Schumaier AP, Fritch JL, Grawe BM: Correlation of Patient-Reported Outcome Measurement Information System physical function upper extremity computer adaptive testing, with the American Shoulder and Elbow Surgeons shoulder assessment form and simple shoulder test in patients with shoulder pain. *Arthroscopy* 2018;34:1430-1436.

14. Strong B, Maloney M, Baumhauer J, et al.: Psychometric evaluation of the Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function and Pain Interference Computer Adaptive Test for subacromial impingement syndrome. *J Shoulder Elbow Surg* 2019;28:324-329.

15. Hajewski CJ, Glass NA, Westermann RW, Bollier M, Wolf BR, Hettrich C: Performance of the PROMIS after operative interventions for shoulder instability. *Orthop J Sports Med* 2019;7: 2325967119846920.

16. Anthony CA, Glass NA, Hancock K,Bollier M, Wolf BR, Hettrich CM: Performance of PROMIS instruments in patients with shoulder instability. *Am J Sports Med* 2017;45:449-453.

17. Morgan JH, Kallen MA, Okike K, Lee OC, Vrahas MS: PROMIS physical function computer adaptive test compared with other upper extremity outcome measures in the evaluation of proximal humerus fractures in patients older than 60 years. *J Orthop Trauma* 2015;29: 257-263.

18. Lu Y, Agarwalla A, Patel BH, et al: Relationship between the Patient-Reported Outcomes Measurement Information System (PROMIS) computer adaptive testing and legacy instruments in patients undergoing isolated biceps tenodesis. *J Shoulder Elbow Surg* 2020;29:1214-1222.

19. Doring AC, Nota SP, Hageman MG, Ring DC: Measurement of upper extremity disability using the Patient-Reported Outcomes Measurement Information System. *J Hand Surg* 2014;39:1160-1165.

20. Overbeek CL, Nota SP, Jayakumar P, Hageman MG, Ring D: The PROMIS physical function correlates with the QuickDASH in patients with upper extremity illness. *Clin Orthop Relat Res* 2015;473:311-317.

21. Patterson BM, Orvets ND, Aleem AW, et al: Correlation of Patient-Reported Outcomes Measurement Information System (PROMIS) scores with legacy patient-reported outcome scores in patients undergoing rotator cuff repair. *J Shoulder Elbow Surg* 2018;27:S17-S23.

22. Kolade O, Ghosh N, Luthringer TA, et al: Correlation of Patient Reported Outcome Measurement Information System (PROMIS) with American Shoulder and Elbow Surgeon (ASES), and Constant (CS) scores in idiopathic adhesive capsulitis. *J Shoulder Elbow Surg* 2021;30:554-560.

23. Hung M, Voss MW, Bounsanga J, Gu Y, Granger EK, Tashjian RZ: Psychometrics of the Patient-Reported Outcomes Measurement Information System Physical Function instrument administered by computerized adaptive testing and the Disabilities of Arm, Shoulder and Hand in the orthopedic elbow patient population. *J Shoulder Elbow Surg* 2018;27:515-522.

24. Meldau JE, Borowsky P, Blanchett J, et al: Impact of patient demographic factors on preoperative Patient-Reported Outcomes Measurement Information System (PROMIS) physical function, pain interference, and depression computer adaptive testing scores in patients undergoing shoulder and elbow surgery. *Orthop J Sports Med* 2019;7:2325967119884543.

25. Randall DJ, Zhang Y, Harris AP, et al: The minimal clinically important difference of the Patient-Reported Outcomes Measurement Information System (PROMIS) physical function and upper extremity computer adaptive tests and QuickDASH in the setting of elbow trauma. *JSES Int* 2021;5:1132-1138.

26. Rojas EO, Glass N, Owens J, et al: Performance of the PROMIS in patients undergoing 3 common elbow procedures. *Orthop J Sports Med* 2019;7:2325967119852595.

27. Miles M, Nadarajah V, Jauregui JJ, et al: Evaluation of the PROMIS physical function computer adaptive test in patients undergoing knee surgery. *J Knee Surg* 2020;33:810-817.

28. Tenan MS, Robins RJ, Sheean AJ, et al: A high-sensitivity international knee documentation committee survey index from the PROMIS system: The next-generation patient-reported outcome for a knee injury population. *Am J Sports Med* 2021;49:3561-3568.

29. Hancock KJ, Glass N, Anthony CA, et al: PROMIS: A valid and efficient outcomes instrument for patients with ACL tears. *Knee Surg Sports Traumatol Arthrosc* 2019;27:100-104.

30. Scott EJ, Westermann R, Glass NA, Hettrich C, Wolf BR, Bollier MJ: Performance of the PROMIS in patients after anterior cruciate ligament reconstruction. *Orthop J Sports Med* 2018;6: 2325967118774509.

31. Lu Y, Beletsky A, Nwachukwu BU, et al: Performance of PROMIS physical function, pain interference, and depression computer adaptive tests instruments in patients undergoing meniscal surgery. *Arthrosc Sports Med Rehabil* 2020;2:e451-e459.

32. Bernholt D, Wright RW, Matava MJ, Brophy RH, Bogunovic L, Smith MV: Patient Reported Outcomes Measurement Information System scores are responsive to early changes in patient outcomes following arthroscopic partial meniscectomy. *Arthroscopy* 2018;34:1113-1117.

33. Nwachukwu BU, Beletsky A, Naveen N, et al: Patient-Reported Outcomes Measurement Information System (PROMIS) instruments correlate better with legacy measures in knee cartilage patients at postoperative than at preoperative assessment. *Arthroscopy* 2020; 36:1419-1428.

34. Shamrock AG, Wolf BR, Ortiz SF, et al: Preoperative validation of the Patient-Reported Outcomes Measurement Information System in patients with articular cartilage defects of the knee. *Arthroscopy* 2020;36:516-520.

35. Hajewski CJ, Baron JE, Glass NA, et al: Performance of the Patient-Reported Outcome Measurement Information System in patients with patellofemoral instability. *Orthop J Sports Med* 2020;8: 2325967120915540.

36. Kenney RJ, Houck J, Giordano BD, Baumhauer JF, Herbert M, Maloney MD: Do Patient Reported Outcome Measurement Information System (PROMIS) scales demonstrate responsiveness as well as disease-specific scales in patients undergoing knee arthroscopy? *Am J Sports Med* 2019;47:1396-1403.

37. Trasolini NA, Korber S, Gipsman A, San AE, Weber AE, Hatch GFR III: Performance of PROMIS computer adaptive testing as compared with established instruments for multiple-ligament knee injuries. *Orthop J Sports Med* 2019;7:2325967119867419.

38. Gulledge CM, Koolmees D, Smith DG, et al: The PROMIS CAT demonstrates responsiveness in patients after ACL reconstruction across numerous health domains. *Orthop J Sports Med* 2021;9:2325967120979991.

39. Bodendorfer BM, DeFroda SF, Clapp IM, Newhouse A, Nwachukwu BU, Nho SJ: Defining clinically significant improvement on the Patient-Reported Outcomes Measurement Information System test at 1-year follow-up for patients undergoing hip arthroscopy for the treatment of femoroacetabular impingement syndrome. *Am J Sports Med* 2021;49:2457-2465.

40. Li DJ, Clohisy JC, Schwabe MT, Yanik EL, Pascual-Garrido C: PROMIS versus legacy patient-reported outcome measures in patients undergoing surgical treatment for symptomatic acetabular dysplasia. *Am J Sports Med* 2020;48:385-394.