

# Use of PROMIS for Patients Undergoing Primary Total Shoulder Arthroplasty

S. Blake Dowdle,<sup>\*†</sup> MD, Natalie Glass,<sup>†</sup> PhD, Chris A. Anthony,<sup>†</sup> MD, and Carolyn M. Hettrich,<sup>†</sup> MD  
*Investigation performed at the Department of Orthopaedic Surgery and Rehabilitation, University of Iowa Hospitals and Clinics, Iowa City, Iowa, USA*

**Background:** The Patient-Reported Outcomes Measurement Information System (PROMIS) consists of question banks for health domains through computer adaptive testing (CAT).

**Hypothesis:** For patients with glenohumeral arthritis, (1) there would be high correlation between traditional patient-reported outcome (PRO) measures and the PROMIS upper extremity item bank (PROMIS UE) and PROMIS physical function CAT (PROMIS PF CAT), and (2) PROMIS PF CAT would not demonstrate ceiling effects.

**Study Design:** Cohort study (diagnosis); Level of evidence, 3.

**Methods:** Sixty-one patients with glenohumeral osteoarthritis were included. Each patient completed the American Shoulder and Elbow Surgeons (ASES) assessment form, Marx Shoulder Activity Scale, Short Form-36 physical function scale (SF-36 PF), EuroQol 5 Dimensions (EQ-5D) questionnaire, Western Ontario Osteoarthritis Shoulder (WOOS) index, PROMIS PF CAT, and the PROMIS UE. Correlation was defined as high (>0.7), moderate (0.4-0.6), or weak (0.2-0.3). Significant floor and ceiling effects were present if more than 15% of individuals scored the lowest or highest possible total score on any PRO.

**Results:** The PROMIS PF demonstrated excellent correlation with the SF-36 PF ( $r = 0.81, P < .0001$ ) and good correlation with the ASES ( $r = 0.62, P < .0001$ ), EQ-5D ( $r = 0.64, P < .001$ ), and WOOS index ( $r = 0.51, P < .01$ ). The PROMIS PF demonstrated low correlation with the Marx scale ( $r = 0.29, P = .02$ ). The PROMIS UE demonstrated good correlation with the ASES ( $r = 0.55, P < .0001$ ), SF-36 ( $r = 0.53, P < .01$ ), EQ-5D ( $r = 0.48, P < .01$ ), and WOOS ( $r = 0.34, P < .01$ ), and poor correlation with the Marx scale ( $r = 0.06, P = .62$ ). There were no ceiling or floor effects observed. The mean number of items administered by the PROMIS PRO was 4.

**Conclusion:** These data suggest that for a patient population with operative shoulder osteoarthritis, PROMIS UE and PROMIS PF CAT may be valid alternative PROs. Additionally, PROMIS PF CAT offers a decreased question burden with no ceiling effects.

**Keywords:** arthritis; glenohumeral joint; patient-reported outcomes; PROMIS; shoulder arthroplasty; computer adaptive testing

Patient-reported outcomes (PROs) are essential in the evaluation and treatment of orthopaedic patients. The aim of orthopaedic treatments is largely to improve physical function and quality of life,<sup>6,7,10</sup> and valid and precise measurement instruments are essential. Orthopaedic surgeons have become increasingly reliant on PROs when discussing surgical options and possible postoperative improvement with patients. With the increased usage of PROs, physicians have begun to decrease or even eliminate previously

utilized PROs that have proven to be unreliable, invalid, or inefficient. To be an effective PRO, the tool of use must demonstrate reliability, validity, and efficiency and be specific to the patient population in question.<sup>7,10</sup>

Previously validated PRO instruments used in the upper extremity include the American Shoulder and Elbow Surgeons (ASES) assessment form,<sup>12</sup> Western Ontario Osteoarthritis of the Shoulder (WOOS) index,<sup>13</sup> and Marx Shoulder Activity Score.<sup>4</sup> Other PROs can be used for overall health-related quality of life—namely the EuroQol 5 Dimensions questionnaire (EQ-5D)<sup>11</sup> and the Short Form-36 Health Survey physical function scale (SF-36 PF).<sup>3</sup> Indeed, several of these PRO instruments provide summary scores that may include different health domains, such as pain, physical function, and mental health. As a result, selection of the most important instrument for patient assessment and interpretation of scores over time or among instruments may be challenging. Another concern is that respondent burden and the reality of busy clinic practices limit the feasibility of administering multiple PRO instruments that might otherwise be considered ideal for comprehensive patient

\*Address correspondence to Spencer Dowdle, MD, 200 Hawkins Drive, Iowa City, IA 52242, USA (email: spencer-dowdle@uiowa.edu).

†Department of Orthopaedic Surgery and Rehabilitation, University of Iowa Hospitals and Clinics, Iowa City, Iowa, USA.

One or more of the authors has declared the following potential conflict of interest or source of funding: C.M.H. receives research support from Tornier.

Ethical approval for this study was obtained from the University of Iowa Human Subjects Office/Institutional Review Board (No. 201211724).

The Orthopaedic Journal of Sports Medicine, 5(9), 2325967117726044

DOI: 10.1177/2325967117726044

© The Author(s) 2017

assessment. With this in mind, the National Institutes of Health developed the Patient-Reported Outcomes Measurement Information System (PROMIS).

PROMIS was developed to advance PRO instruments by developing question banks for major health domains and by providing computerized adaptive testing (CAT) tools.<sup>5</sup> CAT utilizes item banks that have been developed through item response theory, which is used to examine responses to individual questions as well as the relationships among questions in a specific domain.<sup>10,14</sup> It has the ability to reduce respondent burden while preserving test precision by selecting relevant questions from the item bank specific to the patient's level of health in a particular domain, such as physical function. The PROMIS upper extremity item bank (PROMIS UE) consists of 16 questions directed at upper extremity musculoskeletal conditions; the PROMIS physical function CAT (PROMIS PF CAT; version 1.2) is a broader physical function instrument composed of 121 possible questions that assess the upper and lower extremities. Previous work has been performed to validate PROMIS for patients with hand and upper extremity pathology; however, these studies have either excluded shoulder patients altogether or have lumped all upper extremity pathologies into a single cohort, potentially confounding the results and masking the validity of these findings in specific diseases. More recently, Anthony et al<sup>1</sup> evaluated the use of PROMIS in patients with shoulder instability. While they were able to validate the use of PROMIS in this population, the diagnoses of instability and osteoarthritis affect 2 vastly different patient populations. Thus, we find it necessary to validate the use of PROMIS for patients undergoing shoulder arthroplasty for end-stage osteoarthritis of the glenohumeral joint. Furthermore, previous authors have concluded that further work is needed, as PROMIS instruments must be validated in specific patient populations and disease processes.<sup>8,9</sup>

In this study, we hypothesized that for our patients with end-stage glenohumeral arthritis, there would be good to excellent correlation between traditional orthopaedic upper extremity PROs and the PROMIS UE and PROMIS PF CAT; furthermore, we hypothesized that the PROMIS PF CAT would not demonstrate ceiling effects in an older patient population.

## METHODS

The present study was deemed Health Insurance Portability and Accountability Act compliant and approved by our institutional review board. We enrolled 61 patients with a primary diagnosis of shoulder osteoarthritis. Three patients had bilateral symptoms. All patients were scheduled for shoulder arthroplasty for their shoulder condition. Of the 61 patients, 10 had  $\geq 1$  incomplete PROs and were thus excluded from our study. Potential participants were identified by daily review of appointment lists and enrolled at the time of their preoperative clinic visit. All participants provided written informed consent. Participants prospectively completed the ASES, Marx, SF-36 PF, WOOS, PROMIS PF CAT, and PROMIS UE on a computer kiosk during their preoperative office visit. Participant demographic data, including age, sex, body mass

TABLE 1  
Demographics of Included Patients

Variable	No. (%)	Mean $\pm$ SD	Median (Range)
Age, y		60.8 $\pm$ 12.9	62.5 (29-87)
Sex			
Male	31 (58.5)		
Female	22 (41.5)		
Body mass index, kg/m <sup>2</sup>		33.9 $\pm$ 6.8	32.3 (23.1-54.1)

index (kg/m<sup>2</sup>), and operative side, were obtained through chart reviews.

Descriptive analyses were completed (frequency distributions and estimation of summary measures). The normality of variables was assessed via the Shapiro-Wilk test and by evaluating histograms and Q-Q plots. Based on the results of these analyses, the relationships among PROs were described with Pearson or Spearman correlation coefficients. Correlation was defined as excellent ( $>0.7$ ), excellent-good (0.61-0.7), good (0.31-0.6), or poor (0.2-0.3).<sup>9</sup> Floor and ceiling effects were also evaluated by determining the proportion of participants who achieved the highest and lowest possible scores, respectively. Floor and ceiling effects were considered present if more than 15% of individuals scored the lowest or highest possible total score on a PRO.<sup>2,3,6,7</sup> A prospective sample size was estimated. To detect a correlation of 0.4 (moderate) between PROs with 80% power and an alpha of .05, a minimum sample size of 29 was required. Statistical software (SAS, v 9.4; SAS Institute Inc) was utilized for analyses, and a *P* value  $<.05$  was considered statistically significant.

## RESULTS

Of 53 patients included in the final analyses, 22 (41.5%) were female and 31 (58.5%) were male. The mean age was 60.8  $\pm$  12.9 years, and the mean body mass index was 33.9  $\pm$  6.8 kg/m<sup>2</sup> (Table 1). The PROMIS UE demonstrated good correlation with ASES ( $r = 0.55$ ,  $P < .0001$ ), SF-36 PF ( $r = 0.53$ ,  $P < .0001$ ), EQ-5D ( $r = 0.48$ ,  $P = .002$ ), and WOOS ( $r = 0.34$ ,  $P < .01$ ) and poor correlation with Marx ( $r = 0.06$ ,  $P = .62$ ). The PROMIS PF CAT demonstrated excellent correlation with SF-36 PF ( $r = 0.81$ ,  $P < .0001$ ); good-excellent correlation with ASES ( $r = 0.62$ ,  $P < .0001$ ), EQ-5D ( $r = 0.64$ ,  $P < .001$ ), and WOOS ( $r = 0.51$ ,  $P < .01$ ); and poor correlation with Marx ( $r = 0.29$ ,  $P = .02$ ) (Table 2). There were no ceiling or floor effects observed. The mean number of items administered by the PROMIS PF CAT was 4 (range, 4-6; median, 4). The number of items administered by the other PROs was as follows: SF-36 PF, 10; ASES, 11; Marx, 6.5; and WOOS, 28.

## DISCUSSION

PROs play a key role in evaluation of patients and are critical in maintaining a flow of communication among patients, surgeons, and health care systems. The number of PROs available is vast and highly variable in question

TABLE 2  
Correlations Between PROMIS Physical Function and Upper Extremity Score and Validated Patient-Administered Outcome Instruments (Pearson or Spearman Coefficient)<sup>a</sup>

PROMIS	ASES Score	SF-36 PF	EQ-5D Questionnaire	WOOS Index	Marx Shoulder Activity Scale
Upper extremity	0.55 ( $P < .01$ )	0.53 ( $P < .01$ )	0.48 ( $P < .01$ )	0.34 ( $P < .01$ )	0.06 ( $P = .62$ )
Physical function	0.62 ( $P < .01$ )	0.81 ( $P < .01$ )	0.64 ( $P < .01$ )	0.51 ( $P < .01$ )	0.29 ( $P = .02$ )

<sup>a</sup>ASES, American Shoulder and Elbow Surgeons; EQ-5D, EuroQol 5 Dimensions; PROMIS, Patient-Reported Outcomes Measurement Information System; SF-36 PF, Short Form-36 Health Survey Physical Function Scale; WOOS, Western Ontario Osteoarthritis Shoulder.

burden. We hypothesized that there would be good to excellent correlation between traditional PROs and the PROMIS UE and PROMIS PF CAT and that the PROMIS PF CAT would not demonstrate ceiling effects in a population of patients undergoing shoulder arthroplasty for end-stage glenohumeral disease. We report good to excellent correlations between most of the previously validated PROs and the PROMIS UE and PROMIS PF CAT. We also report that no ceiling or floor effects were present.

The findings in this study, although encouraging and supportive to PROMIS UE and PROMIS PF CAT, do not justify the altogether abandonment of the standard PROs in glenohumeral osteoarthritis at this time. These previously validated PROs have the ability to establish small nuances among patients that may not be assessed in the CAT-based models. In addition, 1 limitation is the bias toward inclusion of some, and exclusion of other, commonly used PROs—namely, the University of California at Los Angeles (UCLA) Shoulder Score, the Constant Score, and the Shoulder Activity Score. Per physician preference and familiarity, we included the previously mentioned PROs while excluding some common PROs out of concern for overburdening our patient cohort, while the UCLA Shoulder Score has not yet been validated.<sup>16</sup> However, with the increasing body of literature validating and supportive of PROMIS UE CAT, it does beg the question of whether standard paper-based PROs will be replaced with CAT. The correlation in this study adds to this body of literature and validates the use of PROMIS UE and PROMIS PF CAT for patients with glenohumeral arthritis; however, we would like to see higher correlation prior to the abandonment of traditional PROs.

One of the potential drawbacks to CAT and to PROMIS UE and PROMIS PF CAT in particular is the potential for ceiling or floor effects. Hung et al<sup>10</sup> found that PF CAT was unable to capture differences between upper and lower extremity patients. However, Tyser et al<sup>15</sup> found no ceiling or floor effects when using PF CAT among hand and upper extremity patients. For patients with glenohumeral arthritis, we assume that they would not be pushing the limits of physical function, because of the age and baseline function of the population, and would thus not see a ceiling effect. Importantly, there was no floor effect seen, indicating that the questions asked addressed appropriate functional status for this population.

Most traditional PRO forms suffer 2 critical flaws—namely, the length or duration to complete the form and the narrow scope of applicability. With PROMIS PF CAT, these 2 flaws are directly addressed. We were able to show a

decreased test burden, with patients required to complete only 4 questions on average, while maintaining the correlation to ASES, WOOS, and SF-36. Also, by drawing on a question bank of individually validated questions, PROMIS UE and PROMIS PF CAT can be more widely applied to patient populations. This dramatically reduces the burden for test takers.

Validation, reliability, test response, and test fatigue are all reasons to select one PRO over another. These areas have all been established among the SF-36, Marx, ASES, and WOOS.<sup>2,6,7,10</sup> CAT PROs have the added benefit of being highly efficient, which consequently has been shown to decrease patient burden by as much as 10 times among different PROs.<sup>7</sup> In general terms, computer-based PROs have been shown to have increased test-retest reliability when compared with traditional forms. Computer-based administration has shown better distributional properties, lower means, more variance, higher internal consistency reliabilities, and stronger intercorrelations.<sup>7</sup>

The results of our study were primarily limited by the smaller sample size and the narrow scope of patient diagnosis as compared with previously reported validation studies for the PROMIS.<sup>6,7,15</sup> This study was conducted at a single institution, which could limit the overall generalizability of the results. The results could have been influenced by the nature of PRO evaluation in sequential questioning. This could have led to test fatigue and thus altered the results of each PRO. Also, there is no gold standard to base our testing instrument against. Finally, patients were questioned on a cross-sectional basis at a single point in time, which could have skewed the results of ceiling or floor effects.

## CONCLUSION

The PROMIS PROs have shown good to excellent correlation with previously established PROs among patients with primary osteoarthritis of the shoulder. They also show lower patient burden with no ceiling or floor effects in the same patient population. PROMIS is a good alternative for traditional PROs in this patient population.

## REFERENCES

1. Anthony CA, Glass NA, Hancock KJ, Bollier M, Wolf BR, Hettrich CM. Performance of PROMIS instruments in patients with shoulder instability. *Am J Sports Med.* 2017;45(2):449-453.

2. 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 Elb Surg.* 2015;24:1961-1967.
3. Bohannon RW, DePasquale L. Physical Functioning Scale of the Short Form (SF) 36: internal consistency and validity with older adults. *J Geriatr Phys Ther.* 2010;33(1):16-18.
4. Brophy RH, Beauvais RL, Jones EC, Cordasco FA, Marx RG. Measurement of shoulder activity level. *Clin Orthop Relat Res.* 2005;439:101-108.
5. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res.* 2007;16(suppl 1):133-141.
6. Coster MC, Rosengren BE, Bremander A, Karlsson MK. Surgery for adult acquired flatfoot due to posterior tibial tendon dysfunction reduces pain, improves function and health related quality of life. *Foot Ankle Surg.* 2015;21(4):286-289.
7. Doring AC, Nota SP, Hageman MG, Ring DC. Measurement of upper extremity disability using the Patient-Reported Outcomes Measurement Information System. *J Hand Surg Am.* 2014;39(6):1160-1165.
8. 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.
9. 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.
10. 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 to the short musculoskeletal function assessment in orthopaedic trauma patients. *J Orthop Trauma.* 2014;28(8):439-443.
11. Jansson KA, Granath F. Health-related quality of life (EQ-5D) before and after orthopedic surgery. *Acta Orthop.* 2011;82(1):82-89.
12. Kocher MS, Horan MP, Briggs KK, Richardson TR, O'Holleran J, Hawkins RJ. Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. *J Bone Joint Surg Am.* 2005;87(9):2006-2011.
13. Lo IK, Griffin S, Kirkley A. The development of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: the Western Ontario Osteoarthritis of the Shoulder (WOOS) index. *Osteoarthritis Cartilage.* 2001;9(8):771-778.
14. Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res.* 1997;6(6):595-600.
15. Tyser AR, Beckmann J, Franklin JD, et al. Evaluation of the PROMIS physical function computer adaptive test in the upper extremity. *J Hand Surg Am.* 2014;39(10):2047-2051.e4.
16. Wylie JD, Beckmann JT, Granger E, Tashjian RZ. Functional outcomes assessment in shoulder surgery. *World J Orthop.* 2014;5(5):623-633.