



PROMIS physical function underperforms psychometrically relative to American Shoulder and Elbow Surgeons score in patients undergoing anatomic total shoulder arthroplasty

Michael C. Fu, MD, MHS*, Brenda Chang, MS, MPH, Alexandra C. Wong, BS, Benedict U. Nwachukwu, MD, MBA, Russell F. Warren, MD, David M. Dines, MD, Joshua S. Dines, MD, Frank A. Cordasco, MD, MS, Stephen Lyman, PhD, Lawrence V. Gulotta, MD

Sports Medicine and Shoulder Service, Hospital for Special Surgery, New York, NY, USA

Background: The purpose of this study was to evaluate the psychometric properties of the Patient-Reported Outcomes Measurement Information System (PROMIS) physical function computer adaptive test (PF-CAT) relative to the American Shoulder and Elbow Surgeons (ASES) score in patients with glenohumeral osteoarthritis undergoing primary anatomic total shoulder arthroplasty (TSA).

Methods: A retrospective study of an institutional TSA registry was performed. Preoperative PROMIS PF-CAT and ASES scores were collected. Floor and ceiling effects were determined, and convergent validity was established through Pearson correlations. Rasch partial credit modeling was used for psychometric analysis of the validity of PF-CAT and ASES question items. Person-item maps were generated to characterize the distribution of question responses along the latent dimension of shoulder disability.

Results: Responses from 179 patients (184 shoulders) were included. PF-CAT had a moderate correlation to ASES ($r = 0.487$; $P < .001$), with no floor or ceiling effects; ASES had a 1.1% floor effect and no ceiling effect. With iterative Rasch model item-reduction analysis eliminating poorly fitting question items, all possible PF-CAT items were eliminated after 6 iterations. With ASES, just 1 function question item was dropped. Person-item maps showed ASES to be superior to PROMIS PF-CAT psychometrically, with sequential and improved coverage of the latent dimension of shoulder disability.

Conclusion: Despite moderate correlation with ASES, PROMIS PF-CAT demonstrated inferior validity and psychometric properties in patients undergoing TSA. PF-CAT should not replace the ASES in this population of patients.

Level of Evidence: Basic Science Study; Validation of Outcome Instruments

© 2019 Journal of Shoulder and Elbow Surgery Board of Trustees. All rights reserved.

Keywords: PROMIS; computer adaptive test; ASES; arthroplasty; patient-reported outcomes; physical function; osteoarthritis; shoulder

Approval was received from the Hospital for Special Surgery Institutional Review Board (HSS 2013-014).

*Reprint requests: Michael C. Fu, MD, MHS, Hospital for Special Surgery, 535 E 70th St, New York, NY 10021, USA.
E-mail address: fum@hss.edu (M.C. Fu).

A number of outcomes instruments are used in the shoulder to assess patient function, quality of life, disability, satisfaction, and efficacy of interventions. These instruments range from objective physician measurements to subjective patient-reported outcome measures (PROMs). In addition, outcomes measures can include overall health and physical function assessments as well as region- or condition-specific instruments.^{20,21} One of the most commonly reported measures specific to the shoulder is the American Shoulder and Elbow Surgeons (ASES) score, developed in 1994 by the ASES Research Committee.¹⁵ With psychometric studies, the ASES score has been shown to have acceptable reliability, responsiveness, and validity for glenohumeral osteoarthritis, instability, and rotator cuff disease.^{10,11}

To address a number of barriers against the adoption and use of PROMs, such as administrative burden, reliability, and limited scope of most outcomes instruments, the Patient-Reported Outcomes Measurement Information System (PROMIS) was developed with funding from the National Institutes of Health.⁶ A number of PROMIS instruments are in the form of a computer adaptive test (CAT), which tailors question delivery to the patient according to the answers to previous items. As a measure of overall physical function, the PROMIS physical function CAT (PF-CAT) has been compared against the ASES score for various shoulder conditions, with most studies reporting moderate to excellent correlations and thus concluding that PROMIS PF-CAT may be an acceptable alternative to the ASES.^{2,5,9,13}

In addition to simple correlations between PROMs, however, it is important to determine the validity and psychometric properties of the instrument for a specific condition before widespread adoption. This enables a more comprehensive determination of an instrument's ability to measure the latent characteristic or dimension, in this case, shoulder function and disability. To date, psychometric testing of the PROMIS PF-CAT relative to the ASES score for patients with glenohumeral osteoarthritis has not been reported. Therefore, in a cohort of patients with primary glenohumeral osteoarthritis indicated for anatomic total shoulder arthroplasty (TSA), the purpose of this study was to determine the psychometric properties of the PROMIS PF-CAT with respect to the ASES. We hypothesized that the psychometric properties of the PF-CAT would be inferior to those of ASES in this population of patients.

Materials and methods

Patient cohort

The shoulder arthroplasty research registry at our orthopedic specialty institution began collecting PROMIS PF-CAT in February 2016. With the increasing interest in and adoption of

PROMIS in clinical outcomes research at the time, the decision was made to add the PF-CAT to the registry data collection to specifically compare its performance to our registry's legacy shoulder instrument, the ASES score. Given concerns about responder fatigue and instrument redundancy, the ASES is used in our registry over other shoulder-specific outcomes, given its proven psychometric properties^{10,11} and its inclusion on the list of Instruments for Collection of Orthopaedic Quality Data as recommended by the American Academy of Orthopaedic Surgeons.¹ A total of 277 cases of TSA were consented to be in the registry between February 2016 and April 2017, which is when data analysis for this study began. The inclusion criteria for this study included patients with a diagnosis of osteoarthritis, patients having complete preoperative scores for ASES and PF-CAT, and patients undergoing primary anatomic TSA. After application of these criteria, 184 TSA cases in 179 patients were included in the study cohort.

Statistical analysis

The validity, reliability, and efficiency of the PROMIS PF-CAT were compared with preoperative ASES scores. First, floor and ceiling effects were calculated for the overall ASES and the PROMIS PF-CAT, using the frequency and percentage of patients who had the lowest and highest score reported, respectively. Convergent validity between the instruments was calculated through Pearson correlation coefficients.

To evaluate the psychometric properties of PROMIS PF-CAT and ASES, Rasch partial credit models were fitted using the Extended Rasch Modeling eRm package in R (R version 3.3.1 2016-06-21; R Foundation for Statistical Computing, Vienna, Austria). The partial credit model is an item response model that evaluates whether ordinal variables, such as responses to a question for a patient-reported outcome, reflect a latent variable of interest, such as shoulder function or disability. Because the partial credit model requires scores to have polytomous item responses (ordinal responses on a 5-point Likert scale for each item/question, for example), items with only 1 response, missing items, or only 1 answer at a different response level were excluded from the partial credit model. Rasch modeling is well established in the orthopedic and shoulder literature.^{4,5,8}

Rasch partial credit model outputs of item-fit and person-fit statistics include "infit" and "outfit" measures, which indicate how accurately the model performs psychometrically. For PROMIS PF-CAT and ASES activity of daily living (ADL) questions, item fit statistics were iteratively evaluated by sequentially dropping poorly fitting questions, defined as the outfit mean square statistic <0.80 (underfit) or the infit mean square statistic >1.2 (overfit). Partial credit models were fitted iteratively until the items allowing the most coverage were left and only well-fitting items remained. This methodology has been used in the upper extremity literature as a means of item-reduction in the setting of questionnaire development to eliminate misfitting items and to minimize overlap.⁴ Person-item maps to characterize the distribution of question responses along the latent variable dimension of overall shoulder function were generated for PROMIS PF-CAT and ASES. In an ideal person-item map, responses are positioned sequentially with evenly spread intervals between each response.

Results

Preoperative responses from 179 patients (184 TSA shoulders) were included. Patients had a mean age at surgery of 65.6 ± 9.3 years; 57.5% of the cohort was male, with an average body mass index of 28.5 kg/m^2 (Table I).

The mean preoperative ASES score was 38.6 ± 18.9 , with a 1.1% floor effect and no ceiling effect. PROMIS PF-CAT had a mean preoperative score of 41.4 ± 7.4 , with no floor or ceiling effect (Table II). In terms of convergent validity, PROMIS PF-CAT and ASES demonstrated moderate correlation ($r = 0.487$; $P < .001$).

With iterative analysis of question fit and latent dimension coverage with item fit statistics from the Rasch model, all irrelevant PROMIS PF-CAT questions were dropped after 6 rounds of modeling (Table III). With the ASES ADL questions, just 1 question item was dropped (Table IV). Person-item maps for PROMIS PF-CAT (Fig. 1) and ASES (Fig. 2) showed ASES to be superior psychometrically, with sequential and improved coverage of the latent dimension of shoulder disability. Again, in an ideal person-item map, the individual items in the instrument should cover various domains along the latent dimension of shoulder disability, as demonstrated by the points on the ASES map (Fig. 2). In addition, within each item, the sequential responses (eg, very difficult, somewhat difficult, not difficult) should be progressive along the latent dimension as well, which is also the case for the ASES (Fig. 2). In contrast, the PROMIS PF-CAT person-item map is markedly inferior on both of these measures (Fig. 1).

Discussion

In this study of 179 patients with primary glenohumeral osteoarthritis undergoing TSA, despite a moderate correlation between PROMIS PF-CAT and ASES, the PROMIS PF-CAT underperformed with psychometric analysis relative to the ASES. From a clinical and practical standpoint, this psychometric underperformance is manifested in 2 ways. First, all PROMIS PF question items included in the CAT for our patients were eventually eliminated with iterative item-reduction testing in the Rasch model, whereas just 1 ASES ADL question (ability to reach a high shelf) was dropped. This iterative item-reduction process is typically used in questionnaire development to eliminate redundant items to identify the questions that best cover the various dimensions of the overall variable that the instrument is trying to measure, which in this case is shoulder function or disability. In other words, in this specific population of patients, the PROMIS PF-CAT question items were redundant with one another, and no specific items stood out as being particularly good at distinguishing patients who have significant shoulder disability from those

Table I Demographic and clinical characteristics of study population

	N	%	Mean \pm SD
Total number of patients	179		
Total number of shoulders	184		
Age (yr)			65.6 ± 9.3
Sex			
Female	76	42.5	
Male	103	57.5	
Body mass index (kg/m^2)			28.5 ± 5.4
Laterality			
Right	101	54.9	
Left	83	45.1	
Smoking history	88	47.8	

SD, standard deviation.

Table II Preoperative patient-reported outcome scores

	Mean \pm SD	Minimum	Maximum	Floor effect (%)	Ceiling effect (%)
ASES	38.6 ± 18.9	0.0	84.0	1.1	0.0
PROMIS PF-CAT	41.4 ± 7.4	24.1	58.3	0.0	0.0

SD, standard deviation; ASES, American Shoulder and Elbow Surgeons; PROMIS PF-CAT, Patient-Reported Outcomes Measurement Information System physical function computer adaptive test.

who are less disabled. Second, there was inferior coverage of the latent dimension by the PROMIS PF-CAT on the person-item map relative to the ASES. This indicates that in this population of patients, the PF-CAT questions were unable to cover a wide range of shoulder disability and that the sequential responses to the individual questions did not actually correlate with increasing or decreasing shoulder disability.

First established in 1994 as a shoulder-specific outcome instrument, the original ASES score questionnaire included both a physician-determined objective component and a patient-reported assessment.¹⁵ In the current shoulder clinical outcomes literature, the objective component is rarely reported, and the ASES score is determined by the patient’s pain on a visual analogue scale and the patient’s ability to perform 10 ADLs.²⁰ Given its use for >2 decades, it has been thoroughly validated and characterized in terms of normative values,¹⁶ minimal clinically important difference,^{11,17,18} and patient acceptable symptom state.⁷ It has also been rigorously studied psychometrically in terms of construct validity, responsiveness, and reliability.^{3,10,11,20,21} In a cohort of 137 patients with glenohumeral osteoarthritis, Kocher et al¹⁰ found no floor or ceiling effects with the ASES, which is in line with the findings of this study

Table III PROMIS PF-CAT questions included in Rasch analysis, with questions removed in sequential rounds of analyses on the basis of poor infit and outfit statistics

Number	Question	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6
3	Does your health now limit you in doing heavy work around the house like scrubbing floors or lifting or moving heavy furniture?	●	●	●	●	○	○
9	Are you able to do chores such as vacuuming or yard work?	●	●	●	●	●	○
17	Are you able to run or jog for 2 miles (3 km)?	●	○	○	○	○	○
22	Are you able to do yard work like raking leaves, weeding, or pushing a lawn mower?	●	●	●	●	●	○
37	Are you able to carry a laundry basket up a flight of stairs?	●	●	●	○	○	○
47	Are you able to run errands and shop?	●	●	○	○	○	○
53	Does your health now limit you in hiking a couple of miles (3 km) on uneven surfaces, including hills?	●	○	○	○	○	○
54	Does your health now limit you in doing strenuous activities such as backpacking, skiing, playing tennis, bicycling, or jogging?	●	●	●	●	○	○
60	Are you able to carry a shopping bag or briefcase?	●	●	●	●	●	○
106	Does your health now limit you in walking more than a mile (1.6 km)?	●	○	○	○	○	○
108	Are you able to walk at a normal speed?	●	●	○	○	○	○

PROMIS PF-CAT, Patient-Reported Outcomes Measurement Information System physical function computer adaptive test.

Patients select the answer that indicates their ability to do each activity: 1, unable to do; 2, with much difficulty; 3, with some difficulty; 4, with a little difficulty; 5, without any difficulty.

● indicates that the question was included in the next iteration of analysis, with infit and outfit >0.8 (underfit or redundant) or <1.2 (overfit).

○ indicates items that were removed because of infit or outfit <0.8 or >1.2 .

Table IV ASES function questions included in Rasch analysis, with questions removed in sequential rounds of analyses on the basis of poor infit and outfit statistics

Number	Question	Round 1	Round 2
1	Put on a coat	●	●
2	Sleep on your painful or affected side	●	●
3	Wash back/do up bra in back	●	●
4	Manage toileting	●	●
5	Comb hair	●	●
6	Reach a high shelf	●	○
7	Lift 10 lb above the shoulder	●	●
8	Throw a ball overhand	●	●
9	Do usual work	●	●
10	Do usual sport	●	●

ASES, American Shoulder and Elbow Surgeons.

Patients select the answer that indicates their ability to do each activity: 0, unable to do; 1, very difficult to do; 2, somewhat difficult; 3, not difficult.

● indicates that the question was included in the next iteration of analysis, with infit and outfit >0.8 (underfit or redundant) and <1.2 (overfit).

○ indicates items that were removed because of infit or outfit <0.8 or >1.2 .

overweighting of pain relative to ADLs, and the applicability of the high-level ADL items related to sport and overhand throwing in lower demand patients.

In comparison, the PROMIS PF-CAT has several attractive qualities for clinical outcomes research, including the fewer questions required to complete the questionnaire and the minimization of floor or ceiling effects, given the adaptive nature of the test. In studies examining the PROMIS PF-CAT for rotator cuff disease, glenohumeral osteoarthritis, proximal humerus fractures, and shoulder instability, the number of questions required to complete the PF-CAT ranged from 4 to 4.6 questions compared with 11 for the ASES.^{2,5,9,13} No floor or ceiling effects have been reported for the PF-CAT in patients undergoing TSA or patients with shoulder instability,^{2,9} whereas in patients with rotator cuff disease, Beckmann et al⁵ reported a floor effect of 3.2% and a ceiling effect of 0.53%.

Moderate correlation between PROMIS PF-CAT and ASES was found in our cohort, with a Pearson correlation coefficient of 0.487. Previously reported correlations between PROMIS PF-CAT and ASES have ranged from 0.581 for rotator cuff disease⁵ to 0.67 for shoulder instability.² In the only other study to date comparing PROMIS PF-CAT and ASES in patients undergoing primary TSA, Dowdle et al⁹ reported a correlation coefficient of 0.62 in their cohort of 61 patients. Similar to this study, no floor or ceiling effects were observed for PROMIS PF-CAT. Notably, that study did not include any advanced

(1.1% floor effect, no ceiling effect). Nevertheless, the ASES does have a number of limitations, including the complexity of the score conversion formula, the relative

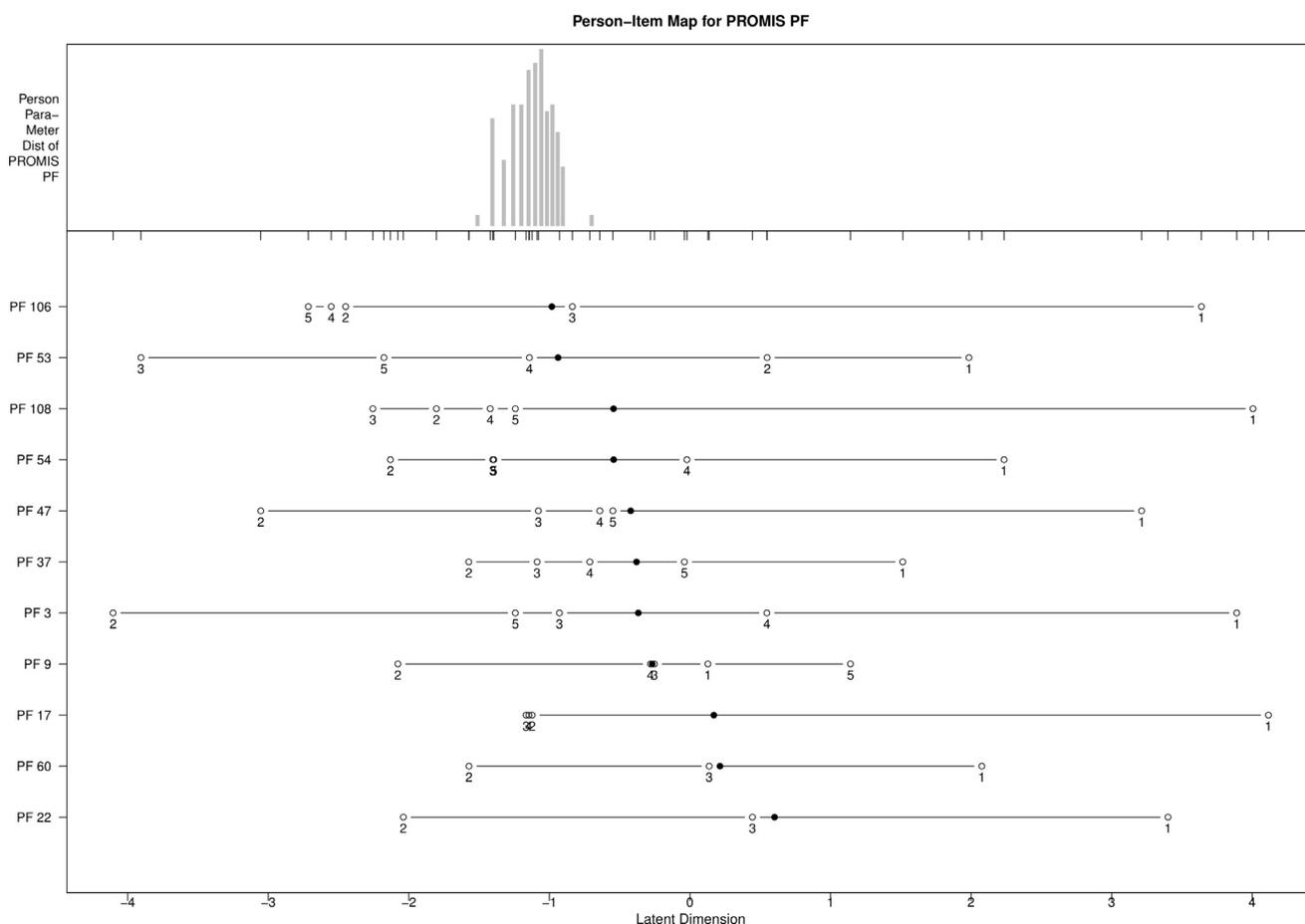


Figure 1 Person-item map for Patient-Reported Outcomes Measurement Information System (*PROMIS*) physical function (*PF*).

psychometric analyses of the validity of PF-CAT relative to ASES but nonetheless concluded that PROMIS PF-CAT may be a good alternative to traditional PROMs in a shoulder arthroplasty population. It is unclear why the correlation in our study was lower than that reported by Dowdle et al, although they did have a younger study cohort (mean age of 60.8 years) with significantly higher mean body mass index (33.9 kg/m²), suggesting a difference in study selection bias.

Interestingly, despite the moderate correlation between PF-CAT and ASES observed in our study, with psychometric analysis using the Rasch model, we found PROMIS PF-CAT to underperform psychometrically relative to ASES. Specifically, the item response partial credit model indicated that the questions included in PROMIS PF-CAT did not adequately cover the latent variable of shoulder disability. This is not surprising, given the items that were pulled by the CAT from the question bank for our TSA population (Table III), as they are mostly not specific to the upper extremity or shoulder. Therefore, despite previous studies proposing the PROMIS PF-CAT as a potential alternative to the ASES based on acceptable correlation coefficients and requiring fewer questions to complete, the

psychometric findings in this study suggest that the PF-CAT should not be used in place of the ASES score.

Despite that it is not specific to either the upper or lower extremities, PROMIS PF-CAT has also been shown to correlate well with other upper extremity PROMs, such as the Disabilities of Arm, Shoulder, and Hand (DASH),^{13,19} QuickDASH,¹⁴ Constant shoulder score,¹³ Simple Shoulder Test,⁵ and Western Ontario Shoulder Instability Index.² Nevertheless, there has been increasing use of extremity-specific PROMIS CAT instruments, such as the PROMIS upper extremity CAT (UE-CAT). In a recent study by Minoughan et al¹² comparing PROMIS UE-CAT and ASES in 52 patients with shoulder osteoarthritis, the authors found a moderate correlation coefficient of 0.57. Similar to the PF-CAT, the authors found that the UE-CAT took significantly less time to complete than the ASES (62.6 seconds vs. 160.6 seconds). As discussed before, despite acceptable correlation coefficients between the PROMs, it is important to rigorously assess the psychometric validity of any instrument before widespread use or replacement of existing validated outcomes such as the ASES. In terms of future research in our registry, we have since replaced the PROMIS PF-CAT with the PROMIS UE-CAT and will

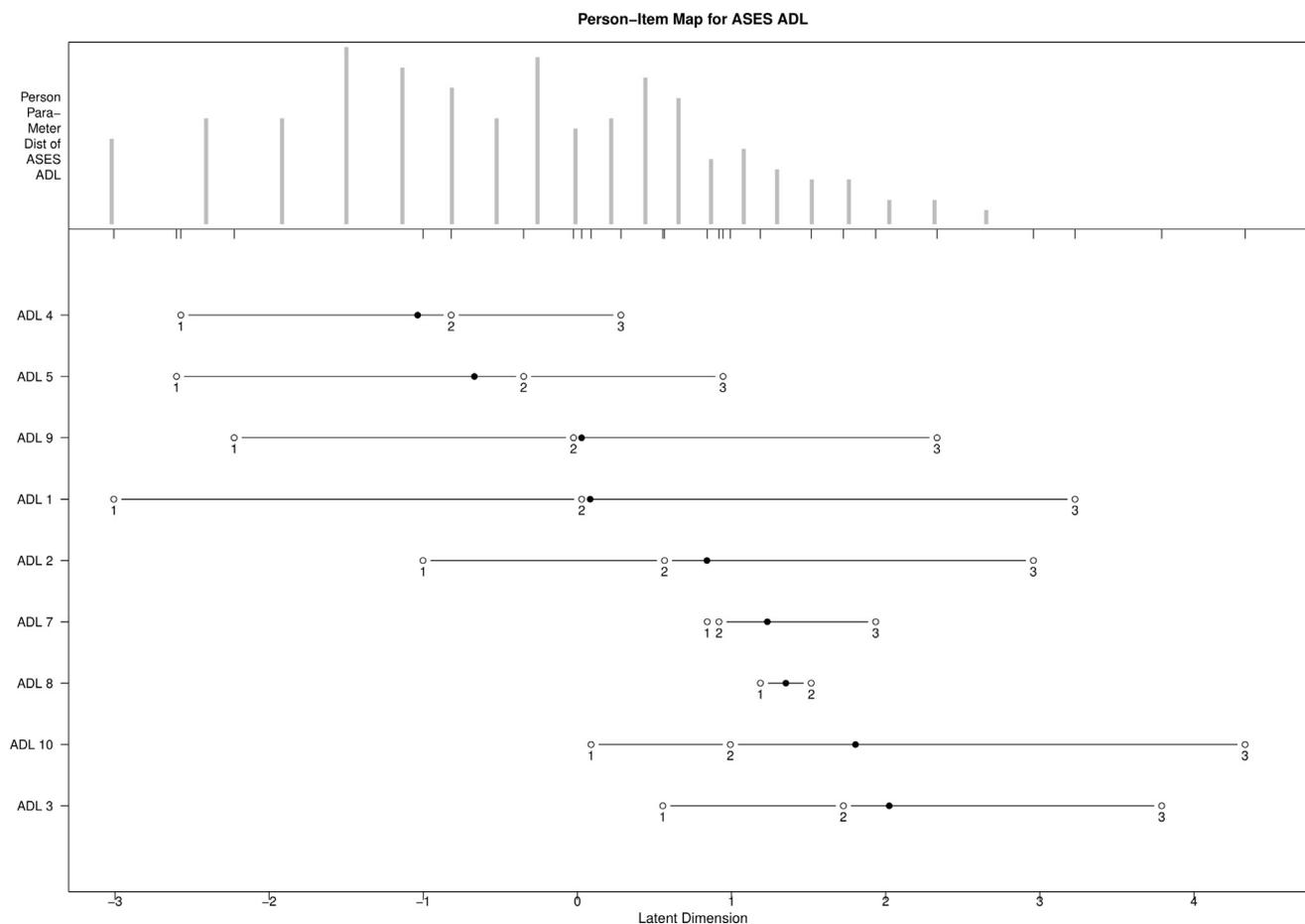


Figure 2 Person-item map for American Shoulder and Elbow Surgeons (ASES) function items. *ADL*, activity of daily living.

continue to collect the ASES to enable similar psychometric comparisons.

This study is subject to several important limitations. First, the scope of our study was limited to patients with primary glenohumeral osteoarthritis undergoing anatomic TSA, and the results may not be valid for patients suffering from other shoulder conditions. In addition, as a single-specialty tertiary referral center, there is potential selection bias in our study population. Furthermore, there is potential for questionnaire fatigue as a result of completing multiple PROMs, which may affect answer responses. Finally, as a cross-sectional study capturing only a single preoperative time point, we were unable to assess the responsiveness of the PROMs to TSA.

Conclusion

Despite moderate correlation with the ASES, PROMIS PF-CAT demonstrated inferior validity and psychometric properties in a population of patients undergoing

primary TSA. PROMIS PF-CAT should not replace the ASES in this population of patients.

Disclaimer

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

References

1. American Academy of Orthopaedic Surgeons. Patient reported outcome measures—instruments for collection of orthopaedic quality data. <https://www.aaos.org/CustomTemplates/landingPage.aspx?id=4294968282&ssopc=1>; 2016. Accessed February 5, 2019
2. 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;45:449-53. <https://doi.org/10.1177/0363546516668304>
3. Beaton D, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. *J Shoulder Elbow Surg* 1998;7:565-72.
 4. Beaton DE, Wright JG, Katz JN, Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am* 2005;87:1038-46. <https://doi.org/10.2106/JBJS.D.02060>
 5. 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;24:1961-7. <https://doi.org/10.1016/j.jse.2015.06.025>
 6. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, 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(Suppl 1):S3-11. <https://doi.org/10.1097/01.mlr.0000258615.42478.55>
 7. Chamberlain AM, Hung M, Chen W, Keener JD, McAllister J, Ebersole G, et al. Determining the patient acceptable symptomatic state for the ASES, SST, and VAS pain after total shoulder arthroplasty. *J Shoulder Elbow Arthroplasty* 2017;1. 247154921772004. <https://doi.org/10.1177/2471549217720042>.
 8. Conaghan PG, Emerton M, Tennant A. Internal construct validity of the Oxford knee scale: evidence from Rasch measurement. *Arthritis Care Res* 2007;57:1363-7. <https://doi.org/10.1002/art.23091>
 9. Dowdle SB, Glass N, Anthony CA, Hettrich CM. Use of PROMIS for patients undergoing primary total shoulder arthroplasty. *Orthop J Sport Med* 2017;5:1-4. <https://doi.org/10.1177/2325967117726044>
 10. 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;87:2006-11. <https://doi.org/10.2106/JBJS.C.01624>
 11. 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;11:587-94. <https://doi.org/10.1067/mse.2002.127096>
 12. Minoughan CE, Schumaier AP, Fritch JL, Grawe BM. Correlation of PROMIS Physical Function Upper Extremity Computer Adaptive Test with American Shoulder and Elbow Surgeons shoulder assessment form and Simple Shoulder Test in patients with shoulder arthritis. *J Shoulder Elbow Surg* 2018;27:585-91. <https://doi.org/10.1016/j.jse.2017.10.036>
 13. 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;29:257-63. <https://doi.org/10.1097/BOT.0000000000000280>
 14. 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;473:311-7. <https://doi.org/10.1007/s11999-014-3840-2>
 15. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. *J Shoulder Elbow Surg* 1994;3:347-52. [https://doi.org/10.1016/s1058-2746\(09\)80019-0](https://doi.org/10.1016/s1058-2746(09)80019-0)
 16. Sallay PI, Reed L. The measurement of normative American Shoulder and Elbow Surgeons scores. *J Shoulder Elbow Surg* 2003;12:622-7. [https://doi.org/10.1016/S1058-2746\(03\)00209-X](https://doi.org/10.1016/S1058-2746(03)00209-X)
 17. Tashjian RZ, Deloach J, Green A, Porucznik CA, Powell AP. Minimal clinically important differences in ASES and Simple Shoulder Test scores after nonoperative treatment of rotator cuff disease. *J Bone Joint Surg Am* 2010;92:296-303. <https://doi.org/10.2106/JBJS.H.01296>
 18. Tashjian RZ, Hung M, Keener JD, Bowen RC, McAllister J, Chen W, et al. Determining the minimal clinically important difference for the American Shoulder and Elbow Surgeons score, Simple Shoulder Test, and visual analog scale (VAS) measuring pain after shoulder arthroplasty. *J Shoulder Elbow Surg* 2017;26:144-8. <https://doi.org/10.1016/j.jse.2016.06.007>
 19. Tyser AR, Beckmann J, Franklin JD, Cheng C, Hon SD, Wang A, et al. Evaluation of the PROMIS physical function computer adaptive test in the upper extremity. *J Hand Surg Am* 2014;39:2047-51. <https://doi.org/10.1016/j.jhsa.2014.06.130>
 20. Wright RW, Baumgarten KM. Shoulder outcomes measures. *J Am Acad Orthop Surg* 2010;18:436-44. <https://doi.org/10.5435/00124635-201007000-00006>
 21. Wylie JD, Beckmann JT, Granger E, Tashjian RZ. Functional outcomes assessment in shoulder surgery. *World J Orthop* 2014;5:623-33. <https://doi.org/10.5312/wjo.v5.i5.623>