



BASIC SCIENCE

Psychometric evaluation of the Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function and Pain Interference Computer Adaptive Test for subacromial impingement syndrome



Benjamin Strong, MD^a, Michael Maloney, MD^a, Judith Baumhauer, MD^a, Joseph Schaffer, MD^a, Jeff R. Houck, PT, PhD^a, Man Hung, PhD^b, Jerry Bounsanga, BS^b, Maren W. Voss, ScD^b, Yushan Gu, BS^b, Ilya Voloshin, MD^{a,*}

^aDepartment of Orthopaedics, University of Rochester Medical Center, Rochester, NY, USA

^bDepartment of Orthopaedics, University of Utah, Salt Lake City, UT, USA

Background: The Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function Computer Adaptive Test (CAT) was previously validated for rotator cuff disease and shoulder instability. This study evaluated the psychometric properties of the PROMIS Physical Function (PF) CAT, PROMIS Pain Interference (PI) CAT, and the American Shoulder and Elbow Surgeons (ASES) Shoulder Function Score for subacromial impingement syndrome.

Methods: PROMIS PF CAT, PI CAT, and ASES (Pain, Function, Total) were collected on all visits for 2 surgeons between January 2016 and August 2016. New patients, aged 18 years and older, were selected by International Classification of Diseases code for impingement syndrome of the shoulder. The mean number of questions answered determined efficiency. Person-item maps were created to determine ceiling and floor effects as well as person reliability. Convergent validity was determined by comparison of PROMIS domains to ASES scores with Pearson correlations.

Results: For PROMIS PF CAT, the mean number of items answered was 4.54 (range 4–12). The ceiling effect was 1.56%, and the floor effect was 3.13%. The person reliability was 0.94. Pearson correlation coefficients between the PF CAT and ASES were 0.664 (ASES Function), 0.426 (ASES Pain), and 0.649 (ASES Total). For PROMIS PI CAT, the mean number of items answered was 4.27 (range 3–11). The ceiling effect was 4.69%, and the floor effect was 8.33%. The person reliability was 0.92. Pearson correlation coefficients between the PI CAT and ASES were: 0.667 (ASES Function), 0.594 (ASES Pain), and 0.729 (ASES Total).

Conclusions: The psychometric properties of PROMIS PF and PI CATs were favorable for subacromial impingement syndrome.

The University of Rochester Research Subjects Review Board approved this study (RSRB00061949).

*Reprint requests: Ilya Voloshin, MD, University of Rochester Medical Center, 601 Elmwood Ave, Box 665, Rochester, NY 14625, USA.
E-mail address: ilya_voloshin@urmc.rochester.edu (I. Voloshin).

Level of evidence: Basic Science Study; Development or Validation of Outcomes Instruments
 © 2018 Journal of Shoulder and Elbow Surgery Board of Trustees. All rights reserved.

Keywords: PROMIS; CAT; physical function; pain interference; shoulder diagnoses; ASES

There has been a growing interest in orthopedics in the use of patient-reported outcomes (PROs). PROs give patients a voice in their care and are designed to measure what is truly meaningful to them. There are now a great number of available PROs, including general PROs, that can be applied broadly for all patients and diseases and specialty-specific PROs that target one anatomic area or condition. The plethora of available PROs can make choosing which outcome measures to use a difficult task.¹² It is imperative that PROs undergo critical evaluation and scrutiny before being adapted into practice. The chosen PROs must be accurate, reliable, valid, and efficient to administer.⁷

Disease-specific PROs have been developed in orthopedics for various subspecialties such as the Foot and Ankle Ability Measure and the American Shoulder and Elbow Surgeons (ASES) score.^{4,13,14} Unfortunately, disease-specific PROs may be relevant only to specific areas and are often cumbersome to administer. For this reason there is value in the development and use of generalizable PROs.

In an effort to establish efficient, valid, and generalizable PROs, the National Institutes of Health developed the Patient-Reported Outcomes Measurement Information System (PROMIS). PROMIS instruments have the ability to use computerized adaptive testing (CAT), which allows efficient administration while maintaining the power and sensitivity of a large question bank. PROMIS instruments cover many independent and unique domains; however, the most popularly applied domains are the Physical Function CAT (PF CAT), Pain Interference CAT (PI CAT), and Depression CAT.^{5,6,16} The PI CAT is of particular interest; which, in contrast to a simple pain scale, assesses how a patient's pain experience affects their overall function.¹ The PI CAT has also been correlated to patients' coping strategies.¹¹

Up to this point, psychometric evaluation and validation of PROMIS instruments have been performed in orthopedic specialties of foot and ankle, trauma, spine, upper extremity, and shoulder. Across the board, PROMIS instruments have routinely matched or outperformed current PROs in terms of reliability and efficiency.^{4,8-10,17} The PROMIS PF CAT has previously been compared to and evaluated against the ASES for rotator cuff disease^{2,4} and shoulder instability.³ This study evaluated the psychometric properties of the PROMIS PF CAT and PROMIS PI CAT for subacromial impingement syndrome compared with the ASES Function subscore.

Materials and methods

Demographic data, along with PROMIS PF CAT and PI CAT, were collected with the use of tablet computers for all clinic visits

for 2 shoulder surgeons between January 2016 and August 2016. Written forms of the ASES Pain subscore, Function subscore, and Total score were collected and manually entered into the institutional database. Patients were selected by International Classification of Diseases 9 code 762.2 and International Classification of Diseases 10 code M75.4 for impingement syndrome of the shoulder. Data were included for all new patients aged older than 18 years who had completed both the PROMIS CATs and ASES. Psychometric analysis was performed by trained statisticians and in accordance with previously established literature standards.^{4,8-10} The respondent burden was determined by the mean number of questions answered and was considered a measure of efficiency.

Person-item maps were created to visualize item hierarchy, shown in Fig. 1. Ceiling and floor effects were calculated as the proportion of patients who achieved the maximal score and minimal score, respectively. Person reliability, or the consistency with which the test is able to rank patients by ability, was determined by the Pearson product-moment coefficient, where $r > 0.8$ was considered good and $r > 0.9$ was considered excellent.⁸ Convergent validity, or the degree to which the test correlates with a previously established validated measure, was analyzed by comparison of PROMIS instruments to the ASES with Pearson correlations. Correlations of $r > 0.7$ were considered strong, $r > 0.5$ moderate, and $r < 0.5$ poor.⁸

Results

We identified 192 patients (102 men, 90 women). Their mean age was 55.8 years, and the dominant extremity was affected in 93.6% of respondents (Table I).

Table I Patient demographics

Variable	Patient data
Age, yr	
Mean	54.3
Median	56.0
Minimum-maximum	18.0-91.0
Sex, No. (%)	
Male	213 (56.2)
Female	166 (43.8)
Hand dominance, No. (%)	
Right	300 (88.8)
Left	38 (11.2)
Affected shoulder, No. (%)	
Right	198 (57.2)
Left	148 (42.8)
Side affected, No. (%)	
Dominant	184 (54.8)
Nondominant	152 (45.2)

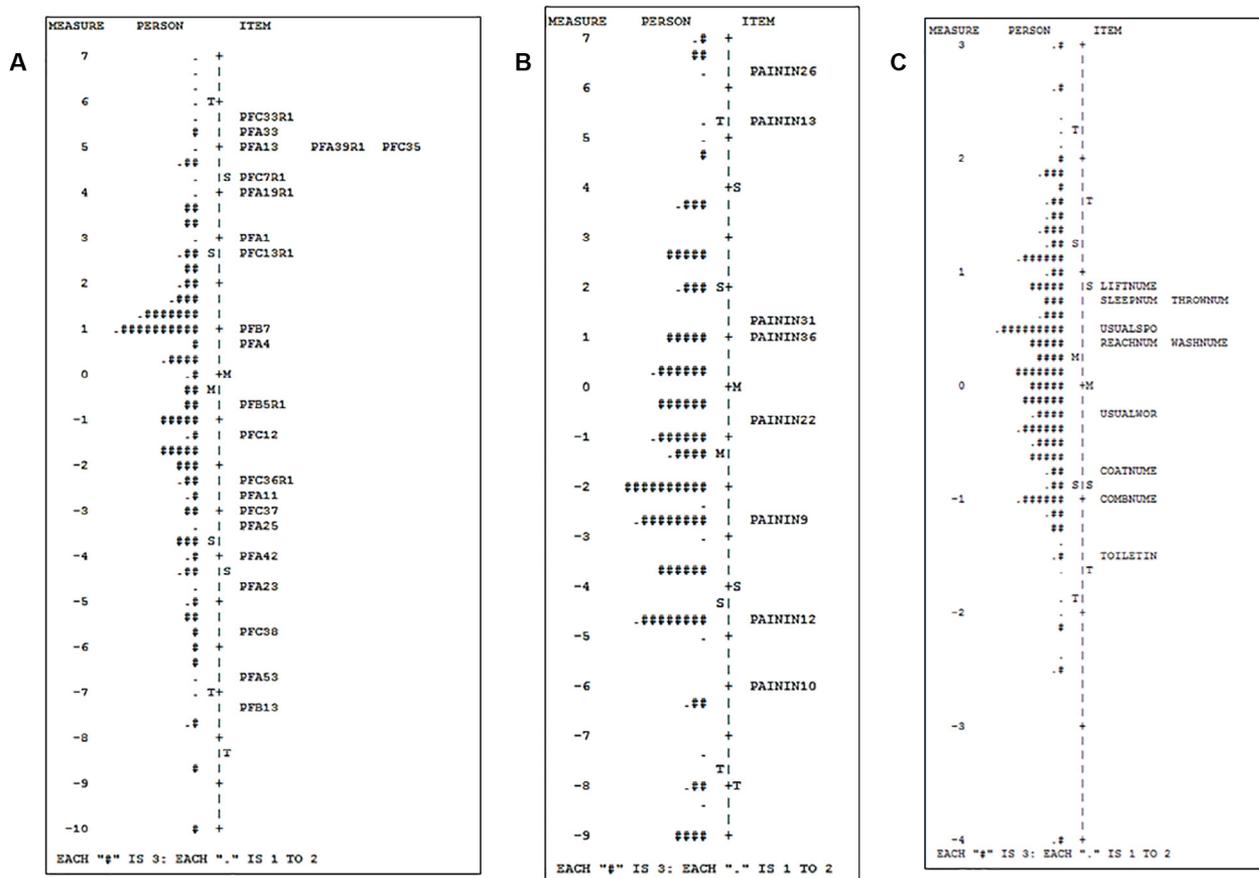


Figure 1 Person-item maps demonstrates item hierarchy. Patient Reported Outcomes Measurement Information System (PROMIS) (A) Physical Function (PF) and (B) Pain Interference (PAININ) computer adaptive test. (C) American Shoulder and Elbow Surgeons Function.

Efficiency

The mean (range) number of questions answered was 4.54 (4-12) for PROMIS PF CAT, 4.27 (3-11) for PROMIS PI CAT, and 9.40 (5-10) for ASES Function.

Ceiling and floor effects

The ceiling effects were 1.56% for PROMIS PF CAT, 4.69% for PROMIS PI CAT, and 22.40% for ASES Function. The floor effects were 3.13% for PROMIS PF CAT, 8.33% for PROMIS PI CAT, and 5.21% for ASES Function.

Person reliability

Person reliability was 0.94 (excellent) for PROMIS PF CAT, 0.92 (excellent) for PROMIS PI CAT, and 0.86 (good) for ASES Function.

Convergent validity

PROMIS PF CAT had moderate correlation to ASES Function (0.664), poor correlation to ASES Pain (0.426), and moderate correlation to ASES Total (0.649), shown in Fig. 2.

PI CAT had moderate correlation to ASES Function (0.667), moderate correlation to ASES Pain (0.594), and strong correlation to ASES Total (0.729), shown in Fig. 3. ASES Function had moderate correlation to ASES Pain (0.513).

The psychometric analysis of PI CAT and ASES Function is summarized in Table II.

Discussion

The PROMIS PF CAT demonstrated high efficiency, minimal ceiling and floor effects, and excellent person reliability. It had moderate correlation to ASES Function and ASES Total, but poor correlation to ASES Pain.

The PROMIS PI CAT demonstrated high efficiency, minimal ceiling and floor effects, and excellent person reliability. It had moderate correlation to ASES Function and ASES Pain and strong correlation to ASES Total. The greater correlation of the PI CAT to ASES Total compared with ASES Function and ASES Pain subscores alone highlights that PI is not just a measure of pain but more so describes the degree to which pain affects function. Overall, PROMIS PI CAT had excellent psychometric performance and the strongest correlation to ASES Total score of all domains tested. This is consistent with the reported results from

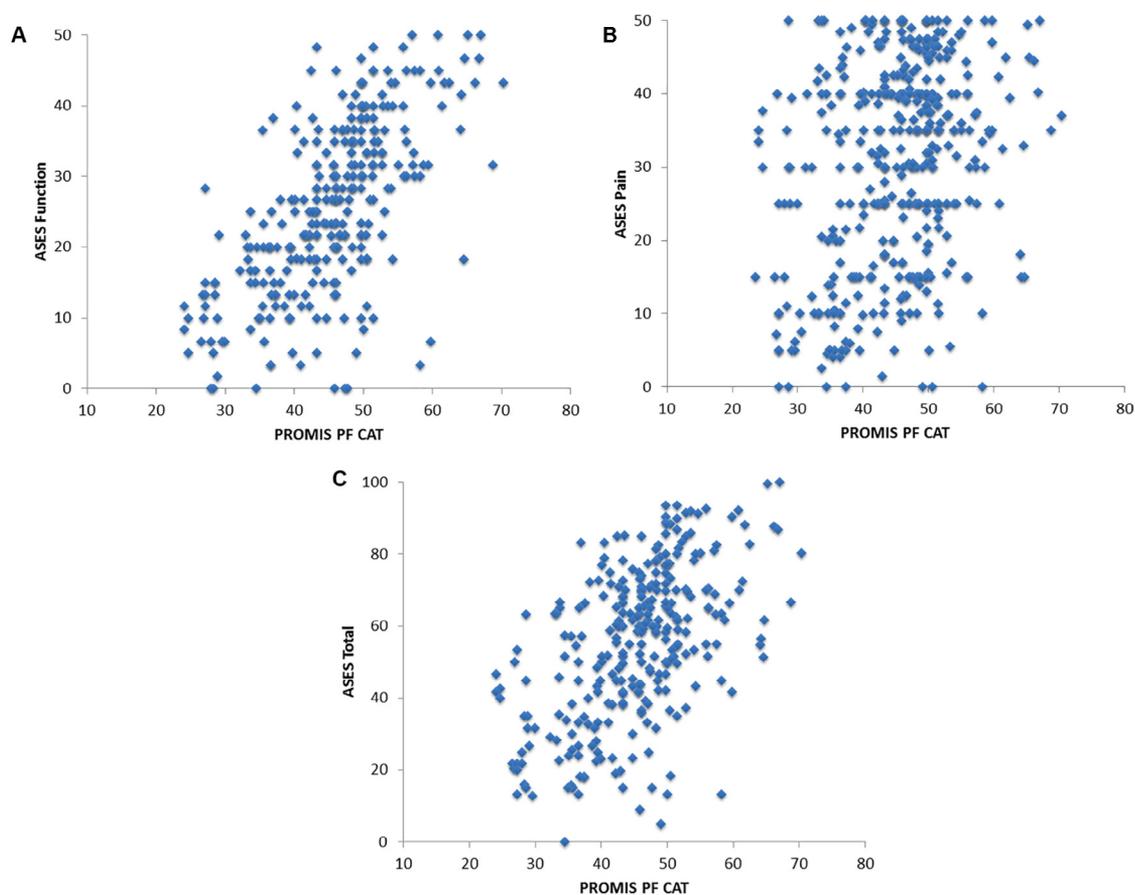


Figure 2 Pearson correlation comparison of Patient Reported Outcomes Measurement Information System (*PROMIS*) Physical Function (*PF*) computer adaptive test (*CAT*) vs. (A) American Shoulder and Elbow Surgeons (*ASES*) Function, (B) *ASES* Pain, and (C) *ASES* Total.

Overbeek et al¹⁵ in which they compared *PROMIS* domains to the 11-item version of the Disabilities of the Arm, Shoulder and Hand (*QuickDASH*) scores for upper extremity disease. They found moderate correlation between *PROMIS* PF and *QuickDASH* scores and strong correlation between *PROMIS* PI and *QuickDASH* scores.¹⁵ The psychometric performance of the PI CAT was also consistent to the prior analysis of Beckman et al⁴ comparing the PF CAT to SST and *ASES* for rotator cuff disease as well as the validation by Anthony et al³ validation of *PROMIS* PF CAT for shoulder instability. This gives confidence that the *PROMIS* PI CAT can be used in addition to the already established *PROMIS* PF CAT.

ASES Function had the greatest respondent burden of all domains tested (9.40 questions). Furthermore, the *ASES* Total score is calculated by adding the single-item *ASES* Pain subscore to the *ASES* Function subscore. The *ASES* also has 6 questions that are a part of the questionnaire but not used in the scoring. The true number of questions answered for a fully completed *ASES* questionnaire is 17.¹⁸ *ASES* Function had the greatest ceiling effect of all tests (22.40%) but had relatively minimal floor effect (5.21%). *ASES* Function had good person reliability (0.86) whereas *PROMIS* PF CAT (0.94) and PI CAT (0.92) both had excellent person reliabil-

ity. This is likely because *ASES* Function questionnaire is a set length of 10 items, which greatly limits its discriminating power. In comparison, the PF CAT is drawn from a 165-item bank, and the PI CAT is drawn from a 40-item bank (question banks: <http://www.healthmeasures.net/search-view-measures>; see Appendices S1 and S2). *ASES* Function had poor correlation to *ASES* Pain.

The psychometric properties of the *ASES* Pain subscore and *ASES* Total score were not specifically evaluated because the *ASES* Pain subscore is derived entirely from the single-item visual analog scale for pain. The *ASES* Total is then calculated as the addition of the *ASES* Function subscore and *ASES* Pain subscore, which makes 50% of the *ASES* Total score dependent on the response to the visual analog scale for pain. Due to the predominant composition of the *ASES* Pain subscore and *ASES* Total on a single item, we felt that separate psychometric analysis was not appropriate.

In all measures of psychometric performance for subacromial impingement syndrome, *PROMIS* PF CAT and PI CAT matched or outperformed the *ASES*. Both tests demonstrate excellent overall psychometric performance but are designed to describe 2 independent domains of the patients' experience. This was well described by Hung et al⁸ in their validation of *PROMIS* PF and PI CATs for foot and ankle

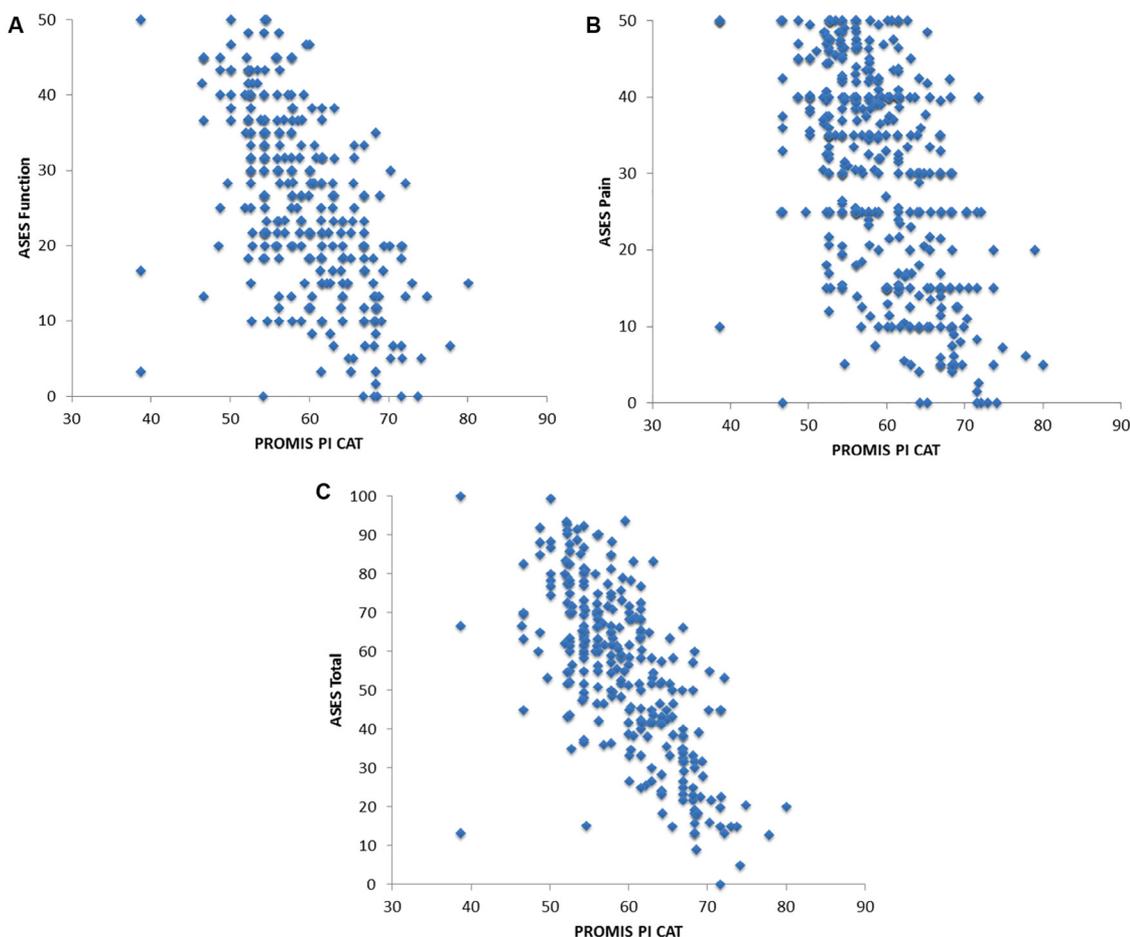


Figure 3 Pearson correlation comparison of Patient Reported Outcomes Measurement Information System (*PROMIS*) Pain Interference (*PI*) computer adaptive test (*CAT*) vs. (A) American Shoulder And Elbow Surgeons (*ASES*) Function, (B) *ASES* Pain, and (C) *ASES* Total.

patients. They performed Rasch modeling, which established the independent, univariate nature of the *PROMIS* PF and *PI* *CAT*s.⁸ In contrast, in their comparison of *PROMIS* PF *CAT* to *ASES* Total, Beckman et al⁴ determined *ASES* Total has a high degree of unexplained variance, meaning that it measures multiple domains. Because of these differences, if *PROMIS* *CAT*s are chosen as a substitute for the *ASES*,

we recommend use of *PROMIS* PF and *PI* *CAT* together to more completely describe the patients’ true experience.

This study has several limitations. Only new patient visits were included, and the data were analyzed at a single time point. Because of this, we could not determine the responsiveness of the *PRO*s. In addition, only patients with a diagnosis of impingement syndrome were

Variable	Efficiency: questions answered Mean (range), No.	Ceiling effect (%)	Floor effect (%)	Person reliability (<i>r</i>)	<i>ASES</i> Function correlation (<i>r</i>)	<i>ASES</i> Pain correlation (<i>r</i>)	<i>ASES</i> Total correlation (<i>r</i>)
<i>PROMIS</i>							
PF <i>CAT</i>	4.54 (4-12)	1.56	3.13	0.9	0.664	0.426	0.649
<i>PI</i> <i>CAT</i>	4.27 (3-11)	4.69	8.33	0.92	0.667	0.594	0.729
<i>ASES</i> Function	9.40 (5-10)	22.40	5.21	0.86	1.00	0.513	0.845

PROMIS, Patient Reported Outcomes Measurement Information System; *PF*, physical function; *CAT*, computer adaptive test; *PI*, pain interference; *ASES*, American Shoulder and Elbow Surgeons.

included. We believe the results would continue to hold across shoulder diagnoses; however, further analysis including additional diagnoses would be required to confirm this.

Conclusion

PROMIS PF CAT and PI CAT demonstrated favorable psychometric properties in efficiency, reliability, and validity for subacromial impingement syndrome. The PROMIS PI CAT had stronger correlation to the ASES Total score than the PROMIS PF CAT.

Disclaimer

Ilya Voloshin is a paid consultant for and receives speaking fees from Zimmer, Arthrex, ArthroSurface, and Smith & Nephew, and receives royalties from Innomed. The other 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.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jse.2018.07.024>.

References

1. Amtmann D, Cook KF, Jensen MP, Chen WH, Choi S, Revicki D, et al. Development of a PROMIS item bank to measure pain interference. *Pain* 2010;150:173-82. <http://dx.doi.org/10.1016/j.pain.2010.04.025>
2. Anthony CA, Glass N, Hancock K, Bollier M, Hettrich CM, Wolf BR. Preoperative performance of the patient-reported outcomes measurement information system in patients with rotator cuff pathology. *Arthroscopy* 2017;33:1770-74.e1. <http://dx.doi.org/10.1016/j.arthro.2017.04.018>
3. 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. <http://dx.doi.org/10.1177/0363546516668304>
4. Beckman 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. <http://dx.doi.org/10.1016/j.jse.2015.06.025>
5. Cappelleri JC, Lundy JJ, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther* 2014;36:648-62. <http://dx.doi.org/10.1016/j.clinthera.2014.04.006>
6. Cook KF, O'Malley KJ, Roddey TS. Dynamic assessment of health outcomes: time to let the CAT out of the bag? *Health Serv Res* 2005;40:1694-711. <http://dx.doi.org/10.1111/j.1475-6773.2005.00446.x>
7. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: a new era in clinical research. *Perspect Clin Res* 2011;2:137-44. <http://dx.doi.org/10.4103/2229-3485.86879>
8. Hung M, Baumhauer JF, Latt LD, Saltzman CL, SooHoo NF, Hunt KJ, et al. Validation of PROMIS® physical function computerized adaptive tests for orthopaedic foot and ankle outcome research. *Clin Orthop Relat Res* 2013;471:3466-74. <http://dx.doi.org/10.1007/s11999-013-3097-1>
9. Hung M, Hon SD, Franklin JD, Kendall RW, Lawrence BD, Neese A, et al. Psychometric properties of the PROMIS physical function item bank in patients with spinal disorders. *Spine* 2014;39:158-63. <http://dx.doi.org/10.1097/BRS.000000000000097>
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 with the short musculoskeletal function assessment in orthopaedic trauma patients. *J Orthop Trauma* 2014;28:439-43. <http://dx.doi.org/10.1097/BOT.0000000000000059>
11. Kortlever JT, Janssen SJ, van Berckel MM, Ring D, Vranceanu AM. What is the most useful questionnaire for measurement of coping strategies in response to nociception? *Clin Orthop Relat Res* 2015;473:3511-8. <http://dx.doi.org/10.1007/s11999-015-4419-2>
12. 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;26:352-7. <http://dx.doi.org/10.1016/j.jse.2016.09.054>
13. Martin R, Irrgang J, Burdett R, Conti S, Van Swearingen J. Evidence of validity of the Foot and Ankle Ability Measure (FAAM). *Foot Ankle Int* 2005;26:968-83. <http://dx.doi.org/10.1177/107110070502601113>
14. 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. <http://dx.doi.org/10.1067/mse.2002.127096>
15. 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. <http://dx.doi.org/10.1007/s11999-014-3840-2>
16. Papuga OM, Dasilva C, McIntyre AW, Mitten D, Kates S, Baumhauer JF. Large-scale clinical implementation of PROMIS computer adaptive testing with direct incorporation into the electronic medical record. *Health Systems* 2017;1-13. <http://dx.doi.org/10.1057/s41306-016-0016-1>
17. 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. <http://dx.doi.org/10.1016/j.jhssa.2014.06.130>
18. Wylie JD, Beckmann JT, Granger E, Tashjian RZ. Functional outcomes assessment in shoulder surgery. *World J Orthop* 2014;5:623-33. <http://dx.doi.org/10.5312/wjo.v5.i5.623>