



# PROMIS: a valid and efficient outcomes instrument for patients with ACL tears

Kyle John Hancock<sup>1</sup> · Natalie Glass<sup>1</sup> · Chris A. Anthony<sup>1</sup> · Brian R. Wolf<sup>1</sup> · Carolyn M. Hettrich<sup>1</sup> · John Albright<sup>1</sup> · Matt Bollier<sup>1</sup> · Annunziato Amendola<sup>2</sup>

Received: 17 May 2017 / Accepted: 25 June 2018 / Published online: 4 July 2018  
© European Society of Sports Traumatology, Knee Surgery, Arthroscopy (ESSKA) 2018

## Abstract

**Purpose** The current study compares the Patient Reported Outcomes Information System Physical Function Computer Adaptive Test (PROMIS PF CAT) to traditional knee PRO instruments in a healthy population undergoing surgery for ACL injuries with the following objectives: (1) identify and determine the strength of any correlations between the scores of PROMIS PF CAT and current knee PROs or their subscales that measure physical function; (2) evaluate PROMIS PF CAT's test burden; and (3) determine if PROMIS PF CAT has any floor or ceiling effects in this population.

**Methods** Patients indicated for ACL surgery completed the Short Form-36 Physical Function (SF-36 PF), Knee Injury and Osteoarthritis Outcome Score (KOOS), Marx Knee Activity Rating Scale (Marx), the EuroQol 5-dimensions Questionnaire (EQ-5D), and PROMIS PF CAT. Correlations between PROs were defined as follows: High ( $\geq 0.7$ ); high-moderate (0.61–0.69); moderate (0.4–0.6); moderate-weak (0.31–0.39); and weak ( $\leq 0.3$ ). Floor or ceiling effects were considered significant if 15% or more patients reported the lowest or highest possible total score, respectively.

**Results** 100 patients participated with a mean age of 26 years (range 11–57). The PROMIS PF CAT demonstrated high correlations with SF-36 PF ( $r=0.82$ ,  $p<0.01$ ), EQ-5D ( $r=-0.70$ ,  $p<0.01$ ) KOOS ADL ( $r=0.74$ ,  $p<0.01$ ), and KOOS Sport ( $r=0.70$ ,  $p<0.01$ ). There were no ceiling or floor effects for PROMIS PF CAT (0%). The mean number of items completed for the PROMIS PF CAT was 4.2 (median 4; range 4–11).

**Conclusions** The PROMIS PF CAT shows a high correlation with commonly employed PROs that also measure physical function with low test burden and without ceiling effects in this relatively young and healthy population.

**Keywords** Patient-reported outcomes · PROMIS · ACL outcomes · Computer adaptive testing

## Introduction

The tools that orthopedic surgeons use to assess patient outcomes have changed substantially over time. In the first half of the twentieth century, surgeons typically reported their findings using anecdotal assessment or arbitrary variables [10, 15]. The next generation of instruments, such as the Harris Hip Score, included more consistent objective data that allowed for comparison between studies, but were

characteristically completed by the physician and, therefore, subjected to the physician's bias [1, 9, 15, 16, 23, 31]. The most recent generation of tools, patient-reported outcomes (PRO) instruments, directly assess the patient's own subjective interpretation of their outcome, a variable that many feel is a more ideal measure of treatment success. PRO instruments today are continually being developed and streamlined, using vigorous methodology to ensure validity and responsiveness and are increasingly employed via electronic media [8, 21, 22, 26, 28].

As part of this ongoing development, the National Institutes of Health funded the development of the Patient Reported Outcome Measurement Information System (PROMIS), a large database of questions relating to various health domains. The questions, or items, in PROMIS are individually validated and developed using Item Response Theory so that they can be used alone or in combination

✉ Kyle John Hancock  
kyle-hancock@uiowa.edu

<sup>1</sup> The Department of Orthopedic Surgery and Rehabilitation, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, 01008 JPP, Iowa City IA 52242, USA

<sup>2</sup> Division of Sports Medicine, Duke University, Durham, North Carolina, USA

with other questions within the same question bank. This contrasts to traditional fixed-length instruments that need to be completed in their entirety to be valid [3, 6, 14]. Each item also has a known relationship to other items so that the patient's response to the first item can be recorded, and from that response the most informative second item can next be determined and presented to the patient. This process is repeated after each item until a predetermined precision level is reached. In practice today, the process described is administered using computers and is known as computer adaptive testing (CAT) [2, 3, 6, 7, 12, 14]. CAT allows for the custom-tailored assessment of each patient using the fewest questions necessary while maintaining measurement precision.

Some of the advantages of PROMIS CATs that have been demonstrated to date include high-content validity and good responsiveness to change, [13] high-reliability [12], and improved efficiency [2, 12, 14]. Initial work has demonstrated that PROMIS CATs have many favorable characteristics, but they have primarily been studied in relatively lower functioning populations and the potential for ceiling effects has been suggested [6]. Currently, there is little work examining populations of higher function and the characteristics of PROMIS CAT in such groups remains unknown.

Currently available PRO instruments include generalized instruments such as the Short Form Healthy Survey (SF-36) [19, 30] or the EuroQol 5-Dimensions Questionnaire (EQ-5D) [24] which are used as overall surveys of health. There are also more focused instruments such as those specific to an anatomical site, examples of which include the Knee Injury and Osteoarthritis Outcomes Score (KOOS) [11, 25] or the Marx Knee Activity Rating Scale (Marx) [17, 18]. Both general and specific instruments have proven useful for measuring patient outcomes, but some are burdensome to administer and complete because of their high numbers of questions (e.g., greater than 30 questions for the KOOS and SF-36).

This study assesses the physical function domain of PROMIS CAT (PROMIS PF CAT) in a higher functioning cohort indicated for ACL surgery with for the following purposes: (1) to identify and quantify any correlations between the scores of PROMIS PF CAT and current knee PROs or their subscales that measure physical function; (2) to evaluate PROMIS PF CAT's test burden; and (3) to determine if PROMIS PF CAT has any floor or ceiling effects in this population.

## Materials and methods

All patients who were indicated for operative management of an ACL injury by any one of five participating surgeons at a single university outpatient clinic over a 10-month period

beginning January 2015 were approached for the study and those who consented were enrolled. Any patient undergoing significant simultaneous operations including microfracture, meniscus repair, osteotomy, or osteochondral allograft were excluded. Each patient completed the PROMIS PF CAT, EQ-5D, KOOS, Marx, and SF-36 subscales of physical function, general health and pain in random order on a computer kiosk during their preoperative clinic visit. Both generic and anatomical site-specific PRO instruments were included to evaluate for both divergent and convergent construct validity. Institutional Review Board approval was obtained for this study (University of Iowa IRB #201,609,839).

## Statistical analysis

Patient demographic data was recorded. The Shapiro–Wilk test was used to evaluate the normality of each PRO instrument's score distribution. All distributions with the exception of the KOOS Symptoms subscale were not normally distributed and thus Spearman correlation coefficients were chosen to identify associations between PROMIS PF CAT and other instruments with  $p < 0.05$ .

The strength of a correlation was categorized as: High ( $\geq 0.7$ ); high-moderate (0.61–0.69); moderate (0.4–0.6); moderate-weak (0.31–0.39); and weak ( $\leq 0.3$ ) [27]. The correlation between PROMIS PF CAT and the other instruments that either measured physical function (convergent validity) or measured some other domain (discriminant validity) was used to evaluate construct validity.

An estimated sample size of 36 for a two-sided test and alpha level of 0.05 was determined to provide 80% discriminatory power to detect a correlation of moderate (0.6) from weak (0.2).

Floor or ceiling effects were calculated as the percentage of patients who obtained the lowest or highest possible score on a given instrument, respectively, and were considered significant if they were at least 15% [29].

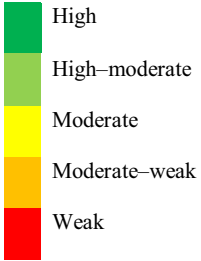
Patients completed the PROMIS PF CAT instrument when they answered enough items to reach a predefined level of significance using the default administration rules for PROMIS [14]. The number of items required for completion of the PROMIS PF CAT was recorded for each patient.

## Results

One hundred patients were enrolled (45% women; 55% men). The mean ( $\pm$  SD) age was  $26.4 \pm 9.2$  years and mean BMI was  $27.2 \pm 6.2$  kg/m<sup>2</sup>.

Table 1 illustrates correlations between PROMIS PF CAT scores and scores of the other PRO instruments. Correlations were high for each of the other PRO instruments/subscales aimed at measuring physical function, including SF-36 PF

**Table 1** Correlations between PROMIS PF CAT and other PROs

	PRO instrument	r value	Median raw score	
Phys. function PROs	PROMIS PF CAT	1.00	41.3	Correlation strength: 
	SF-36 PF	0.82 ( $p < 0.01$ )	55.0	
	KOOS Sport	0.70 ( $p < 0.01$ )	97.0	
	KOOS ADL	0.74 ( $p < 0.01$ )	99.1	
Other PROs	SF-36 GH	0.12 ( $p = 0.22$ )	80.0	
	SF-36 Pain	0.51 ( $p < 0.01$ )	55.0	
	KOOS Sym	0.54 ( $p < 0.01$ )	98.2	
	KOOS Pain	0.58 ( $p < 0.01$ )	98.6	
	KOOS QOL	0.49 ( $p < 0.01$ )	97.3	
	Marx	0.08 ( $p = 0.46$ )	14.0	
	EQ-5D	0.70 ( $p < 0.01$ )	0.76	

\*Definitions: SF-36 Physical Functioning, SF-36 General Health Perceptions, SF-36 Bodily Pain, Knee Injury and Osteoarthritis Outcome Score (KOOS) Symptoms, KOOS Daily Living, KOOS Sports and Recreational Activities, KOOS Pain, KOOS Quality of Life, Marx Knee Activity Rating Scale, Euro-Qol-5 Dimension

( $r=0.82, p < 0.01$ ), KOOS Sport ( $r=0.70, p < 0.01$ ), and KOOS ADL ( $r=0.74, p < 0.01$ ). With the exception of EQ-5D, all correlations that reached statistical significance between PROMIS PF CAT and instruments not specifically aimed at physical function were moderate. Correlation between EQ-5D and PROMIS PF CAT was high ( $r=0.70, p < 0.01$ ). PROMIS PF CAT scores and those of the MARX and SF-36 GH did not have a significant relationship.

PROMIS PF CAT demonstrated no floor or ceiling effects, with no patients attaining either the highest or lowest score (Table 2). Both the SF-36 GH and MARX demonstrated significant ceiling effects with 15.0% and 42.0% of patients scoring at the extreme high end of these instruments, respectively.

By default rules, participants answered a minimum of 4 questions for PROMIS PF CAT. Ninety-four percent of the 100 patients enrolled were only asked the minimum of 4 items (range 4–11), with the average number of questions answered being  $4.2 \pm 0.9$ .

### Discussion

The most important finding of the present study is that PROMIS PF CAT scores have a high correlation with currently used knee PRO instruments or their subscales that are designed to measure physical function including KOOS

**Table 2** PRO instrument ceiling and floor effects

PRO instrument*	Floor effects	Ceiling effects
PROMIS PF CAT	0 (0.0%)	0 (0.0%)
SF-36 PF	1 (1.0%)	5 (5.0%)
SF-36 GH	0 (0.0%)	15 (15.0%)
SF-36 pain	1 (1.0%)	3 (3.0%)
KOOS Sym	0 (0.0%)	1 (1.0%)
KOOS ADL	0 (0.0%)	8 (8.0%)
KOOS sport	0 (0.0%)	2 (2.0%)
KOOS pain	0 (0.0%)	2 (2.0%)
KOOS QOL	0 (0.0%)	1 (1.0%)
Marx	9 (9.0%)	42 (42.0%)
ED-5D	0 (0.0%)	3 (3.0%)

\*Definitions: SF-36 Physical Functioning, SF-36 General Health Perceptions, SF-36 Bodily pain, Knee Injury and Osteoarthritis Outcome Score (KOOS) Symptoms, KOOS Daily Living, KOOS Sports and Recreational Activities, KOOS Pain, KOOS Quality of Life, Marx Knee Activity Rating Scale, EuroQol-5 Dimension

Sport, KOOS ADL, and SF-36 PF, and a lesser correlation with those not specifically aimed at measuring physical function. Commenting on related work, other authors have highlighted the need to make this correlation in order for PROMIS PF CAT to make the jump into clinical usefulness [14]. The findings herein indicate that PROMIS PF CAT maintains construct validity and likely provides similar

information as some of the traditional physical function PRO instruments compared here.

Previous work has suggested that the PROMIS PF CAT has several positive traits including high-content validity and reliability, good responsiveness to change, and improved efficiency that can potentially reduce test fatigue and facilitate data collection [2, 5, 6, 12–14, 20]. The finding of high correlations with other PF instruments herein builds on previous work by demonstrating high *construct* validity in addition to those characteristics described by others.

The current study also indicates that in the ACL injury population PROMIS PF CAT is inclusive with no floor or ceiling effects, despite early suspicion to the contrary [6] and in accordance with more recent work [12, 14]. Instruments that cover all patients including those at the extreme high or low end of functioning are essential in the applicability of these tools. This study suggests that PROMIS PF CAT performs well in this regard in this specific population. Evaluation of the responsiveness to change and inclusiveness of this instrument as function improves in this cohort postoperatively may be of further interest.

In agreement with other reports, PROMIS PF CAT in the current study also exhibited high efficiency relative to other common PRO instruments [5, 12, 14]. The large majority of patients herein only had to complete four items prior to the conclusion of the PROMIS PF CAT, compared to a total of 88 total items (36 for SF-36, 6 for ED-5D, 42 for KOOS, 4 for Marx) in the other traditional PROs used in this study. Being able to precisely record patient function using fewer items than most instruments is a key advantage of PROMIS PF CAT and has the potential to minimize test burden, facilitate data collection, and improve response rates [3, 4, 6, 7, 14]. Future work evaluating patient and administrator satisfaction and time and cost analysis with the use of PROMIS is indicated.

There are some notable limitations of the current study. It is recognized that the specificity of the patient population analyzed may limit its generalizability, but it was in line with the goals of the study to evaluate PROMIS in a relatively young and healthy population. Additionally, this work required patients to complete several PRO instruments one after another which could possibly have caused test fatigue and may arguably alter scores. An attempt to control for this limitation by randomizing the order of administration of the instruments was performed. Instead of timing how long it took each patient to complete each instrument, this study also substituted the number of items asked prior to completion as a marker of test burden. This was a necessity because of the administering clinic's design where patients may be interrupted during instrument administration, making time data inaccurate in many cases. It is the authors' opinion, however, that the relatively few number of items for PROMIS PF CAT was representative of its efficiency and

indeed this idea has been consistently supported throughout the literature to date. Finally, there was an unexpectedly strong correlation between PROMIS PF CAT and the EQ-5D, which is an instrument aimed at measuring general health rather than specifically physical function. Some may argue that this might indicate problems with discriminant validity. One possible alternative explanation for this, however, is that EQ-5D scores may correlate strongly with the physical function of the patient's measured.

In summary, the present work suggests that PROMIS PF CAT is a valid and efficient instrument for routine clinical assessment and may have advantages over some traditional fixed-length PROs. The strong correlations found here indicate that PROMIS PF CAT accurately measures physical function. Though the use of CAT, it is able to do so with high precision while using fewer questions, reducing burden on the patient. Efficient and precise measurements allow for easier use in the clinical setting and may improve response rates and allow for more comprehensive assessment of patient function before and after treatment, as patients may be more likely to complete a shorter test particularly when they are asked to complete these instruments at serial clinic visits over time.

## Conclusions

PROMIS PF CAT scores and other currently used PRO instruments measuring physical function in patients indicated for operative management of ACL injuries correlate strongly. The instrument is inclusive with no floor or ceiling effects in this population, and is a viable and efficient alternative to measuring physical function in the ACL injury population.

**Funding** No funding was provided for the completion of this study.

## Compliance with ethical standards

**Conflict of interest** Each author certifies that he or she, or a member of his or her immediate family, has no funding or commercial associations (e.g., consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article.

**Ethical approval** Each author certifies that his or her institution approved or waived approval for the human protocol for this investigation and that all investigations were conducted in conformity with ethical principles of research.

## References

1. Amstutz HC, Sew Hoy AL, Clarke IC (1981) UCLA anatomic total shoulder arthroplasty. *Clin Orthop Relat Res* 155:7–20

2. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C (2001) Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther* 14:128–146
3. Cella D, Gershon R, Lai JS, Choi S (2007) The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res* 16(Suppl 1):133–141
4. Chakravarty EF, Bjorner JB, Fries JF (2007) Improving patient reported outcomes using item response theory and computerized adaptive testing. *J Rheumatol* 34:1426–1431
5. Doring AC, Nota SP, Hageman MG, Ring DC (2014) Measurement of upper extremity disability using the patient-reported outcomes measurement information system. *J Hand Surg Am* 39:1160–1165
6. Fries J, Rose M, Krishnan E (2011) The PROMIS of better outcome assessment: responsiveness, floor and ceiling effects, and Internet administration. *J Rheumatol* 38:1759–1764
7. Fries JF, Bruce B, Cella D (2005) The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. *Clin Exp Rheumatol* 23:S53–S57
8. Garrard L, Price LR, Bott MJ, Gajewski BJ (2015) A novel method for expediting the development of patient-reported outcome measures and an evaluation of its performance via simulation. *BMC Med Res Methodol* 15:77
9. Harris WH (1969) Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am* 51:737–755
10. Hoppe DJ, Bhandari M (2008) Evidence-based orthopaedics: a brief history. *Indian J Orthop* 42:104–110
11. Huber EO, Meichtry A, de Bie RA, Bastiaenen CH (2016) Construct validity of change scores of the Chair Stand Test versus Timed Up and Go test, KOOS questionnaire and the isometric muscle strength test in patients with severe knee osteoarthritis undergoing total knee replacement. *Man Ther* 21:262–267
12. Hung M, Baumhauer JF, Brodsky JW, Cheng C, Ellis SJ, Franklin JD et al (2014) Psychometric comparison of the PROMIS physical function CAT with the FAAM and FFI for measuring patient-reported outcomes. *Foot Ankle Int* 35:592–599
13. Hung M, Baumhauer JF, Latt LD, Saltzman CL, SooHoo NF, Hunt KJ (2013) Validation of PROMIS (R) physical function computerized adaptive tests for orthopaedic foot and ankle outcome research. *Clin Orthop Relat Res* 471:3466–3474
14. Hung M, Stuart AR, Higgins TF, Saltzman CL, Kubiak EN (2014) 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* 28:439–443
15. Hurwitz S (2003) Evidence-based medicine in orthopaedic surgery—a way to the future. *Iowa Orthop J* 23:61–65
16. Marchetti P, Binazzi R, Vaccari V, Girolami M, Morici F, Impalomeni C et al (2005) Long-term results with cementless Fitek (or Fitmore) cups. *J Arthroplasty* 20:730–737
17. Marx RG (2003) Knee rating scales. *Arthroscopy* 19:1103–1108
18. Marx RG, Stump TJ, Jones EC, Wickiewicz TL, Warren RF (2001) Development and evaluation of an activity rating scale for disorders of the knee. *Am J Sports Med* 29:213–218
19. McHorney CA, Ware JE Jr, Raczek AE (1993) The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 31:247–263
20. Mohtadi NG, Chan DS, Hollinshead RM, Boorman RS, Hiemstra LA, Lo IK et al (2014) A randomized clinical trial comparing open and arthroscopic stabilization for recurrent traumatic anterior shoulder instability: two-year follow-up with disease-specific quality-of-life outcomes. *J Bone Jt Surg Am* 96:353–360
21. Nguyen J, Marx R, Hidaka C, Wilson S, Lyman S (2016) Validation of electronic administration of knee surveys among ACL-injured patients. *Knee Surg Sports Traumatol Arthrosc* 25:3116–3122
22. Noyes FR, Barber SD, Mooar LA (1989) A rationale for assessing sports activity levels and limitations in knee disorders. *Clin Orthop Relat Res* 246:238–249
23. Nutton RW, McBirmie JM, Phillips C (1997) Treatment of chronic rotator-cuff impingement by arthroscopic subacromial decompression. *J Bone Joint Surg Br* 79:73–76
24. Rabin R, de Charro F (2001) EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 33:337–343
25. Roos EM, Lohmander LS (2003) The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes* 1:64–64
26. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynon BD (1998) Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther* 28:88–96
27. Shoukri M (1998) *Statistical methods for health sciences*, 2nd edn. CRC Press, Pause
28. Tegner Y, Lysholm J (1985) Rating systems in the evaluation of knee ligament injuries. *Clin Orthop Relat Res* Sep:43–49
29. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J et al (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60:34–42
30. Ware JE Jr, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 30:473–483
31. Wyer PC, Silva SA (2009) Where is the wisdom? I—a conceptual history of evidence-based medicine. *J Eval Clin Pract* 15:891–898