

2018

## Validation and Generalizability of Preoperative PROMIS Scores to Predict Postoperative Success in Foot and Ankle Patients

Michael R. Anderson

Jeff Houck

Charles L. Saltzman

Man Hung

Florian Nickisch

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.georgefox.edu/pt\\_fac](https://digitalcommons.georgefox.edu/pt_fac)

 Part of the [Physical Therapy Commons](#)

---

---

**Authors**

Michael R. Anderson, Jeff Houck, Charles L. Saltzman, Man Hung, Florian Nickisch, Alexej Barg, Timothy Beals, and Judith F. Baumhauer

---

# Validation and Generalizability of Preoperative PROMIS Scores to Predict Postoperative Success in Foot and Ankle Patients

Michael R. Anderson, DO<sup>1</sup>, Jeff R. Houck, PT, PhD<sup>2</sup>, Charles L. Saltzman, MD<sup>3</sup>, Man Hung, PhD<sup>3</sup>, Florian Nickisch, MD<sup>3</sup>, Alexej Barg, MD<sup>3</sup>, Timothy Beals, MD<sup>3</sup>, and Judith F. Baumhauer, MD, MPH<sup>1</sup>

## Abstract

**Background:** A recent publication reported preoperative Patient-Reported Outcomes Measurement Instrumentation System (PROMIS) scores to be highly predictive in identifying patients who would and would not benefit from foot and ankle surgery. Their applicability to other patient populations is unknown. The aim of this study was to assess the validation and generalizability of previously published preoperative PROMIS physical function (PF) and pain interference (PI) threshold *t* scores as predictors of postoperative clinically meaningful improvement in foot and ankle patients from a geographically unique patient population.

**Methods:** Prospective PROMIS PF and PI scores of consecutive patient visits to a tertiary foot and ankle clinic were obtained between January 2014 and November 2016. Patients undergoing elective foot and ankle surgery were identified and PROMIS values obtained at initial and follow-up visits (average, 7.9 months). Analysis of variance was used to assess differences in PROMIS scores before and after surgery. The distributive method was used to estimate a minimal clinically important difference (MCID). Receiver operating characteristic curve analysis was used to determine thresholds for achieving and failing to achieve MCID. To assess the validity and generalizability of these threshold values, they were compared with previously published threshold values for accuracy using likelihood ratios and pre- and posttest probabilities, and the percentages of patients identified as achieving and failing to achieve MCID were evaluated using  $\chi^2$  analysis.

**Results:** There were significant improvements in PF ( $P < .001$ ) and PI ( $P < .001$ ) after surgery. The area under the curve for PF (0.77) was significant ( $P < .01$ ), and the thresholds for achieving MCID and not achieving MCID were similar to those in the prior study. A significant proportion of patients (88.9%) identified as not likely to achieve MCID failed to achieve MCID ( $P = .03$ ). A significant proportion of patients (84.2%) identified as likely to achieve MCID did achieve MCID ( $P < .01$ ). The area under the curve for PROMIS PI was not significant.

**Conclusions:** PROMIS PF threshold scores from published data were successful in classifying patients from a different patient and geographic population who would improve with surgery. If functional improvement is the goal, these thresholds could be used to help identify patients who will benefit from surgery and, most important, those who will not, adding value to foot and ankle health care.

**Level of evidence:** Level II, Prospective Comparative Study

**Keywords:** patient-reported outcomes, PROMIS, foot and ankle, pain, physical function

The utility of patient-reported outcomes (PROs) has begun to expand beyond the scope of research by demonstrating value in direct patient care, however, the adoption of PROs in daily clinical practice is not widespread. Systematic barriers to PRO adoption are numerous and include technological, financial, and work flow challenges to realizing clinical integration of PROs for physicians.<sup>1</sup> Widespread adoption, therefore, will hinge upon the realization that PROs

---

<sup>1</sup>University of Rochester, Rochester NY, USA

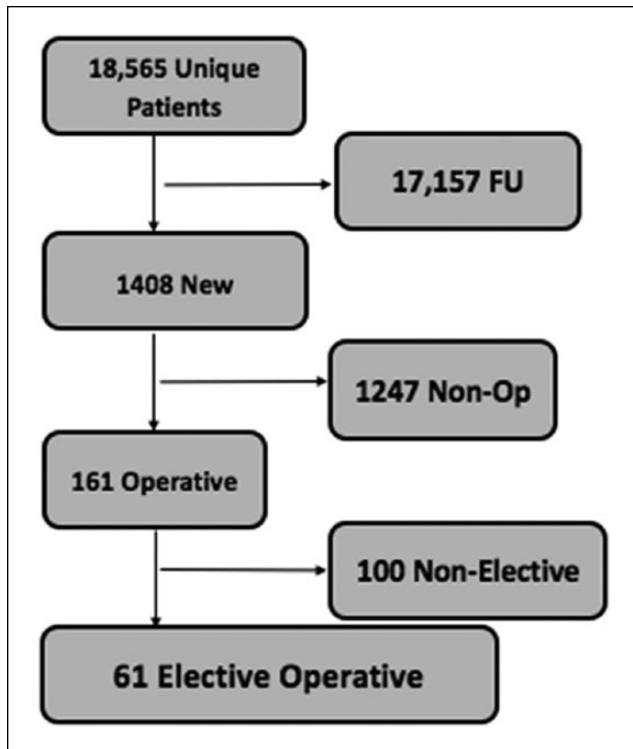
<sup>2</sup>George Fox University, Newberg, OR, USA

<sup>3</sup>University of Utah, Salt Lake City, UT, USA

## Corresponding Author:

Judith F. Baumhauer, MD, MPH, University of Rochester, School of Medicine and Dentistry, 601 Elmwood Avenue, Box 665, Rochester, NY 14642, USA.

Email: judy\_baumhauer@urmc.rochester.edu



**Figure 1.** Patients selected for inclusion.

facilitate clinically relevant improvement in care for both physicians and patients while being efficient and easy to use clinically.<sup>6</sup> The Patient-Reported Outcomes Measurement Instrumentation System (PROMIS), a validated measure that uses computerized adaptive testing (CAT) and item response theory, decreases administration time, avoids floor and ceiling effects, and achieves high accuracy.<sup>7</sup> In head-to-head comparisons, PROMIS scales demonstrate higher accuracy and efficiency than disease-specific legacy scales across the spectrum of orthopedics.<sup>3,9-12,14,17</sup> The benefits to applying PROMIS scales in clinical decision making remain uncertain.

Using PROMIS scales to predict how an individual patient will respond to a proposed treatment is of great interest to physicians and patients. Ho et al<sup>9</sup> recently reported preoperative PROMIS scores to be highly predictive in identifying patients who would and would not improve from foot and ankle surgery. In that study, PROMIS physical function (PF), pain interference (PI), and depression (D) scores were prospectively collected at an academic tertiary foot and ankle center. Having the pre- and postoperative data available and defining the minimal clinically important difference (MCID), the sensitivity (true-positive rate) and 100 minus specificity (false-positive rate) can be obtained and plotted on a curve called the receiver operating characteristic (ROC) curve, and predictive percentages can be determined. ROC curve analysis demonstrated that

preoperative PF, PI, and D scores were predictive of postoperative improvement. Furthermore, patients with poor preoperative function, high preoperative pain, and high preoperative depression were more likely to reach MCID than less affected patients, allowing the identification of PROMIS score thresholds with 95% accuracy. PROMIS PF scales have been compared with other foot and ankle legacy scales and found to be more efficient, reliable, and responsive to change.<sup>10</sup> The strength of PROMIS scales, which are global, non-disease-specific scales, in predicting MCID improvement was surprising given the differences across procedures, diagnoses, and severity.<sup>9</sup> The derived threshold values from that study have not been validated in a separate sample of foot and ankle patients and are not known to be generalizable. Although similar findings of preoperative PROMIS scores were found to be highly predictive of outcomes from hip,<sup>5</sup> knee,<sup>4</sup> and shoulder<sup>19</sup> arthroplasty patients, the proposed thresholds relative to foot and ankle surgery remain speculative when applied to a separate cohort of patients.

The purpose of this analysis was to assess the validity and generalizability of previously published preoperative PROMIS PF and PI *t* scores as predictors of postoperative clinically meaningful improvement in foot and ankle patients in a separate, geographically distinct cohort. We tested 3 hypotheses: (1) PROMIS scales (PF and PI) would improve at postoperative follow-up, as in other studies<sup>9,14</sup>; (2) preoperative PROMIS PF and PI scores would accurately predict postoperative improvement (MCID), resulting in similar threshold values (specifically, the hypothesis was that patients with high preoperative pain would be more likely to reach MCID than those with low preoperative pain and that those with low preoperative function would be more likely to reach MCID than those with high preoperative function); and (3) applying the preoperative PROMIS PF and PI thresholds reported previously to this data set would successfully identify patients experiencing postoperative MCID change in this validation sample over chance alone.

## Methods

PROMIS PF and PI scores were prospectively collected from consecutive patient visits to a multiple-surgeon tertiary foot and ankle clinic between January 2014 and November 2016. This consisted of 18565 unique patient visits, of which 1408 were new patient visits. Patients undergoing elective operative intervention were identified by Current Procedural Terminology codes. Application of exclusion criteria identical to those of Ho et al<sup>9</sup> (new patients, elective surgery, complete data sets with a minimum of 6 months of follow-up) resulted in a sample of 61 subjects (Figure 1). PROMIS PF and PI scores were assessed at initial and follow-up visits (minimum, 6 months;

**Table 1.** Comparison of Derivation and Validation Sample Characteristics.<sup>a</sup>

Variable	Validation Sample (n = 61)	Derivation Sample (Reference) (n = 61)
Age, y	50.4 ± 16.7	53.2 ± 13.6
Female gender	49.2%	67.2%
Average follow-up, mo	7.9 ± 2.9	7.3 ± 1.6
PROMIS PF score		
Pre	34.4 ± 8.7	34.2 ± 8.4
Post	40.4 ± 9.0	42.6 ± 7.1
Change	6.0 ± 11.6	8.4 ± 9.9
Percentage above MCID	44.3	34
PROMIS PI score		
Pre	64.4 ± 6.9	61.0 ± 7.5
Post	57.4 ± 9.5	57.1 ± 8.5
Change	-7.0 ± 8.4	-3.9 ± 9.9
Percentage above MCID	55.7	33.0

Abbreviations, MCID, minimal clinically important difference; PF, physical function; PI, pain interference; PROMIS, Patient-Reported Outcomes Measurement Instrumentation System.

<sup>a</sup>Data are expressed as mean ± SD or as percentages.

**Table 2.** Breakdown of Procedures Between Study Groups.

Procedure	Validation Sample (n = 61)	Derivation Sample (Reference) (n = 61)
Achilles tendon debridement	16.3%	9.8%
Ankle arthroscopy	13.1%	9.8%
Subtalar fusion	13.1%	
Hallux valgus correction	11.5%	8.2%
Flatfoot correction		9.8%
First metatarsal phalangeal fusion	8.2%	
Ankle arthrodesis	8.2%	
Lateral ligament reconstruction		8.2%

average, 7.9 months) (Table 1). The most common procedures were Achilles tendon debridement with secondary repair (16.3%), ankle arthroscopy (13.1%), subtalar fusion (13.1%), hallux valgus correction (11.5%), first metatarsal phalangeal fusion (8.2%), and ankle arthrodesis (8.2%) (Table 2).

The PROMIS CAT PF and PI scales were administered in the waiting room using an iPad, with scores stored and retrieved for later analysis. Item response theory and CAT maximize precision (eg, items are selected at appropriate difficulty), decrease patient burden (eg, each scale usually takes less than 1 minute to complete with only 4–7 questions), and increase efficiency by tailoring items on the basis of the level of response to previous items.<sup>8</sup> The CAT approach yields high levels of accuracy in less time than traditional scales.<sup>13</sup> The resulting PROMIS scales are referenced to *t* scores, where 50 equals the mean of the general US population in terms of sex, age, and race (according to the 2000 US census), and ±10 points represents 1 standard deviation.<sup>15,18</sup> A lower PROMIS PI *t* score indicates less pain interference in daily activities, and a higher PROMIS PF *t* score indicates greater physical function.

### Statistical Analysis

PROMIS PF and PI score improvement from the pre- to the postoperative visit was evaluated using repeated-measures 2-way analysis of variance. The 2 repeated factors were time (pre to post) and PROMIS scale (PF and PI). Gender and age were entered as covariates in the analysis. A lower PROMIS PI score and a higher PROMIS PF score indicate improvement. An interaction effect was consistent with improvement on both scales. To evaluate whether significant improvement was occurring, pairwise comparisons were pursued separately for PROMIS PF and PROMIS PI scores.

For the second and third hypotheses, it was necessary to establish improvement (ie, MCID) on the PROMIS PF and PI scales at postoperative follow-up. The distributive method of estimating the MCID was used.<sup>9</sup> There are multiple validated methods to determine the MCID in the literature, including the distribution-based, anchor-based, and Delphi approaches. Distribution-based approaches are based on statistical characteristics of the obtained samples. In this case, the distributive method was calculated as half

**Table 3.** Receiver Operating Characteristic Curve Analysis Thresholds From a Previous Study.<sup>9</sup>

	Area Under the Curve	95% Confidence Interval	P Value
Physical function	0.77	0.64–0.90	<.01
Pain interference	0.57	0.42–0.72	.33

of the standard deviation of the change scores observed (ie, the preoperative to postoperative changes in PROMIS PF and PI scores).<sup>16</sup>

Two approaches were used to validate the proposed thresholds that were previously published. Initially, ROC analysis was completed on PROMIS PF and PI scores from the current data set and compared with the previous analysis. This includes ROC analysis to determine the area under the curve (AUC) for PROMIS PF and PI scores. The AUC represents the percentage of patients who could be correctly identified as achieving or failing to achieve MCID for all possible thresholds. If the AUC was significant (ie, >0.5), new thresholds were determined by selecting the thresholds closest to 95% specificity for determining that the patient did achieve an MCID improvement and 95% specificity for concluding that the patient did not achieve a MCID improvement. Once the thresholds were identified, likelihood ratios (the odds a patient would achieve or fail to achieve MCID), pretest probabilities, and posttest probabilities were calculated. These are the probabilities that an individual patient would be correctly identified preoperatively as meeting MCID or failing to meet MCID at the postoperative time point, prior to application of the threshold (pretest probability) and after the threshold is applied (posttest probability). To compare the effect of applying the previous thresholds to the current data set, the thresholds published (PROMIS PF, 29.7 and 42; PROMIS PI, 67.2 and 55) were applied to the present study data, and the same set of variables were calculated for comparison.<sup>9</sup> A subsequent second approach evaluated the categorization of the new data set using published thresholds. Using 2 × 2 tables, thresholds from the previously published derivation sample were used to categorize patients into 1 of 3 groups (rows)—(1) expected to not achieve MCID, (2) expected to achieve MCID, and (3) ambiguous as to MCID achievement<sup>9</sup>—and this was compared with the known MCID improvement (columns). Chi-square analysis was used to test whether the proportion of patients placed in these 6 categories was due to chance.

## Results

There was significant improvement in PROMIS PF and PI scores over time. The 2-way analysis of variance showed a significant interaction of time and PROMIS scores ( $P < .001$ ). Neither age ( $P = .069$ ) nor gender ( $P = .33$ ) significantly influenced the results. Pairwise comparisons for PROMIS

PI were significant ( $P < .001$ ), with an average pre- to postoperative improvement of 7.0 (95% confidence interval [CI], 4.9–9.1). Similarly, pairwise comparisons for PROMIS PF were significant ( $P < .001$ ), with average pre- to postoperative improvement of 6.0 (95% CI, 3.1–9.0).

ROC analysis showed that PROMIS PF and not PROMIS PI was predictive of postoperative determination of meeting MCID (Table 3). The PROMIS PF AUC of 0.77 was significant ( $P < .001$ ). A PROMIS PF score of 23.8 or lower yielded 94.1% specificity for identifying patients who would achieve MCID improvement at follow-up. A PROMIS PF score of 41.6 yielded 96.3% specificity for identifying those patients who did not achieve MCID improvement at follow-up. The likelihood ratios were 6.3 and 7.2 for ruling in achieving MCID improvement and not achieving MCID improvement, respectively. As a consequence, the posttest probabilities were also high (83.3% for achieving MCID improvement and 90% for failing to achieve MCID improvement). Applying the threshold scores from Ho et al<sup>9</sup> resulted in similar values for specificity, likelihood ratios, and pretest and posttest probabilities (Table 4). In contrast, the PROMIS PI AUC of 0.57 was not significant ( $P = .33$ ). Therefore, no similar analysis of PROMIS PI was pursued. An unplanned, post hoc Pearson correlation analysis of the preoperative PROMIS PI score with the change from pre- to postoperative PROMIS PI produced a low and nonsignificant correlation coefficient ( $r = -0.25$ ,  $P = .06$ ).

The  $\chi^2$  analysis of the 2 × 2 analysis of the categories on the basis of MCID improvement was significant (PROMIS PF  $\chi^2 = 18.9$ ,  $P < .01$ ). The proportion of patients placed in each category is shown in Table 5. The total percentage of patients classified as achieving or not achieving MCID using the thresholds for this study was 45.8%. Applying the threshold values from Ho et al<sup>9</sup> to this data set, 39.3% of the sample was categorized correctly (8 patients [13.1%] achieving MCID and 16 patients [26.2%] failing to achieve MCID). In contrast only 6.5% of the sample (4 patients) was categorized incorrectly, and 54.2% (33 patients) could not be classified.

## Discussion

This validation study supports the use of preoperative PROMIS PF scales as accurate global health PRO instruments while giving patients a voice in their health care. These data in a geographically distinct location replicate the

**Table 4.** Threshold Values and Likelihood Ratios for Predicting MCID Change.

	MCID	Threshold	Observed Specificity (95% CI)	Likelihood Ratio (95% CI)	Pretest Probability (%)	Posttest Probability (%)	Change From Pretest Probability (%)
Present ROC analysis							
95% specificity for achieving MCID	5.8	<28.0	94.1 (80.3–99.3)	8.8 (2.2–35.5)	44.3	87.5	43.2
95% specificity for not achieving MCID	5.8	>41.6	96.3 (81.0–99.9)	7.2 (1.0–53.0)	55.7	90.0	34.3
Ho et al <sup>9</sup> criteria applied to data							
95% specificity for achieving MCID	4.2 <sup>a</sup>	<29.7	91.2 (76.3–98.1)	6.7 (2.2–20.7)	44.3	84.2	39.9
95% specificity for not achieving MCID	4.2 <sup>a</sup>	>42.0	96.3 (81.0–99.9)	6.4 (0.9–47.7)	55.7	88.9	33.2

Abbreviations, CI, confidence interval; MCID, minimal clinically important difference; ROC, receiver operating characteristic.

<sup>a</sup>MCIDs from Ho et al<sup>9</sup> are noted, but the MCID thresholds from the present data set were used.

**Table 5.** Chi-Square Analysis of Proportions Using Thresholds From Ho et al.<sup>9</sup>

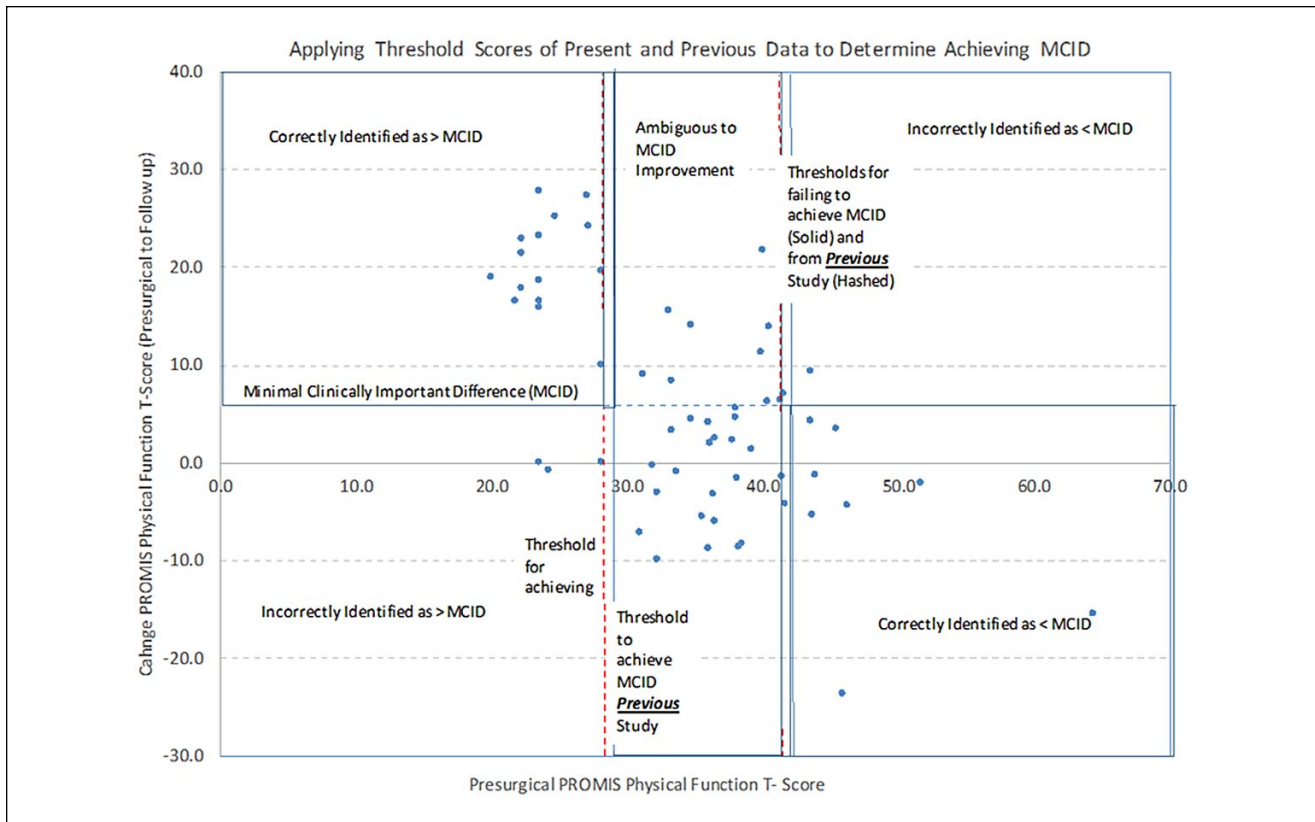
		Failed to Achieve MCID	Achieved MCID	$\chi^2$	P Value
Baseline					
Physical function					
Failed to achieve MCID	>42.0	16 (26.2%)	3 (4.9%)		
Ambiguous range	29.7–42	10 (16.4%)	23 (37.7%)		
Achieved MCID	<29.7	1 (1.6%)	8 (13.1%)	18.9	<.001

positive findings of a previous study. Although validation of a scale is an evolving process, this study validates (1) the ability of PROMIS PF and PI to detect pre- to postoperative change in physical ability and pain, (2) the ability of preoperative PROMIS PF to predict postoperative improvement, and (3) the ability to correctly classify preoperatively 39.3% of patients. This provides guidance to 40% of patients as to whether they would benefit from surgery. The reverse is also true: approximately 60% of patients could not be preoperatively identified by PROMIS scores as benefiting or failing to benefit from surgery. The ability of PROMIS PI to predict postoperative improvement was not supported. The analysis also supports the threshold values published in the derivation study.<sup>9</sup> Other derivation studies also support the use of preoperative PRO measures for surgical decisions, but none were tested in unique other populations to determine their validity.<sup>4,5,9,19</sup> Although the development of PROMIS scales is occurring rapidly and with high rigor, studies evaluating their clinical effectiveness are also needed to allow translation of *t* scores into clinical usefulness for providers and patients.

The validation sample of patients was similar in terms of Current Procedural Terminology codes, diagnoses, follow-up length, and PROMIS scores but differed in gender distribution (Table 1). Much as in the work of Ho et al,<sup>9</sup> the

patient population studied was representative of a tertiary academic foot and ankle practice with a wide range of diagnoses and procedures (Table 2). Furthermore, the percentages of the various procedures were similar to those in a large population of common foot and ankle procedures published by the American Board of Orthopaedic Surgery oral examination.<sup>14</sup> The follow-up period was also similar between the derivation sample and this validation sample. Although prognosis data are limited, the follow-up length might be characterized as beyond the rapid recovery that occurs after foot and ankle procedures but short of the gradual progress that may occur from 6 to 24 months.<sup>2</sup> The preoperative PROMIS scores suggest significant pain (PROMIS PI score = 64.4) and loss of function (PROMIS PF score = 34.4), indicating that these patients were appropriate for surgery. These preoperative PROMIS PF and PI *t* scores are within 3.4 points of the derivation sample scores, suggesting similar preoperative severity. With the exception of gender, the samples are similar on the metrics documented, and therefore similar findings between samples could reasonably be expected.

The improvements in PROMIS PF and PI scores after foot and ankle surgery were within the variation of other published studies.<sup>14</sup> Similar to previous studies, PROMIS PF and PI scores significantly improved at postoperative



**Figure 2.** Patients identified as achieving and failing to achieve a minimal clinically important difference (MCID) using presurgical Patient-Reported Outcomes Measurement Instrumentation System (PROMIS) physical function on the basis of the present data (dots) and a previous study (solid and dashed lines).

follow-up.<sup>9,14</sup> Raising the specter of interinstitutional differences, there were some similarities and potentially important differences. PROMIS PF pre- to postoperative improvement was more than half a standard deviation for both PROMIS PI (7.0; 95% CI, 4.9–9.1) and PROMIS PF (6.0; 95% CI, 4.9–9.1). However, the improvement in PROMIS PF was higher and that in PROMIS PI lower compared with the derivation sample. In a large multicenter 6-month outcomes study for specific Current Procedural Terminology codes, improvement in PROMIS PI and PF scores showed wide variation (7.5 for hallux valgus procedures compared with 24.3 for ankle arthrodesis), suggesting that specific procedures may yield different responses. Nevertheless, these data are consistent with previous studies in documenting the ability of PROMIS PI and PF to determine aggregate changes in pain and function after foot and ankle surgery.

New to this study is the validation of the ability of preoperative PROMIS PF but not PROMIS PI to predict improvement in patients undergoing foot and ankle surgery. The derivation sample noted the ability of both PROMIS PF and PI to predict MCID improvement at follow-up.<sup>9</sup> In this validation analysis, PROMIS PI was not predictive of MCID

improvement at follow-up (Table 3). The correlation between preoperative PROMIS PI score and change in PROMIS PI score was too low to yield a useful prediction ( $r = -0.25$ ,  $P = .06$ ). However, PROMIS PF was significant and yielded very similar thresholds for MCID improvement as in a previous study (Table 4). Remarkably, the threshold value for determining preoperatively patients who were not likely to achieve MCID was 41.6 in the present study and 42.0 in the previous study. This suggests that patients who self-report as higher than 42 on PROMIS PF are not likely to experience MCID improvement at an average follow-up of 7 to 8 months. The preoperative PROMIS PF thresholds of the derivation sample were within 1.7 points (28.0 vs 29.7) for achieving an MCID (Table 4, Figure 2). To refine the thresholds further, a large data set that heavily populates the full range of the scale is necessary. Nevertheless, patients with preoperative PROMIS PF values lower than 29.7 were highly likely to achieve an MCID change at 7- to 8-month follow-up.

A key way to judge the value of applying the threshold preoperatively is by evaluating the difference between the pretest and posttest probability. Large differences suggest a high value in applying the threshold clinically. The



differences in pre- to posttest probabilities range from 33.3% to 43.2%, suggesting large shifts in probabilities. Additionally, the posttest probabilities are all higher than 83.3%, suggesting that the preoperative thresholds yield higher than 80% probabilities that a patient will or will not achieve MCID improvement on the basis of preoperative data. If these thresholds were applied to the present data set, 10 procedures (16.4%) that did not produce MCID improvement might have been avoided and may have saved health care dollars allocated to surgery (Table 5, Figure 2). In total, the  $\chi^2$  analysis confirms that proportions of patients based on previous thresholds lead to meaningful groups beyond what can be expected from chance (Table 5). This analysis shows that if used, clinically preoperative PROMIS PF data may influence 45.8% of patient encounters.

There were several important limitations to this study. First, multiple disparate diagnoses were included. As a validation study, it was important to match the inclusion criteria of Ho et al,<sup>9</sup> but there may be more optimal PROMIS PF or PI thresholds for each particular diagnosis.<sup>9,14</sup>

Second, the inclusion criteria allowed for patients with 6-month follow-up. Patients doing poorly following surgery from delayed union, persistent swelling, infection, or other complications may be overrepresented and skew the change scores down compared with other studies.<sup>14</sup> Prospective follow-up data would eliminate bias from patients who did not attend follow-up for various reasons (eg, a smooth recovery). Longer term follow-up also may demonstrate differences, with some patients continuing to experience improvement.

Third, a larger sample size would increase the precision of the threshold values. The moderate correlation ( $r = 0.63$ ; Figure 2) between preoperative PROMIS PF scores and change in preoperative to postoperative PROMIS PF scores makes it necessary to increase the sample size to more accurately define optimal thresholds.

Fourth, MCID was defined by using the distribution method, specifically, 0.5 standard deviations. Additional studies are being performed to correlate this method with an anchor-based method.

Fifth, this study appropriately classified approximately 40% of patients as benefiting or failing to benefit from surgery. The goal of any test is to have high sensitivity and specificity to identify the trait (or disease) of interest in a population. Although this study did validate the prior threshold values (nearly identical), it did not aid in increasing the value of these PROMIS PF findings. However, correctly identifying even 40% of patients is a great aid to patients, clinicians, and the health care system.

Finally, applicability of the threshold values to unique populations remains unknown. For example, elite athletes likely have a higher threshold for preoperative PF or may respond favorably to less than 0.5 standard deviations of

MCID given the thin margin of function that can affect performance in athletes. Caution should be used when applying these findings to unique patient populations.

## Conclusions

This study replicated the ability of PROMIS PF and PI scores to discriminate aggregate improvement after foot and ankle surgery. In addition, this validation study supports the use of preoperative PROMIS PF but not PROMIS PI for clinical decision making with patients with high or low PROMIS PF  $t$  scores and considering foot and ankle surgery. Foot and ankle surgeons should consider including PROMIS PF assessment as part of their preoperative workup when engaging patients in clinical decisions regarding foot and ankle surgery.

## Acknowledgments

We thank Jerry Bounsanga, BS, and Yushan Gu, BS, at the University of Utah Department of Orthopaedics Quality Outcomes Research and Assessment for their assistance in data preparation, cleaning, and processing.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. ICMJE forms for all authors are available online.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was partly funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health under award number U01AR067138. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

1. Basch E. Patient-reported outcomes—harnessing patients' voices to improve clinical care. *N Engl J Med*. 2017;376(2):105–108.
2. Beckenkamp PR, Lin CW, Chagpar S, et al. Prognosis of physical function following ankle fracture: a systematic review with meta-analysis. *J Orthop Sports Phys Ther*. 2014;44(11):841–851, B2.
3. Beckmann JT, Hung M, Bounsanga J, et al. 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(12):1961–1967.
4. Berliner JL, Brodke DJ, Chan V, SooHoo NF, Bozic KJ. John Charnley Award: preoperative patient-reported outcome measures predict clinically meaningful improvement in function after THA. *Clin Orthop Relat Res*. 2016;474(2):321–329.

5. Berliner JL, Brodke DJ, Chan V, SooHoo NF, Bozic KJ. Can preoperative patient-reported outcome measures be used to predict meaningful improvement in function after TKA? *Clin Orthop Relat Res.* 2017;475(1):149–157.
6. Brodke DJ, Saltzman CL, Brodke DS. PROMIS for orthopaedic outcomes measurement. *J Am Acad Orthop Surg.* 2016;24(11):744–749.
7. Cella D, Riley W, Stone A, et al. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol.* 2010;63(11):1179–1194.
8. 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(5 Pt 2):1694–1711.
9. Ho B, Houck JR, Flemister AS, et al. Preoperative PROMIS scores predict postoperative success in foot and ankle patients. *Foot Ankle Int.* 2016;37(9):911–918.
10. 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.
11. Hung M, Hon SD, Franklin JD, et al. Psychometric properties of the PROMIS physical function item bank in patients with spinal disorders. *Spine (Phila Pa 1976).* 2014;39(2):158–163.
12. Hung M, Nickisch F, Beals TC, Greene T, Clegg DO, Saltzman CL. New paradigm for patient-reported outcomes assessment in foot & ankle research: computerized adaptive testing. *Foot Ankle Int.* 2012;33(8):621–626.
13. 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(8):439–443.
14. Hunt KJ, Alexander I, Baumhauer J, et al. The Orthopaedic Foot and Ankle Outcomes Research (OFAR) network: feasibility of a multicenter network for patient outcomes assessment in foot and ankle. *Foot Ankle Int.* 2014;35(9):847–854.
15. Liu H, Cella D, Gershon R, et al. Representativeness of the patient-reported outcomes measurement information system Internet panel. *J Clin Epidemiol.* 2010;63(11):1169–1178.
16. Norman GR, Sloan JA, Wywich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care.* 2003;41(5):582–592.
17. Papuga MO, Beck CA, Kates SL, Schwarz EM, Maloney MD. Validation of GAITRite and PROMIS as high-throughput physical function outcome measures following ACL reconstruction. *J Orthop Res.* 2014;32(6):793–801.
18. Pilkonis PA, Choi SW, Reise SP, et al. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS<sup>®</sup>): depression, anxiety, and anger. *Assessment.* 2011;18(3):263–283.
19. Wong SE, Zhang AL, Berliner JL, Ma CB, Feeley BT. Preoperative patient-reported scores can predict postoperative outcomes after shoulder arthroplasty. *J Shoulder Elbow Surg.* 2016;25(6):913–919.