8-2018

Preoperative PROMIS Scores Predict Postoperative PROMIS Score Improvement for Patients Undergoing Hand Surgery

David N. Bernstein

Jeff Houck

Ronald M. Gonzalez

Danielle M. Wilbur

Richard J. Miller

See next page for additional authors

Authors

David N. Bernstein, Jeff Houck, Ronald M. Gonzalez, Danielle M. Wilbur, Richard J. Miller, David J. Mitten, and Warren C. Hammert

Preoperative PROMIS Scores Predict Postoperative PROMIS Score Improvement for Patients Undergoing Hand Surgery

David N. Bernstein[1], Jeff R. Houck[2], Ronald M. Gonzalez[1], Danielle M. Wilbur[1], Richard J. Miller[1], David J. Mitten[1], and Warren C. Hammert[1]

Abstract

Background: Patient-Reported Outcomes Measurement Information System (PROMIS) can be used alongside preoperative patient characteristics to set postsurgery expectations. This study aimed to analyze whether preoperative scores can predict significant postoperative PROMIS score improvement. **Methods:** Patients undergoing hand and wrist surgery with initial and greater than 6-month follow-up PROMIS scores were assigned to derivation or validation cohorts, separating trauma and nontrauma conditions. Receiver operating characteristic curves were calculated for the derivation cohort to determine whether preoperative PROMIS scores could predict postoperative PROMIS score improvement utilizing minimal clinically important difference principles. **Results:** In the nontrauma sample, patients with baseline Physical Function (PF) scores below 31.0 and Pain Interference (PI) and Depression scores above 68.2 and 62.2, respectively, improved their postoperative PROMIS scores with 95%, 96%, and 94% specificity. Patients with baseline PF scores above 52.1 and PI and Depression scores below 49.5 and 39.5, respectively, did not substantially improve their postoperative PROMIS scores with 94%, 93%, and 96% sensitivity. In the trauma sample, patients with baseline PF scores below 34.8 and PI and Depression scores above 69.2 and 62.2, respectively, each improved their postoperative PROMIS scores with 95% specificity. Patients with baseline PF scores above 52.1 and PI and Depression scores below 46.6 and 44.0, respectively, did not substantially improve their postoperative scores with 95%, 94%, and 95% sensitivity. **Conclusions:** Preoperative PROMIS PF, PI, and Depression scores can predict postoperative PROMIS score improvement for a select group of patients, which may help in setting expectations. Future work can help determine the level of true clinical improvement these findings represent.

Keywords: hand surgery, patient-reported outcomes, PROMIS, shared decision-making, value-based health care

Introduction

Hand surgeons often face the challenge of preoperatively setting patient expectations for symptomatic and functional improvement following a surgical procedure. The challenge is magnified given the many variables, including patient risk factors (eg, body mass index [BMI], age, mental health status) and type of surgical intervention needed, playing a role in clinical improvement post surgery. Recent initiatives have begun to focus more on patient viewpoints instead of health care provider and system perspectives when determining treatment success. At the core of this movement is the development of patient-reported outcome (PRO) tools. In 2004, the National Institutes of Health (NIH) initiated large scale support for the creation of the Patient-Reported Outcomes Measurement Information System (PROMIS), which was developed to be a validated general PRO measure using computerized adaptive testing based on item response theory.[8] PROMIS has since been implemented in a number of medical settings, and research in a number of surgical areas is taking place.[19]

In the upper extremity literature, PROMIS is correlated strongly with "gold standard" legacy PRO tools like Disabilities of the Arm, Shoulder and Hand (DASH)/Quick-DASH.[21,23,25,28] Furthermore, in general, PROMIS has been shown to be more efficient and perform better (ie, better ceiling and floor effects) than traditional PRO tools.[13,21,28] However, there continues to be areas of needed improvement

[1]University of Rochester Medical Center, NY, USA
[2]George Fox University, Newberg, OR, USA

Corresponding Author:
Warren C. Hammert, Department of Orthopaedic Surgery, University of Rochester Medical Center, 601 Elmwood Avenue, Box 665, Rochester, NY 14642, USA.
Email: warren_hammert@urmc.rochester.edu

with PROMIS, most notable with the PROMIS Upper Extremity (UE) domain. Beckmann et al noted that PROMIS UE had a much higher ceiling effect (10.82%) compared with PROMIS Physical Function (PF) (1.32%) or DASH (5.28%).[3] Additional prior work has shown PROMIS UE ceiling effects that appear to limit patients from improving above 0.6 of a standard deviation (SD) above the assumed population mean score of 50.[4] In addition, PROMIS instruments are updated as new versions for each domain become available. While PROMIS has provided a great means of tracking patient status in a general sense, it remains unclear how to most effectively interpret and utilize PROMIS scores. Ideally, the surgeon would be able to use preoperative PROMIS scores, along with more traditional clinical insights and tools, to confidently predict which patients would improve following surgery. This question is being researched across a number of orthopedic subspecialties utilizing a variety of PRO tools including the PROMIS and Knee Injury and Osteoarthritis Outcome Score/Hip Disability and Osteoarthritis Outcome Score. Initial results suggest preoperative PRO scores are potentially important for predicting postoperative outcomes in foot and ankle surgery and total joint arthroplasty.[5,6,17] Of note, a follow-up study in the foot and ankle literature demonstrated that the same predictive rule using preoperative PROMIS scores applied to a separate patient cohort from a different institution could still be used to predict postoperative improvement.[2]

Surgeons and patients seeking hand surgery care can benefit from a more complete understanding of how PROMIS scores can better inform shared decision-making and appropriate treatment options during office visits. Specifically, determining the value of PROMIS in setting more specific postoperative outcome expectations would help both surgeons and patients enter treatment with a concrete idea of the attainable goals given each patient's unique preoperative status. To accomplish this goal, we utilized a general hand surgery patient population undergoing procedures for both atraumatic and traumatic conditions at a single academic, urban tertiary care center. The principle objectives of this study were to prospectively analyze: (1) whether PROMIS domains (PF, Pain Interference [PI], and Depression) improve from preoperative to postoperative follow-up; and (2) whether preoperative PROMIS scores can be used to predict clinical success (defined as improvement greater than the minimal clinically important difference [MCID]) following hand surgery to determine whether certain outcome expectations can be set prior to a procedure. We hypothesized that: (1) preoperative PROMIS scores in all 3 domains could predict substantial postoperative PROMIS score improvement, defined as improvement greater than the traditional MCID, which may or may not signify true meaningful clinical improvement; and (2) preoperative PROMIS scores in all 3 domains could predict patients who failed to attain substantial postoperative PROMIS score improvement at follow-up.

Materials and Methods

PROMIS PF v1.0, PI v1.0, and Depression v1.0 data were collected prospectively on all consecutive patients presenting to the orthopedic hand clinic at a single urban tertiary academic medical center between February 2015 and October 2016. PROMIS UE was not used because it was unavailable at the time of data collection. PROMIS domains are assumed to follow a normal distribution at a population level with a mean *T*-score of 50 and SD of 10. Higher PROMIS PF scores symbolize improved physical function, while lower PROMIS PI and Depression scores reflect decreased activity due to pain and worse mental health, respectively. To be included in our study, a patient was required to have PROMIS PF, PI, and Depression scores at both the initial and final follow-up visit at least 6 months after initial presentation to be included. Using *International Classification of Diseases, Ninth Revision* (*ICD-9*) and *Tenth Revision* (*ICD-10*) and *Current Procedural Terminology* (*CPT*) codes, patients were split into 2 samples: (1) hand surgery for traumatic cause (eg, fracture) and (2) hand surgery for a nontraumatic cause. This was done as we thought there might be a difference in traumatic and elective conditions. Patients who had only received nonsurgical treatments (eg, corticosteroid injections) without a surgical procedure and at least a 6-month follow-up visit were removed from the dataset. For each sample subset, all patients were required to have completed 3 PROMIS domains (PF, PI and Depression) at both the initial and follow-up visits to be included. Data from the initial visits were denoted as the preoperative PROMIS scores, while data from the follow-up visits were denoted as the postoperative PROMIS scores. In the hand surgery for trauma sample, a total of 231 patients were analyzed (Table 1), while in the hand surgery for nontrauma sample, 157 patients met the inclusion criteria for this study (Table 1). Using a random number generator, patients in each sample subset were randomly assigned to a derivation (nontrauma, $n = 79$; trauma, $n = 116$) or validation cohort (nontrauma, $n = 78$; trauma, $n = 115$).

Patients completed 3 PROMIS domains on Apple iPads upon check-in to their preoperative clinic appointment for nontraumatic or nonemergent traumatic surgical cases. The traumatic conditions consist of tendon injuries, closed fractures, and similar conditions that are initially evaluated in the emergency department or at an urgent care and then present to the office prior to definitive treatment, at which time PROMIS questionnaires are completed. The use of Apple iPads to collect PROs in hand clinics has been shown to be efficient and a preferable to traditional pen and paper data collection.[32] PROMIS utilizes item

Table 1. Breakdown of Surgical Procedures in Traumatic and Nontraumatic Samples.

Name of procedure	n (%)
Surgery for traumatic cause	
Fracture treatment	57 (24.7)
Nerve repair/neurolysis	33 (13.9)
Triangular fibrocartilage complex repairs	28 (12.1)
Arthroscopic treatments (repairs or debridements)	21 (9.1)
Amputations	15 (6.5)
Other	77 (33.7)
Surgery for nontraumatic cause	
Carpal tunnel release	48 (30.6)
Trigger finger release	31 (19.7)
Primary arthroplasty	18 (11.5)
Other	60 (38.2)

response theory as part of a computerized adaptive test (CAT) and has demonstrated reliability and precision as a PRO tool.[7] Furthermore, the PROMIS domains selected (PF, PI, and Depression) have been validated against many legacy orthopedic PRO tools and can be completed quickly during a busy clinic workflow.[9,18,21,25-27] The Research Subjects Review Board at our institution approved the study.

Statistical Analysis

To place postoperative PROMIS score improvement into context, an MCID was determined using the distribution-based method. To date, no anchor-based methods have determined MCID for PROMIS PF, PI, and Depression domains in orthopedic surgery. However, previous work has focused on determining the MCID using the distribution-based method.[24] Indeed, the MCID was set as one half the SD of the respective PRO measure based on our collected data similar to recent studies in orthopedic surgery.[5,6,17,31]

Using the distribution-based approach for the nontrauma sample subset, the substantial postoperative PROMIS score improvement values for PF, PI, and Depression were 4.7, 4.5, and 4.5, respectively. Using the same approach for the trauma sample subset, the substantial postoperative PROMIS score improvement values for PROMIS PF, PI, and Depression were calculated and set at 5.4, 5.3, and 5.8, respectively.

The nontrauma and trauma sample subsets were analyzed separately using a split sample approach to derive and validate which preoperative scores classify patients substantially having improved postoperative PROMIS scores or not. First, randomly determined derivation and validation cohorts were compared between samples on key variables (Tables 2 and 3). Second, receiver operating characteristic (ROC) curves were then calculated for the derivation cohort

Table 2. Comparison of Patient Characteristics in the Derivation and Validation Cohorts for Those Undergoing Surgery for a Nontraumatic Cause.

	Derivation cohort (n = 79)	Validation cohort (n = 78)
Age	53.4 (14.7)	55.0 (14.6)
Sex, n (%)		
Men	29 (36.7)	37 (47.4)
Women	50 (63.3)	41 (52.3)
Follow-up (months) (SD)	11.1 (3.0)	11.4 (3.4)
PROMIS score, baseline (SD)		
PF	46.3 (10.0)	46.9 (8.8)
PI	57.5 (8.4)	58.5 (7.2)
Depression	48.5 (9.9)	48.0 (9.4)
PROMIS score, follow-up (SD)		
PF	45.7 (8.7)	46.2 (9.3)
PI	55.8 (9.5)	55.3 (8.9)
Depression	46.4 (10.5)	45.9 (10.2)
PROMIS score, change (SD)		
PF	−0.7 (9.3)	−0.6 (8.1)
PI	−1.7 (8.9)	−3.2 (8.6)
Depression	−2.1 (8.9)	−2.1 (9.0)

Note. PROMIS = Patient-Reported Outcomes Measurement Information System; PF = Physical Function; PI = Pain Interference.

Table 3. Comparison of Patient Characteristics in the Derivation and Validation Cohorts for Those Undergoing Surgery for a Traumatic Cause.

	Derivation cohort (n = 116)	Validation cohort (n = 115)
Age	46.0 (18.7)	46.5 (16.3)
Sex, n (%)		
Men	54 (46.6)	61 (53.0)
Women	62 (53.4)	54 (47.0)
Follow-up (months) (SD)	11.1 (3.4)	11.3 (3.8)
PROMIS score, baseline		
PF	44.0 (9.8)	44.8 (10.8)
PI	58.3 (9.9)	59.4 (10.1)
Depression	49.0 (10.6)	49.2 (10.5)
PROMIS score, follow-up (SD)		
PF	47.4 (10.2)	46.7 (10.4)
PI	53.9 (10.1)	54.7 (10.3)
Depression	46.9 (11.5)	46.1 (11.7)
PROMIS score, change (SD)		
PF	3.3 (10.8)	1.9 (10.9)
PI	−4.4 (10.5)	−4.7 (11.3)
Depression	−2.1 (11.5)	−3.1 (10.6)

Note. PROMIS = Patient-Reported Outcomes Measurement Information System; PF = Physical Function; PI = Pain Interference.

Table 4. Receiver Operating Characteristic Curve Analysis, Surgery for Nontraumatic Cause Sample (Derivation Cohort, n = 79).

PROMIS domain	Area under the curve	*P* value	95% confidence interval
PF	0.73	.01	(0.57-0.86)
PI	0.69	< .01	(0.57-0.82)
Depression	0.63	< .05	(0.51-0.75)

PROMIS domain	Cutoff, achieve MCID	Specificity	Ambiguous range	Cutoff, fail to achieve MCID	Sensitivity
PF	< 31.0	95%	30.0-52.1	> 52.1	94%
PI	> 68.2	96%	49.5-68.2	< 49.5	93%
Depression	> 62.2	94%	39.5-62.2	< 39.5	96%

Note. PROMIS = Patient-Reported Outcomes Measurement Information System; PF = Physical Function; PI = Pain Interference; MCID = minimal clinically important difference.

to determine the preoperative cutoff scores for each PROMIS domain that would or would not lead to a patient substantially improving PROMIS scores postoperatively with around 95% specificity and sensitivity, respectively. The area under the curve (AUC), or *c*-statistic, was also determined for each PROMIS domain for the derivation cohort. A *c*-statistic of 0.5 means that the model performs no better than chance, while a *c*-statistic of 1 means that the model performs perfectly in predicting the outcome each time. A *c*-statistic over 0.7 is considered reasonably accurate, while a *c*-statistic greater than 0.8 is considered excellent.[15] Finally, chi-square analysis was used to test whether patients categorized (failing to substantially improve PROMIS scores, ambiguous range, substantially improve PROMIS scores) using the derived cutoffs in the validation cohort matched their known MCID category more consistently than chance. All analyses were conducted with significance set at $\alpha = 0.05$.

Because the ROC analysis is sensitive to prevalence of significant change, this was used to estimate the adequacy of the sample size. With a minimum sample of 75 participants and an AUC of 0.7, we estimated the 95% confidence interval for the AUC at 10% increments for the prevalence of significant change (ie, MCID). For prevalences of 20% or higher, the 95% confidence interval for the AUC excluded 0.5.

Results

When analyzing the nontraumatic sample subset using the derivation cohort, the *c*-statistics for PROMIS PF, PI, and Depression were 0.73 (95% CI, 0.57-0.86), 0.69 (95% CI, 0.57-0.82), and 0.63 (95% CI, 0.51-0.75), respectively (Table 4). Patients with a PROMIS PF score below 31.0 substantially improved their PROMIS scores with 95% specificity. Patients with a PROMIS PI score above 68.2 substantially improved their PROMIS scores with 96% specificity. Patients with a PROMIS Depression score above 62.2 substantially improved their PROMIS scores with 94% specificity. Patients with PROMIS PF scores above 52.1 did not substantially improve their PROMIS scores with 94% sensitivity. Patients with PROMIS PI scores below 49.5 did not substantially improve their PROMIS scores with 93% sensitivity. Patients with PROMIS Depression scores below 39.5 did not substantially improve their PROMIS scores with 96% sensitivity.

The chi-square tests for categorization of the validation cohort using each PROMIS scale were significant for the nontrauma sample subset (Table 5). Using the same nontrauma sample subset and applying the cutoffs determined from the derivation cohort, our results showed that baseline PF predicted 22% of patients failing to substantially improve their PROMIS scores (chi-square $P = .03$) (Table 5). Our model predicted those who would fail to substantially improve their PROMIS scores with 100% accuracy, meaning that all patients above our baseline PF cutoff (> 53.2) failed to substantially improve their PROMIS scores. Baseline PF predicted 2.5% of patients substantially improving their PROMIS scores but with poor accuracy (50% accuracy). Baseline PI predicted (100% accuracy) 10% of patients who substantially improved PROMIS scores (chi-square $P = .01$). Baseline PI also predicted (71% accuracy) 6.4% of patients who would fail to substantially improve their PROMIS scores. Baseline Depression predicted (88 % accuracy) 18% of patients who failed to substantially improve their PROMIS scores and (80% accuracy) 5.1% of patients who substantially improved their PROMIS scores (chi-square $P = .01$). A total of 57 (73%), 63 (81%), and 57 (73%) patients in PROMIS PF, PI, and Depression domains, respectively, were in the ambiguous range of PROMIS scores that did not allow for an accurate prediction.

When analyzing the trauma sample subset using the derivation cohort, the *c*-statistic for PROMIS PF, PI, and

Table 5. Chi-Square Analysis of Proportions Applied to Validation Cohort Using Derivation Cohort Cutoffs, Surgery for Nontraumatic Cause Sample (n = 78).

	Failed to achieve MCID	Achieved MCID	Chi-square statistic	*P* value
PROMIS PF, baseline				
Failed to achieve MCID, > 53.2	17	0		
Ambiguous range, 31.0-53.2	44	13		
Achieved MCID, < 31.0	2	2	7.0	.03
PROMIS PI, baseline				
Failed to achieve MCID, < 49.5	5	2		
Ambiguous range, 49.5-68.2	35	28		
Achieved MCID, > 68.2	0	8	10.0	< .01
PROMIS Depression, baseline				
Failed to achieve MCID, < 39.5	14	2		
Ambiguous range, 39.5 - 62.2	32	25		
Achieved MCID, > 62.2	1	4	8.8	.01

Note. MCID = minimal clinically important difference; PROMIS = Patient-Reported Outcomes Measurement Information System; PF = Physical Function; PI = Pain Interference.

Table 6. Receiver Operating Characteristic Curve Analysis (Derivation Cohort, Surgery for Traumatic Cause Sample [n = 116]).

PROMIS domain	Area under the curve	*P* value	95% confidence interval		
PF	0.69	< .01	(0.59-0.79)		
PI	0.69	< .01	(0.59-0.78)		
Depression	0.76	< .01	(0.69-0.84)		
PROMIS domain	**Cutoff, achieve MCID**	**Specificity**	**Ambiguous range**	**Cutoff, fail to achieve MCID**	**Sensitivity**
PF	< 34.8	95%	(34.8-52.1)	>52.1	95%
PI	> 69.2	95%	(46.6-69.2)	<46.6	94%
Depression	> 62.2	95%	(44.0-62.2)	<44.0	95%

Note. PROMIS = Patient-Reported Outcomes Measurement Information System; PF = Physical Function; PI = Pain Interference; MCID = minimal clinically important difference.

Depression scales were 0.69 (95% CI: 0.59-0.79; $P < .01$), 0.69 (95% CI: 0.59-0.78; $P < .01$), and 0.76 (95% CI: 0.69-0.84; $P < .01$), respectively (Table 6). Patients with PF scores < 34.8, PI scores > 69.2, and Depression scores > 62.2 substantially improved their PROMIS scores with near 95% specificity. Patients with PF scores > 52.1, PI scores < 46.6 and Depression scores < 44.0 failed to substantially improve their PROMIS scores with 95%, 94%, and 95% sensitivity, respectively. An ambiguous PROMIS scale range that did not allow for accurate PROMIS score predictions were defined as 33.8 to 52.1 for PF, 49.9 to 69.2 for PI, and 44.2 to 62.2 for Depression.

Further analysis of the trauma sample subset showed that the chi-square tests for categorization of the validation cohort using each PROMIS scale were significant (Table 7). The PF cutoffs applied to the validation cohort predicted 22% of patients (93% accuracy) failing to substantially improve their PROMIS scores and 10% of patients (63% accuracy) substantially improving their PROMIS scores (Table 7). The PI cutoffs applied to the validation cohort predicted 9.6% of patients (92% accuracy) failing to substantially improve their PROMIS scores and 13% of patients (88% accuracy) substantially improving their PROMIS scores. The Depression cutoffs applied to the validation cohort predicted 28% of patients (82% accuracy) failing to substantially improve their PROMIS scores and 6.1% of patients (44% accuracy) substantially improving their PROMIS scores. A total of 60%, 75%, and 57% of patients fell into the ambiguous PROMIS score ranges for PF, PI, and Depression, respectively.

Discussion

While a number of site-specific PRO tools have existed for years, the development of PROMIS offers a single universal

Table 7. Chi-Square Analysis of Proportions Applied to Validation Cohort Using Derivation Cohort Cutoffs, Surgery for Traumatic Cause Sample (n = 115).

	Failed to achieve MCID	Achieved MCID	Chi-square statistic	*P* value
PROMIS PF, baseline				
Failed to achieve MCID, > 52.1	25	2		
Ambiguous range, 34.8 – 52.1	44	25		
Achieved MCID, < 34.8	7	12	15.9	<.01
PROMIS PI, baseline				
Failed to achieve MCID, < 46.6	11	1		
Ambiguous range, 46.6 – 69.2	53	33		
Achieved MCID, > 69.2	2	15	20.9	<.01
PROMIS depression, baseline				
Failed to achieve MCID, < 44.0	28	6		
Ambiguous range, 44.0 – 62.2	36	29		
Achieved MCID, > 62.2	9	7	7.4	.02

Note. MCID = minimal clinically important difference; PROMIS = Patient-Reported Outcomes Measurement Information System; PF = Physical Function; PI = Pain Interference.

PRO with the goal that the tool captures patient well-being for all illnesses. PROMIS domains (PF, PI, and Depression) had acceptable *c*-statistics using the derivation cohort, and when applied to the validation cohort, preoperative PROMIS scores were able to predict a select group of patients who had their PROMIS scores substantially improved following hand surgery. For example, our results show that for a notable number of patients, those with preoperative PROMIS PF, PI, and Depression scores above or below determined cutoffs predicted substantial postoperative PROMIS score that may indicate clinical improvement. This approach achieved accuracy in identifying patients likely and unlikely to achieve notable PROMIS score improvement for patients above and below the identified thresholds. It is important to note that these scales are meant to augment clinical decisions, which must also incorporate a variety of additional factors (eg, psychological, financial, social) for accurate preoperative assessment. Furthermore, much more work is needed to determine what is truly clinically substantial improvement from a PROMIS perspective.

For patients falling into specific preoperative PROMIS score ranges, our derivation cohort derived models were fair to good at predicting patients who would and would not substantially improve their PROMIS scores. When the cutoffs were applied to the validation cohort, patients who had poor preoperative PF scores (around one and one half SDs below the PROMIS mean) were more likely to substantially improve their PROMIS scores. Furthermore, we found that patients who had very poor (greater pain and depression) preoperative Pain Interference and Depression scores (ie, around one and one half SDs above the PROMIS PI and Depression means) were more likely to substantially improve their PROMIS scores. Our findings suggest that patients who are severely impacted by their hand illness according to PROMIS scores may demonstrate substantial PROMIS score improvement from surgery. While such a finding may appear intuitive, our work helps to determine previously undetermined cutoffs. Prior work by Ho et al analyzing PROMIS data in a similar way for elective foot and ankle procedures concluded that only those severely impaired by their illness benefited (improvement in PROMIS scores) from surgery.[17] This does not necessarily mean that treatment should be withheld until severe impairment, but when patients present with more advanced disease, they are more likely to have substantial changes in the PROMIS scores. In general, comparable analyses conducted regarding anterior crucuate ligament reconstruction, lower extremity total joint arthroplasty, and shoulder arthroplasty have all found that patients with severe preoperative physical function and pain interference issues experienced greater clinical improvement compared with those with mild impairments, as measured by patients in each study reaching calculated MCID cutoffs.[5,6,26,30,31]

The ability to predict whether patients will substantially improve their PROMIS scores following hand surgery can be used to guide future work aimed at assisting surgeons in setting patient expectations prior to intervention. Our work shows that such a prediction using PROMIS is achievable in a measurable number of patients undergoing surgery for hand and wrist conditions. Although we could not predict score improvement in a majority of patients, our study does demonstrate that the surgeon can use PROMIS data in nearly a quarter of patients to predict accurately substantial postoperative PROMIS score improvement. While we hoped that a greater percentage of patients could have their PROMIS score improvement predicted, this provides evidence that PROMIS offers valuable patient-related insight but continues to have room for improvement as development of the

PRO tool proceeds. Nonetheless, for those select patients in a specific preoperative PROMIS score range, the ability to predict substantial PROMIS score improvement is encouraging. Traditional risk factors considered preoperatively (eg, obesity) are correlated with increased complication rates, but BMI cannot be used to predict a complication with high sensitivity and specificity. As PROMIS domains are improved, the hope is that they can play such a predictive role in outcome predictions most important to patients.

To date, most PRO analyses focus on risk factors associated with complications and readmissions.[20,22,29] However, our work provides additional insight from the patients' perspective. Prior work by Ghomrawi et al demonstrated that surgeons were not able to discriminate between patients who would or would not benefit from total knee arthroplasty.[14] Furthermore, prior research has raised the concern that PROMIS—both PF and UE domains—may provide less information for patients who are highly functioning with their illnesses.[4,16] This study used PROMIS PF scale v1.0. However, a new version of the PROMIS PF scale (v2.0) includes a larger number of items in the higher range of physical function and therefore may yield different results. The alternative is that PROMIS PF (v1.0) is reflecting a lack of improvement. In addition, it is important to note that we were unable to use PROMIS UE is this work, as it was not available at our institution at the time of data collection. However, an understanding of both PROMIS PF and UE domains is valuable in hand and wrist surgery, as not every surgical clinic conducting related procedures will use only PROMIS UE. In addition, PROMIS PF, PI, and DASH are highly correlated.[25] It will be important that future work utilizes the most commonly used PROMIS domain versions in hand clinics to ensure that the research is relevant to practical clinical care.

Our initial thought is that PROs that capture more location- and/or disease-specific findings (eg, upper extremity versus lower extremity) or other patient characteristics, like self-efficacy, may clarify or help predict outcomes for patients in the ambiguous range. A PROMIS UE domain was recently created following the initial release of general domains and is now an area of active research. Beckmann et al analyzed PRO data from 379 upper extremity patients and found that PROMIS UE was strongly correlated with PROMIS PF and DASH questionnaires but had a high ceiling effect.[3] In another study of 84 patients, PROMIS UE was found to have good correlation with the QuickDASH with no ceiling or floor effects.[12] These initial studies with the PROMIS UE domain suggest a strong correlation between it and the PROMIS PF domain; this makes future work analyzing the PROMIS UE domain valuable, as differences found will further inform stakeholders of the potential value and concerns of utilizing a more location-specific PROMIS domain.

Our study must be considered with its limitations in mind. First, preoperative decisions should consider many other patient factors as well as PRO assessment. This study does not evaluate how combinations of preoperative factors including PROs predict PROMIS score improvements. Second, because our patients come from a single, urban tertiary care center, it is possible that our findings may not be generalizable externally. Third, there is selection bias in our patient population, as patients who received hand surgery and were discharged prior to 6 months or those who did not have follow-up appointments were excluded. Some conditions, particularly those associated with trauma, may take greater than 6 months to fully recover following surgery; thus, such patients may not have substantial PROMIS score improvements in our study time frame but could have at a later time.[10,11] Of note, we also included only those traumatic cases that were nonemergent and had a preoperative clinic appointment prior to surgical intervention. In addition, there are a number of diagnoses in our dataset and while all patients analyzed received a hand surgery, there are differences in the recovery following the different hand surgery procedures and the impact the diagnosis may have on pain or function. Those that fully resolve shortly after surgery may bias the sample, as this may exclude patients who are doing well. Many patients with nontraumatic conditions (trigger digits, carpal tunnel, and others) do not return after one or two postoperative visits and thus were not in this cohort. PROMIS is a general outcome measure, and patients may be doing well following their hand surgery procedure but have other medical or personal conditions affecting their score, creating a false impression the procedure did not result in a meaningful change. The variation in timing of follow-up PROMIS PF, PI, and Depression scores may add bias as well. Scores may continue to improve past 6 months, causing some patients with PROMIS PF, PI, and Depression scores measured closer to 1 year to substantially improve clinically when they had not at 8 months. In addition, diagnoses that are considered atraumatic or traumatic are debatable; therefore, there may be some bias in the subgroup analyses. Furthermore, PROMIS may not be sensitive enough to detect change for common hand conditions when compared with other PROs. This may be more relevant to PROMIS PF, which may not capture hand physical function as well as the PROMIS UE domain. However, PROMIS PF remains commonly utilized in hand and other orthopedic clinics. Finally, we defined substantial PROMIS score improvement as reaching or exceeding MCID using the distribution-based method of one half of the SD. There are additional approaches, such as the anchor-based method, which may have slightly altered our findings. In one study of patients with low back pain, an anchor-based approach determined PROMIS PI MCID to likely range from 3.4 to

5.5 *T*-score points.[1] This overlaps with our findings that the MCID for the domains analyzed ranged from 4.5 to 5.8 *T*-score points. Nonetheless, the distribution-based method is commonly used in the literature as an appropriate approximation and a good cutoff to utilize when deciding which PROMIS scores substantially improved or not. While commonly used and agreed upon as a means to estimate clinically substantial improvement following an intervention, ongoing debate remains about how accurate such an approach is in hand surgery. Thus, the focus of this work must be to measure those reaching this traditional MCID milestone but leave the door open for future research to truly delineate the level of clinical change associated with these score changes. Therefore, this work begins the necessary analyses required to better grasp the notable benefits and shortcomings of PROMIS in hand care. We categorized those patients reaching the MCID postoperatively as being patients who substantially improved PROMIS scores with surgical intervention.

Our work offers important insight into the potential benefits of PROMIS in assessing hand and wrist conditions, but must be further researched prior to its use in predicting clinically relevant improvement. Preoperative PROMIS scores in all 3 domains were predictive of substantially improving PROMIS scores or not in a select group of hand or wrist surgery patients. Unfortunately, this is only a small percentage of patients and may be due to multiple factors, including the limits of PROMIS domains to detect meaningful change across the hand surgery continuum. Indeed, many hand conditions differ greatly in their presenting symptoms (eg, physical function, pain); thus, lumping all procedures together with a single MCID may not capture all patients who truly improve. Nonetheless, we still feel this is important work, as it helps surgeons understand the benefits, but also the notable limitations, of PROMIS in hand care as this time. By incorporating our study's findings paired with a surgeon's knowledge of more traditional clinical factor, future research can work to truly determine what PROMIS change denotes "clinical improvement" for patients and implement that knowledge into improved preoperative patient counseling may occur. Once further researched, this understanding can help improve patient care and enhance shared decision-making between surgeons and patients by allowing for more accurate preoperative surgical outcome expectation setting by the surgeon prior to hand or wrist surgery. Future work can examine whether our findings are replicable in a variety of individualized hand pathologies. Furthermore, a better understanding of how domains may or may not interact with each other could offer additional valuable clinical insight. Also, additional research can help determine whether substantial clinical improvement in PROMIS PF, PI, and Depression domains correlate to patient satisfaction. The current study analyzing PROMIS score improvements assists in laying the groundwork for predicting clinical improvement and setting patient expectations, which will hopefully translate to improved patient satisfaction following surgery.

Acknowledgments

The authors thank Allison McIntyre, MPH, Chris DaSilva, BS, and the University of Rochester Health Lab for their assistance with collection and reporting of patient-reported outcomes via UR VOICE platform.

Ethical Approval

The institutional review board of the University of Rochester Medical Center approved this study.

Statement of Human and Animal Rights

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008.

Statement of Informed Consent

This is a database review. We have IRB approval for review of the data in the hand repository but individual patient consent is not required.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: D.N.B. is a recipient of Alpha Omega Alpha Carolyn L. Kuckein Student Research Fellowship ($5000). All other authors (J.R.H, R.M.G., D.M.W., R.J.M., D.J.M., W.C.H.) certify that he or she has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: David N. Bernstein, MBA, MA, has received grants from Alpha Omega Alpha, outside of the submitted work.

ORCID iD

DN Bernstein https://orcid.org/0000-0002-1784-3288

References

1. Amtmann D, Kim J, Chung H, et al. Minimally important differences for Patient Reported Outcomes Measurement Information System pain interference for individuals with back pain. *J Pain Res*. 2016;9:251-255.
2. Anderson MR, Houck JR, Saltzman CL, et al. Validation and generalizability of preoperative PROMIS scores to predict postoperative success in foot and ankle patients. *Foot Ankle Int*. 2018;39:763-770.

3. Beckmann JT, Hung M, Voss MW, et al. Evaluation of the patient-reported outcomes measurement information system upper extremity computer adaptive test. *J Hand Surg Am*. 2016;41:739-744.e4.
4. Beleckas CM, Padovano A, Guattery J, et al. Performance of Patient-Reported Outcomes Measurement Information System (PROMIS) Upper Extremity (UE) versus Physical Function (PF) Computer Adaptive Tests (CATs) in upper extremity clinics. *J Hand Surg Am*. 2017;42:867-874.
5. Berliner JL, Brodke DJ, Chan V, et al. Can preoperative patient-reported outcome measures be used to predict meaningful improvement in function after TKA? *Clin Orthop Relat Res*. 2017;475:149-157.
6. Berliner JL, Brodke DJ, Chan V, et al. John Charnley Award: preoperative patient-reported outcome measures predict clinically meaningful improvement in function after THA. *Clin Orthop Relat Res*. 2016;474:321-329.
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;63:1179-1194.
8. Cella D, Yount S, Rothrock N, 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:S3-S11.
9. Choi SW, Schalet B, Cook KF, et al. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. *Psychol Assess*. 2014;26:513-527.
10. Chung KC, Kotsis SV, Kim HM. Predictors of functional outcomes after surgical treatment of distal radius fractures. *J Hand Surg Am*. 2007;32:76-83.
11. Chung KC, Squitieri L, Kim HM. Comparative outcomes study using the volar locking plating system for distal radius fractures in both young adults and adults older than 60 years. *J Hand Surg Am*. 2008;33:809-819.
12. Doring AC, Nota SP, Hageman MG, et al. Measurement of upper extremity disability using the Patient-Reported Outcomes Measurement Information System. *J Hand Surg Am*. 2014;39:1160-1165.
13. Gausden EB, Levack AE, Sin DN, et al. Validating the Patient Reported Outcomes Measurement Information System (PROMIS) computerized adaptive tests for upper extremity fracture care. *J Shoulder Elbow Surg*. 2018;27:1191-1197.
14. Ghomrawi HMK, Mancuso CA, Dunning A, et al. Do surgeon expectations predict clinically important improvements in WOMAC scores after THA and TKA? *Clin Orthop Relat Res*. 2017;475:2150-2158.
15. Gortmaker S, Hosmer D, Lemeshow S. Applied logistic regression. *Contemp Sociol*. 1994;23:159.
16. Hays RD, Liu H, Spritzer K, et al. Item response theory analyses of physical functioning items in the medical outcomes study. *Med Care*. 2007;45:S32-S38.
17. Ho B, Houck JR, Flemister AS, et al. Preoperative PROMIS scores predict postoperative success in foot and ankle patients. *Foot Ankle Int*. 2016;37:911-918.
18. Hung M, Franklin JD, Hon SD, et al. Time for a paradigm shift with computerized adaptive testing of general physical function outcomes measurements. *Foot Ankle Int*. 2014;35: 1-7.
19. Jones RS, Stukenborg GJ. Patient-Reported Outcomes Measurement Information System (PROMIS) use in surgical care: a scoping study. *J Am Coll Surg*. 2017;224: 245-254.
20. Katz JN, Losina E, Amick BC, et al. Predictors of outcomes of carpal tunnel release. *Arthritis Rheum*. 2001;44:1184-1193.
21. Kortlever JT, Janssen SJ, van Berckel MM, et al. What is the most useful questionnaire for measurement of coping strategies in response to nociception? *Clin Orthop Relat Res*. 2015;473:3511-3518.
22. Leit ME, Weiser RW, Tomaino MM. Patient-reported outcome after carpal tunnel release for advanced disease: a prospective and longitudinal assessment in patients older than age 70. *J Hand Surg Am*. 2004;29:379-383.
23. Menendez ME, Bot AG, Hageman MG, et al. Computerized adaptive testing of psychological factors: relation to upper-extremity disability. *J Bone Joint Surg Am*. 2013;95:e149.
24. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41: 582-592.
25. Overbeek CL, Nota SP, Jayakumar P, et al. The PROMIS physical function correlates with the QuickDASH in patients with upper extremity illness. *Clin Orthop Relat Res*. 2015;473:311-317.
26. Papuga MO, Beck CA, Kates SL, et al. Validation of GAITRite and PROMIS as high-throughput physical function outcome measures following ACL reconstruction. *J Orthop Res*. 2014;32:793-801.
27. Papuga MO, Mesfin A, Molinari R, et al. Correlation of PROMIS physical function and pain CAT instruments with Oswestry Disability Index and Neck Disability Index in Spine patients. *Spine (Phila Pa 1976)*. 2016;41:1153-1159.
28. Tyser AR, Beckmann J, Franklin JD, et al. Evaluation of the PROMIS physical function computer adaptive test in the upper extremity. *J Hand Surg Am*. 2014;39:2047-2051.e4.
29. Werner BC, Boatright JD, Chhabra AB, et al. Trigger digit release: rates of surgery and complications as indicated by a United States Medicare database. *J Hand Surg Eur Vol*. 2016;41:970-976.
30. Werner BC, Chang B, Nguyen JT, et al. What change in American shoulder and elbow surgeons score represents a clinically important change after shoulder arthroplasty? *Clin Orthop Relat Res*. 2016;474:2672-2681.
31. Wong SE, Zhang AL, Berliner JL, et al. Preoperative patient-reported scores can predict postoperative outcomes after shoulder arthroplasty. *J Shoulder Elbow Surg*. 2016;25: 913-919.
32. Yaffe M, Goyal N, Kokmeyer D, et al. The use of an iPad to collect patient-reported functional outcome measures in hand surgery. *Hand (N Y)*. 2015;10:522-528.