J Shoulder Elbow Surg (2019) 28, e40–e48

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

The influence of mental health on Patient-Reported Outcomes Measurement Information System (PROMIS) and traditional outcome instruments in patients with symptomatic glenohumeral arthritis

Eitan M. Kohan, MD*, J. Ryan Hill, MD, Maria Schwabe, BS, Alexander W. Aleem, MD, Jay D. Keener, MD, Aaron M. Chamberlain, MD

Department of Orthopaedic Surgery, Washington University in St Louis, St Louis, MO, USA

Background: The Patient-Reported Outcomes Measurement Information System (PROMIS) assessment includes computerized adaptive tests (CATs) that assess function, pain, depression, and anxiety. The influence of mental health on patients' self-reported pain and function has not been explored using PROMIS in patients with symptomatic glenohumeral osteoarthritis.

Methods: This cross-sectional study included 284 shoulders in 276 patients presenting with isolated glenohumeral osteoarthritis. All patients completed the American Shoulder and Elbow Surgeons (ASES) score, Simple Shoulder Test (SST), Visual Analog Pain Scale (VAS), and PROMIS CATs at the time of presentation. PROMIS anxiety and depression scores were converted into Generalized Anxiety Disorder-7 and Patient Health Questionnaire-9 scores, respectively, using the PROsetta Stone "crosswalk" tool. Mean pain and functional scores were compared between patients with and without PROMIS-converted scores corresponding to a diagnosis of anxiety or depression, as well as between scores corresponding to varying degrees of anxiety or depression.

Results: Patients with scores corresponding to a diagnosis of anxiety or depression reported lower functional and higher pain scores compared to those with scores in the normal range ($P < .001$). Analysis of variance showed progressively lower functional and higher pain scores as anxiety severity increased ($P < .001$). Similar results were seen with ASES, upper extremity CAT, and pain scores as depression severity increased ($P < .001$). Functional ASES ($P = .004$), SST ($P < .001$), and physical function CAT ($P = .002$) scores were statistically significantly lower in patients with moderate to severe depression than those without depression or with mild depression.

Discussion: In patients with glenohumeral osteoarthritis, PROMIS-reported anxiety and depression scores, particularly in those with moderate-to-severe scores, correlate with lower functional and higher pain scores. Further investigation is necessary to examine the influence that mental health has on outcomes after operative intervention in this population.

This study has been approved under Washington University in St Louis Institutional Review Board protocol 201611040.

*Reprint requests: Eitan M. Kohan, MD, Department of Orthopaedic Surgery, Washington University in St Louis, 660 S Euclid Ave, Campus Box 8233, St Louis, MO 63110, USA.

E-mail address: kohane@wustl.edu (E.M. Kohan).

1058-2746/$ - see front matter
https://doi.org/10.1016/j.jse.2018.07.033

Level of evidence: Basic Science Study; Development or Validation of Outcome Instruments

Keywords: Glenohumeral arthritis; shoulder; outcomes; PROMIS; depression; anxiety; function; mental health

Clinical outcome measures in patients with shoulder pathology have traditionally focused on parameters that are objectively determined by the physician, such as range of motion or radiographic appearance. Over the past 2 decades, there has been increased focus on understanding patients' perceptions of their illness and its effect on their quality of life.[1] This focus on patient-reported outcomes (PROs) has led to the development of over 30 outcome measure instruments for shoulder pathology alone.[32,33] No individual patient-reported outcome measure (PROM) is comprehensive in assessing all outcomes for all shoulder conditions. In addition, these outcome measures may have large floor and ceiling effects,[26] making them unable to distinguish between patients at the highest and lowest ends of the scale. Thus, patients are often asked to complete multiple PROMs. High numbers of questions coupled with redundant questions across PROMs can lead to increased responder burden, which can lead to incomplete or inaccurate responses from patients.[20]

The Patient-Reported Outcomes Measurement Information System (PROMIS) was developed by the National Institutes of Health to address many of the limitations of traditional PROMs.[6,7] PROMIS is a method of assessing PROs using computerized adaptive testing, which uses item-response theory to prompt patients with sequential questions from a standardized question bank.[14] Each response adjusts the probability curve of the patient's estimated ability. For example, if a patient is able to throw a ball, he or she has a higher probability of performing toward the upper end of physical function (PF). Subsequent questions will narrow the patient's placement within the "upper end" area of the scale. PROMIS instruments have been developed to assess function (including PF and upper extremity [UE]), pain (pain interference [PI]), and mental health (depression and anxiety) in addition to other domains of health.

The PROMIS PF computerized adaptive test (CAT) has previously been shown to have a moderate to high correlation with traditional PROMs including the quick version of the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire; the DASH questionnaire; the American Shoulder and Elbow Surgeons (ASES) score; and the Simple Shoulder Test (SST) in patients with rotator cuff disease, as well as non-shoulder UE complaints.[3,15,24,30] The floor and ceiling effects of the PF CAT were decreased compared with traditional outcome instruments. The UE CAT was developed to isolate upper extremity physical function and has been shown to highly correlate with the PF CAT and DASH questionnaire in non-shoulder upper extremity patients, albeit with a relatively large ceiling effect.[4]

The data comparing PROMIS with established PROMs are limited in patients presenting with glenohumeral osteoarthritis (GHOA). Only 2 studies have evaluated the correlation of PROMIS instruments with traditional scales in this population. One study showed the PF CAT and UE CAT had a moderate correlation with the ASES score, Short Form 36 PF scale, and Western Ontario Osteoarthritis of the Shoulder index.[16] The other study showed the UE CAT had a moderate correlation with the ASES score and SST.[23]

Recently, the National Institutes of Health has developed CATs (PROMIS scores) to assess various facets of mental health as well, specifically anxiety and depression. Given the known influence of mental health on various disease states, including in patients with upper extremity conditions, there exists a need to define these potential influences on patient perceptions of pain and function in GHOA.[10,11,22,25] The purposes of this study were to explore the correlation between the PF CAT and UE CAT with established PROMs in patients presenting with GHOA and to evaluate the influence that mental health has on patient-perceived pain and function, as measured by PROMIS.

Materials and methods

Patient selection

This study was a retrospective cross-sectional case series. The clinic visits of our high-volume regional-referral shoulder arthroplasty center from August 2015 to August 2017 were reviewed. Patients with an isolated *International Classification of Diseases, Tenth Revision* (ICD-10) code of M19.011 (primary osteoarthritis, right shoulder), M19.012 (primary osteoarthritis, left shoulder), or M19.019 (primary osteoarthritis, unspecified shoulder) were included for review. Clinic notes and radiographs were reviewed to ensure that patients had an isolated diagnosis of GHOA. The exclusion criteria were patients with coexisting diagnoses (eg, rotator cuff tear, rheumatoid arthritis, or adhesive capsulitis), traditional outcome scores collected more than 1 year before PROMIS scores, and incomplete medical records.

Data collection

All patient information was acquired at the initial clinic visit. Patient demographic characteristics collected included sex, age, body mass index, Charlson Comorbidity Index,[8] smoking status, and laterality of GHOA. Traditional PROMs collected included a visual analog scale (VAS) for pain, the ASES score, and the SST. The ASES score was separated into its total and functional components for assessment.

PROMIS measures collected included the PF, UE, depression, anxiety, and PI CAT instruments. PROMIS instruments are administered as CATs. The response to the first prompt guides the system's

choice of the next question. The CAT continues until the standard error drops below a T-score metric of 3.0 or the patient has answered 12 questions, whichever occurs first.[35] PROMIS instruments are calibrated with a score of 50 as the average for the US general population with a standard deviation of 10. A higher PROMIS score corresponds to more of the concept being measured (eg, a higher PI CAT score translates to more pain and a lower PROMIS depression score translates to less depression).

By use of the PROsetta Stone linking table, preoperative PROMIS depression scores were transformed into Patient Health Questionnaire-9 (PHQ-9) scores using the previously validated "crosswalk" tool.[13,28] PHQ-9 scores of 0-4, 5-9, 10-14, 15-19, and 20 or greater represent no depression, mild depression, moderate depression, moderately severe depression, and severe depression, respectively. For statistical analysis, the latter 3 categories were grouped as moderate to severe depression.

By use of the corresponding linking table, preoperative PROMIS anxiety scores were transformed into Generalized Anxiety Disorder-7 (GAD-7) scores.[12] GAD-7 scores of 0-4, 5-9, 10-14, and 15 or greater represent no anxiety, mild anxiety, moderate anxiety, and severe anxiety, respectively. For statistical analysis, the latter 2 categories were grouped as moderate to severe anxiety.

Statistical analysis

No a priori sample size determination was performed as this was a retrospective study and all available patients were included. Data normality was assessed using the Shapiro-Wilk test. Correlations between the ASES, SST, UE CAT, and PF CAT scores were determined using Pearson correlation coefficients. The correlations were defined as excellent (>0.7), moderate (0.5-0.7), low-moderate (0.3-0.5), or poor (<0.3). These correlation values are based on previous studies evaluating the correlation of PROMs.[3] Ceiling and floor effects for each measure were determined by the number of responses that approached the maximum and minimum attainable scores on each scale.

Equality of variance between groups was assessed with the Levene test. Independent-samples *t* testing was performed between patients with and without depression or anxiety to assess for differences in patient-reported pain and function, as measured by traditional scales and PROMIS. The same outcome measures were assessed among patients with varying degrees of depression and anxiety as measured by the PHQ-9 and GAD-7, respectively, using 1-way analysis of variance (ANOVA) testing with a post hoc Tukey analysis. Correlation between anxiety and depression was determined using the Pearson correlation coefficient.

Results

Patient demographic characteristics

During the study period, 1497 patients presented to our center with an ICD-10 code corresponding to GHOA. After we excluded those with concomitant diagnoses as well as those with incomplete survey data, 284 shoulders in 276 patients were eligible for inclusion. The reasons for exclusion are shown in Figure 1. Demographic characteristics are shown in Table I.

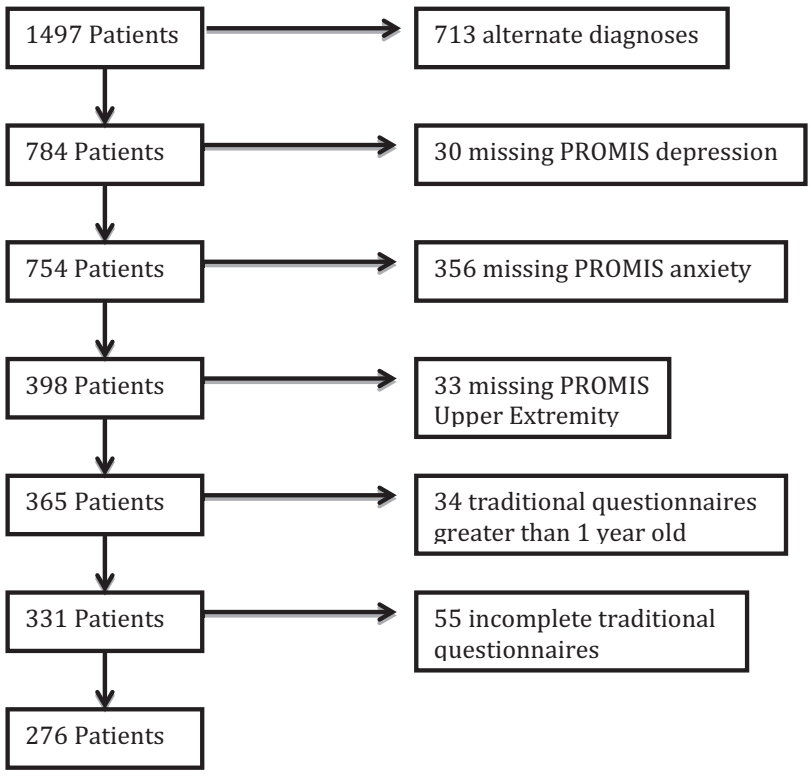

Figure 1 Exclusion criteria. *PROMIS*, Patient-Reported Outcomes Measurement Information System.

Table I Patient demographic characteristics

	Data
Sex, n	
Male	164 (59.4%)
Female	112 (40.6%)
Age, mean ± SD	65.1 ± 11.2
Laterality, n	
Right	150 (52.8%)
Left	134 (47.2%)
Dominant extremity involved	51.4%
Body mass index, mean ± SD	31.3 ± 7.3
Charlson Comorbidity Index, mean ± SD	0.61 ± 1.1
Smoking status	
Active	5.1%
Former	36.2%
Nonsmoker	58.7%

SD, standard deviation.

Functional outcome correlations

Means and standard deviations for total ASES, functional ASES, SST, PF CAT, and UE CAT scores are shown in Table II. The PROMIS PF CAT score showed a low-moderate correlation with both total ASES ($r = 0.42$ [95% confidence interval (CI), 0.32-0.51]; $P < .001$) and functional ASES ($r = 0.45$ [95% CI, 0.35-0.54]; $P < .001$) scores and a moderate correlation with SST score ($r = 0.53$ [95% CI, 0.44-0.61]; $P < .001$). The PROMIS UE CAT score showed a moderate correlation with total ASES ($r = 0.61$ [95% CI, 0.53-0.68]; $P < .001$), functional ASES ($r = 0.65$ [95% CI, 0.58-0.71]; $P < .001$), and SST ($r = 0.64$ [95% CI, 0.57-0.70]; $P < .001$) scores. Correlations are shown in Supplementary Figs. S1-S6.

There were no significant ceiling effects identified with the ASES and SST scores. The floor effect was 1.4% for total ASES score, 2.5% for functional ASES score, and 5.6% for SST score. There were no significant ceiling or floor effects for either the PF CAT or UE CAT.

Depression and anxiety analysis

The overall mean PROMIS depression score was 46.7 ± 8.9. There were 183 patients without depression (mean PROMIS depression score, 41.4 ± 5.7) and 101 patients with depression (mean PROMIS depression score, 56.3 ± 4.3), as measured by PROMIS depression conversion to the PHQ-9. The overall mean PROMIS anxiety score was 51.9 ± 9.2. There were 152 patients without anxiety (mean PROMIS anxiety score, 45.1 ± 5.9) and 132 patients with anxiety (mean PROMIS anxiety score, 59.7 ± 5.3), as measured by PROMIS anxiety conversion to the GAD-7. The PROMIS depression and anxiety scores had a high correlation ($r = 0.760$, $P < .001$).

Of the patients with depression, 74 had mild depression (mean PROMIS depression score, 54.1 ± 1.9) and 27 had moderate to severe depression (mean PROMIS depression score, 62.4 ± 3.1). There was no significant ceiling effect seen with PROMIS depression, but 20.4% of patients had a score correlating with the floor effect. The scores that create this large floor effect were previously shown to be associated with hasty completion of the questionnaire (4 ± 3 seconds), indicating that the resulting scores are likely not representative of the patients' true level of depression.[18] Accordingly, statistical analysis was performed both including and excluding those with the lowest PROMIS depression score. After exclusion of those with the lowest depression score, there were 125 patients without depression (mean PROMIS depression score, 44.7 ± 3.5).

In all analyses, patients with depression had lower total ASES, functional ASES, SST, PF CAT, and UE CAT scores and higher VAS and PI CAT scores than those without depression. Descriptive statistics and *t* test results are shown in Supplementary Tables S1 and S2.

When all patients were included in the analysis, there was a statistically significant difference between levels of depression for all outcome measures as determined by ANOVA testing. The results of the Tukey post hoc test are shown in Table III. We found statistically significantly lower total ASES, PF CAT, and UE CAT scores as depression increased. There was no statistically significant difference in functional ASES or SST scores between patients without depression and those with mild depression, but those with moderate to severe depression had significantly lower scores on both measures. Pain, as measured by the VAS and PI CAT, was statistically significantly higher in patients with mild depression and moderate to severe depression than in those without depression. The VAS showed no statistically significant difference in scores between patients with mild depression and those with moderate to severe depression, although a difference was seen between these groups on the PI CAT.

ANOVA testing and the Tukey post hoc test showed overall similar patterns to the inclusive analysis when the lowest depression scores were excluded. The results are shown in Table IV. As depression increased, total ASES and UE CAT scores decreased. There was no statistically significant difference in functional ASES, SST, or PF CAT scores between patients without depression and those with mild depression, but all 3 measures were significantly lower in those with moderate to severe depression. The pain instruments showed the

Table II Descriptive statistics of functional outcome measures

Outcome measure	Mean ± SD
ASES score	35.6 ± 16.8
ASES function score	19.8 ± 9.9
SST score	4.4 ± 2.8
PF CAT score	41.8 ± 8.1
UE CAT score	32.8 ± 6.9

SD, standard deviation; *ASES*, American Shoulder and Elbow Surgeons; *SST*, Simple Shoulder Test; *PF*, physical function; *CAT*, computerized adaptive test; *UE*, upper extremity.

Table III Tukey post hoc analysis based on depression including depression floor

Outcome measure and depression group	n	Mean	SD	95% CI for mean difference		Tukey HSD comparison, *P* value	
				None	Mild	None	Mild
Total ASES score							
None	183	38.7	16.8				
Mild	74	32.8	15.2	0.7-11.1		.02	
Mod-sev	27	21.6	13.2	9.3-24.9	2.7-19.7	<.001	.006
ASES function score							
None	183	21.2	10.0				
Mild	74	19.1	8.8	−1.0 to 5.2		.26	
Mod-sev	27	12.3	7.7	4.2-13.5	1.7-11.8	<.001	.005
SST score							
None	183	4.7	2.9				
Mild	74	4.4	2.6	−0.6 to 1.2		.72	
Mod-sev	27	2.1	2.1	1.2-3.9	0.8-3.7	<.001	.001
VAS score							
None	183	6.5	2.1				
Mild	74	7.3	1.7	−1.4 to −0.1		.01	
Mod-sev	27	8.1	1.8	−2.6 to −0.7	−1.9 to 0.1	<.001	.11
PI CAT score							
None	183	59.7	5.3				
Mild	74	62.8	5.8	−4.8 to −1.3		<.001	
Mod-sev	27	66.3	5.5	−9.2 to −3.9	−6.4 to −0.6	<.001	.01
PF CAT score							
None	183	43.4	7.7				
Mild	74	40.5	7.5	0.4-5.4		.02	
Mod-sev	27	34.5	8.1	5.2-12.7	1.9-10.1	<.001	.002
UE CAT score							
None	183	34.1	6.5				
Mild	74	31.7	6.6	0.3-4.6		.02	
Mod-sev	27	27.1	6.9	3.9-10.3	1.1-8.1	<.001	.006

SD, standard deviation; *CI*, confidence interval; *HSD*, honestly significant difference; *ASES*, American Shoulder and Elbow Surgeons; *Mod-sev*, moderate to severe; *SST*, Simple Shoulder Test; *VAS*, visual analog scale; *PI*, pain interference; *CAT*, computerized adaptive test; *PF*, physical function; *UE*, upper extremity.

same findings as the inclusive analysis: higher pain scores in patients with depression than in those without depression on both measures and higher pain scores in those with moderate to severe depression than in those with mild depression on the PI CAT but no difference between those with mild pain and those with moderate to severe pain on the VAS.

Of the patients with anxiety, 87 had mild anxiety (mean PROMIS anxiety score, 56.4 ± 2.2) and 45 had moderate to severe anxiety (mean PROMIS anxiety score, 66.0 ± 3.5). There was no significant ceiling or floor effect seen with the PROMIS anxiety scale, so all patients were included for analysis.

Similar to the depression analysis, patients with anxiety had lower total ASES, functional ASES, SST, PF CAT, and UE CAT scores and higher VAS and PI CAT scores than those without anxiety. Descriptive statistics and *t* test results are shown in Supplementary Table S3.

There was a statistically significant difference between levels of anxiety for all outcome measures as determined by ANOVA testing. The results of the Tukey post hoc test are shown in Table V. Across all functional outcome measures, there were statistically significantly lower scores as grade of anxiety increased. There was no statistically significant difference in VAS scores between patients without anxiety and those with mild anxiety. There were statistically significantly increased VAS pain scores in the moderate to severe anxiety group compared with patients without anxiety or with mild anxiety, as well as across all levels of anxiety for the PI CAT.

Discussion

PROMIS CATs have been designed to accurately and efficiently assess various PRO domains. The PROMIS UE, PF, and PI CATs are reliable and consistent. Traditional PROMs including the ASES score and SST are also reliable and consistent in the setting of various pathologic conditions affecting the shoulder.[2,21] PROMIS CATs correlate with the more traditionally used ASES and SST measures in the setting of rotator cuff disease while requiring less time and having a lower responder burden.[3] The PROMIS UE CAT also showed

Table IV Tukey post hoc analysis based on depression excluding depression floor

Outcome measure and depression group	n	Mean	SD	95% CI for mean difference		Tukey HSD comparison, *P* value	
				None	Mild	None	Mild
Total ASES score							
None	125	39.1	16.6				
Mild	74	32.8	15.2	0.8-11.7		.02	
Mod-sev	27	21.6	13.2	9.6-25.3	2.9-19.6	<.001	.005
ASES function score							
None	125	21.0	9.9				
Mild	74	19.1	8.8	−1.3 to 5.1		.34	
Mod-sev	27	12.3	7.7	4.0-13.3	1.8-11.7	<.001	.004
SST score							
None	125	4.8	2.9				
Mild	74	4.4	2.6	−0.5 to 1.3		.58	
Mod-sev	27	2.1	2.1	1.3-4.0	0.8-3.7	<.001	.001
VAS score							
None	125	6.4	2.0				
Mild	74	7.3	1.7	−1.5 to −0.2		.005	
Mod-sev	27	8.1	1.8	−2.7 to −0.8	−1.8 to 0.1	<.001	.09
PI CAT score							
None	125	60.0	5.0				
Mild	74	62.8	5.8	−4.6 to −0.9		.002	
Mod-sev	27	66.3	5.5	−9.0 to −3.6	−6.4 to −0.7	<.001	.01
PF CAT score							
None	125	42.9	7.8				
Mild	74	40.5	7.5	−0.3 to 5.0		.09	
Mod-sev	27	34.5	8.1	4.5-12.2	1.9-10.1	<.001	.002
UE CAT score							
None	125	34.2	6.1				
Mild	74	31.7	6.6	0.4-4.8		.02	
Mod-sev	27	27.1	6.9	4.0-10.4	1.2-8.0	<.001	.004

SD, standard deviation; *CI*, confidence interval; *HSD*, honestly significant difference; *ASES*, American Shoulder and Elbow Surgeons; *Mod-sev*, moderate to severe; *SST*, Simple Shoulder Test; *VAS*, visual analog scale; *PI*, pain interference; *CAT*, computerized adaptive test; *PF*, physical function; *UE*, upper extremity.

a moderate correlation with ASES and SST scores in the setting of GHOA in a smaller patient cohort.[23] Our study showed a low-moderate correlation between the PROMIS PF CAT and the ASES and SST scores and a moderate correlation between the PROMIS UE CAT and the ASES and SST scores. Although both tools were previously reported to be acceptable to use in patients with shoulder arthritis, the UE CAT appears to better correlate with upper extremity function in this population, as measured by traditional scales.

Although all PROMIS instruments are designed to be standardized to a mean of 50 with a standard deviation of 10 in a normative population, the UE CAT has previously been shown to produce a non-normal distribution with a maximum achievable score of 56.4.[5] This can create a large ceiling effect, limiting the utility of the UE CAT in a young, healthy population.[4] Although our results report the same maximum achievable score, only 3 of 276 patients achieved this score. Despite the known limitation of the instrument, patients with GHOA likely have decreased isolated upper extremity function and are unlikely to achieve scores close to this ceiling.

When both the UE CAT and PF CAT were administered to patients with GHOA, they indicated differential levels of absolute impairment. In our population of 276 patients, UE scores consistently indicated greater disability compared with the PF CAT. This finding is consistent with findings in a previously published study of PROMIS scores in over 5000 patients with various upper extremity complaints.[5] Döring et al[15] found a similar direction of disparity within 84 patients attending a hand clinic, with the UE score being on average 10 points lower than the PF score. The mean difference of 8 points between the UE and PF scores found in our study is presumably clinically relevant with an effect size for that difference of 0.8, suggesting function nearly a full standard deviation worse on the UE CAT relative to the PF CAT.

The influence that mental health has on patients' perceptions of their disease, as measured by PROMs, has become increasingly apparent. In the setting of rotator cuff disease, mental health scores had a higher correlation with pain and functional outcome measures than either tear severity

Table V Tukey post hoc analysis based on anxiety

Outcome measure and anxiety group	n	Mean	SD	95% CI for mean difference		Tukey HSD comparison, *P* value	
				None	Mild	None	Mild
Total ASES score							
None	152	40.9	17.3				
Mild	87	33.5	12.6	2.5-12.3		.001	
Mod-sev	45	21.5	13.5	13.2-25.5	5.3-18.6	<.001	<.001
ASES function score							
None	152	22.9	10.2				
Mild	87	18.4	7.4	1.7-7.4		.001	
Mod-sev	45	11.9	7.5	7.4-14.7	2.6-10.4	<.001	<.001
SST score							
None	152	5.2	3.0				
Mild	87	3.9	2.3	0.4-2.1		.002	
Mod-sev	45	2.3	2.2	1.7-3.9	0.4-2.7	<.001	.004
VAS score							
None	152	6.4	2.1				
Mild	87	7.0	1.7	−1.1 to 0.4		.075	
Mod-sev	45	8.1	1.8	−2.4 to −0.9	−1.9 to −0.3	<.001	.007
PI CAT score							
None	152	58.9	5.3				
Mild	87	62.0	5.1	−4.8 to −1.6		<.001	
Mod-sev	45	67.1	4.0	−10.2 to −6.2	−7.3 to −2.9	<.001	<.001
PF CAT score							
None	152	44.7	7.4				
Mild	87	40.3	7.3	2.0-6.6		<.001	
Mod-sev	45	34.9	7.0	6.9-12.7	2.3-8.7	<.001	<.001
UE CAT score							
None	152	35.3	6.5				
Mild	87	31.9	6.0	1.4-5.3		<.001	
Mod-sev	45	26.4	5.2	6.4-11.3	2.8-8.1	<.001	<.001

SD, standard deviation; *CI*, confidence interval; *HSD*, honestly significant difference; *ASES*, American Shoulder and Elbow Surgeons; *Mod-sev*, moderate to severe; *SST*, Simple Shoulder Test; *VAS*, visual analog scale; *PI*, pain interference; *CAT*, computerized adaptive test; *PF*, physical function; *UE*, upper extremity.

or tear size.[34] Low resilience, a psychometric property characterized by an ability to cope with stress, predicted poorer postoperative outcomes after total shoulder arthroplasty.[29] This is the first study to use PROMIS tools to evaluate the correlation between mental health and traditional outcome scores in the symptomatic glenohumeral arthritis population.

In our cohort, in patients with PROMIS depression scores that correlated with PHQ-9 scores of varying degrees of depression, we noted correlations with lower functional scores and higher pain scores. This inverse relationship was especially significant when we analyzed patients with more severe depressive symptoms; the SST and PF CAT instruments showed no statistical difference between patients with mild depression and those with no depression, but functional scores were significantly lower across all instruments in patients with moderate to severe depressive symptoms.

Similar to our depression findings, functional outcome and pain scores were significantly worse in patients with PROMIS anxiety scores that correlated with a diagnosis of anxiety. Those with more severe anxiety symptoms reported worse functional outcome and pain scores. PROMIS depression and PROMIS anxiety scores correlated highly.

With large patient cohorts, it is important to distinguish between statistical significance and clinical relevance. The minimal clinically important difference (MCID) is often used as a threshold to determine clinically relevant differences. Although designed as an instrument to compare differences within groups of patients before and after an intervention or over time, in a cross-sectional analysis, the MCID can be used as a proxy to estimate clinical relevance between groups. Prior studies of shoulder arthritis populations have identified MCIDs of 13.6 points for the total ASES score, 6.5 points for the functional ASES score, 1.5 points for the SST score, and 1.6 points for the VAS score.[27,31] No MCID has been identified for the PROMIS instruments in a shoulder arthritis population, although several studies have identified MCIDs in the range of 3.5-5.0 points for the PROMIS instruments.[9,17,19]

Using these MCID values as proxies for clinical relevance shows that despite the statistically significant differences

between patients with and without depression or anxiety, the values are not likely to be clinically relevant. However, subgroup analysis shows that patients with moderate to severe anxiety or depression reported lower functional scores and higher pain scores that were not only statistically significant but also clinically relevant.

It should be noted that PROMIS depression and anxiety scores have not been validated as instruments to diagnose clinical depression or anxiety. Rather, we used the crosswalk to the PHQ-9 and GAD-7 in this study as an instrument to attempt to correlate PROMIS depression and anxiety scores with degrees of depression severity as defined by the PHQ-9 and anxiety as defined by the GAD-7.

Our study benefits from the relatively large number of patients included in the cohort. To date, this is the largest study evaluating PROMIS in patients with GHOA. In addition, to our knowledge, this is the first study using PROMIS outcomes in mental health domains to evaluate correlations of mental health states with traditional outcome scores in the patient population with symptomatic glenohumeral arthritis.

Study limitations include the retrospective design and cross-sectional analysis. The retrospective nature required identification of patients via an isolated ICD-10 code corresponding to GHOA. This may have led to the exclusion of patients whose visits were not coded appropriately. The cross-sectional analysis allows us to determine association but not causation. Without an objective physical examination and radiographic analysis, it is impossible to differentiate whether worse mental health is related to a more severe GHOA disease process or whether the patients' anxiety and depressive symptoms create a more intense perception of their limitations. In addition, we are unable to draw conclusions on responsiveness to treatment and how mental health may affect outcomes after operative intervention.

Further investigations currently under way are evaluating the effect that preoperative mental health has on electing to undergo operative intervention for symptomatic glenohumeral arthritis; the influence of mental health on postoperative outcomes; and whether operative intervention causes any postoperative change in mental health, as measured by PROMIS.

Conclusion

PROMIS UE and PF CATs show acceptable correlations with traditionally used functional outcome scores including ASES and SST scores in patients with symptomatic GHOA, although the UE CAT appears to correlate slightly better with traditional scores in this population. After we controlled for relevant patient comorbidities and demographic factors, patients with moderate to severe depression and anxiety symptoms score significantly lower on the ASES and SST measures and report significantly higher VAS and PROMIS pain scores.

Disclaimer

Jay D. Keener is a paid consultant for Arthrex and Wright Medial/Tornier, receives research support from the National Institutes of Health and Zimmer/Biomet, receives royalties from Shoulder Innovations and Imascap, and is on the editorial board of the *Journal of Shoulder and Elbow Surgery*.

Aaron M. Chamberlain is a paid consultant for DePuy Synthes, Wright Medical, and Arthrex and received research support from the National Institutes of Health, Zimmer/Biomet, and the Orthopaedic Research and Education Foundation.

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

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jse.2018.07.033.

References

1. Bayley KB, London MR, Grunkemeier GL, Lansky DJ. Measuring the success of treatment in patient terms. Med Care 1995;33(Suppl.): AS226-35.
2. Beaton D, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. J Shoulder Elbow Surg 1998;7:565-72.
3. Beckmann JT, Hung M, Bounsanga J, Wylie JD, Granger EK, Tashjian RZ. Psychometric evaluation of the PROMIS Physical Function Computerized Adaptive Test in comparison to the American Shoulder and Elbow Surgeons score and Simple Shoulder Test in patients with rotator cuff disease. J Shoulder Elbow Surg 2015;24:1961-7. http://dx.doi.org/10.1016/j.jse.2015.06.025
4. Beckmann JT, Hung M, Voss MW, Crum AB, Bounsanga J, Tyser AR. Evaluation of the patient-reported outcomes measurement information system upper extremity computer adaptive test. J Hand Surg Am 2016;41:739-44.e4. http://dx.doi.org/10.1016/j.jhsa.2016.04.025
5. Beleckas CM, Padovano A, Guattery J, Chamberlain AM, Keener JD, Calfee RP. 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;42:867-74. http://dx.doi.org/10.1016/j.jhsa.2017.06.012
6. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, 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;63:1179-94. http://dx.doi.org/10.1016/j.jclinepi.2010.04.011
7. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007;45(Suppl. 1):S3-11. http://dx.doi.org/10.1097/01.mlr.0000258615.42478.55
8. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373-83.

9. Chen RE, Papuga MO, Voloshin I, Nicandri GT, Goldblatt JP, Bronstein RD, et al. Preoperative PROMIS scores predict postoperative outcomes after primary ACL reconstruction. Orthop J Sports Med 2018;6:2325967118771286. http://dx.doi.org/10.1177/2325967118771286
10. Cho CH, Jung SW, Park JY, Song KS, Yu KI. Is shoulder pain for three months or longer correlated with depression, anxiety, and sleep disturbance? J Shoulder Elbow Surg 2013;22:222-8. http://dx.doi.org/10.1016/j.jse.2012.04.001
11. Cho CH, Seo HJ, Bae KC, Lee KJ, Hwang I, Warner JJ. The impact of depression and anxiety on self-assessed pain, disability, and quality of life in patients scheduled for rotator cuff repair. J Shoulder Elbow Surg 2013;22:1160-6. http://dx.doi.org/10.1016/j.jse.2013.02.006
12. Choi SW, Podrabsky T, McKinney N, Schalet BD, Cook KF, Cella D. PROsetta Stone. PROMIS anxiety and generalized anxiety disorder scale, <http://www.prosettastone.org/LinkingTables/Documents/PROMIS%20Anxiety%20and%20GAD-7%20Full%20Report.pdf>; 2015, accessed March 15, 2018.
13. Choi SW, Podrabsky T, McKinney N, Schalet BD, Cook KF, Cella D. PROsetta Stone. PROMIS depression and PHQ-9, <http://www.prosettastone.org/LinkingTables/Documents/PROMIS%20Depression%20and%20PHQ-9%20Full%20Report.pdf>; <http://www.prosettastone.org/LinkingTables/Documents/PROMIS%20Anxiety%20and%20GAD-7%20Full%20Report.pdf>; 2015, accessed March 15, 2018.
14. 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(Pt. 2):1694-711. http://dx.doi.org/10.1111/j.1475-6773.2005.00446.x
15. Döring AC, Nota SP, Hageman MG, Ring DC. Measurement of upper extremity disability using the Patient-Reported Outcomes Measurement Information System. J Hand Surg Am 2014;39:1160-5. http://dx.doi.org/10.1016/j.jhsa.2014.03.013
16. Dowdle SB, Glass N, Anthony CA, Hettrich CM. Use of PROMIS for patients undergoing primary total shoulder arthroplasty. Orthop J Sports Med 2017;5:2325967117726044. http://dx.doi.org/10.1177/2325967117726044
17. Gausden EB, Levack A, Nwachukwu BU, Sin D, Wellman DS, Lorich DG. Computerized adaptive testing for patient reported outcomes in ankle fracture surgery. Foot Ankle Int 2018:1071100718782487. http://dx.doi.org/10.1177/1071100718782487
18. Guattery JM, Dardas AZ, Kelly M, Chamberlain A, McAndrew C, Calfee RP. Floor effect of PROMIS depression CAT associated with hasty completion in orthopaedic surgery patients. Clin Orthop Relat Res 2018;476:696-703. http://dx.doi.org/10.1007/s11999.0000000000000076
19. Ho B, Houck JR, Flemister AS, Ketz J, Oh I, DiGiovanni BF, et al. Preoperative PROMIS scores predict postoperative success in foot and ankle patients. Foot Ankle Int 2016;37:911-8. http://dx.doi.org/10.1177/1071100716665113
20. Jepson C, Asch DA, Hershey JC, Ubel PA. In a mailed physician survey, questionnaire length had a threshold effect on response rate. J Clin Epidemiol 2005;58:103-5. http://dx.doi.org/10.1016/j.jclinepi.2004.06.004
21. Kocher MS, Horan MP, Briggs KK, Richardson TR, O'Holleran J, Hawkins RJ. Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. J Bone Joint Surg Am 2005;87:2006-11. http://dx.doi.org/10.2106/JBJS.C.01624
22. Menendez ME, Baker DK, Oladeji LO, Fryberger CT, McGwin G, Ponce BA. Psychological distress is associated with greater perceived disability and pain in patients presenting to a shoulder clinic. J Bone Joint Surg Am 2015;97:1999-2003. http://dx.doi.org/10.2106/JBJS.O.00387
23. Minoughan CE, Schumaier AP, Fritch JL, Grawe BM. Correlation of PROMIS Physical function upper extremity computer adaptive test with American Shoulder and Elbow Surgeons shoulder assessment form and simple shoulder test in patients with shoulder arthritis. J Shoulder Elbow Surg 2018;27:585-91. http://dx.doi.org/10.1016/j.jse.2017.10.036
24. Overbeek CL, Nota SP, Jayakumar P, Hageman MG, Ring D. The PROMIS physical function correlates with the QuickDASH in patients with upper extremity illness. Clin Orthop Relat Res 2015;473:311-7. http://dx.doi.org/10.1007/s11999-014-3840-2
25. Ring D, Kadzielski J, Fabian L, Zurakowski D, Malhotra LR, Jupiter JB. Self-reported upper extremity health status correlates with depression. J Bone Joint Surg Am 2006;88:1983-8. http://dx.doi.org/10.2106/JBJS.E.00932
26. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE Jr. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. J Clin Epidemiol 2014;67:516-26. http://dx.doi.org/10.1016/j.jclinepi.2013.10.024
27. Simovitch R, Flurin PH, Wright T, Zuckerman JD, Roche CP. Quantifying success after total shoulder arthroplasty: the minimal clinically important difference. J Shoulder Elbow Surg 2018;27:298-305. http://dx.doi.org/10.1016/j.jse.2017.09.013
28. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999;282:1737-44.
29. Tokish JM, Kissenberth MJ, Tolan SJ, Salim TI, Tadlock J, Kellam T, et al. Resilience correlates with outcomes after total shoulder arthroplasty. J Shoulder Elbow Surg 2017;26:752-6. http://dx.doi.org/10.1016/j.jse.2016.12.070
30. Tyser AR, Beckmann J, Franklin JD, Cheng C, Hon SD, Wang A, et al. Evaluation of the PROMIS physical function computer adaptive test in the upper extremity. J Hand Surg Am 2014;39:2047-51, e2044. http://dx.doi.org/10.1016/j.jhsa.2014.06.130
31. 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;25:913-9. http://dx.doi.org/10.1016/j.jse.2016.01.029
32. Wright RW, Baumgarten KM. Shoulder outcomes measures. J Am Acad Orthop Surg 2010;18:436-44.
33. Wylie JD, Beckmann JT, Granger E, Tashjian RZ. Functional outcomes assessment in shoulder surgery. World J Orthop 2014;5:623-33. http://dx.doi.org/10.5312/wjo.v5.i5.623
34. Wylie JD, Suter T, Potter MQ, Granger EK, Tashjian RZ. Mental health has a stronger association with patient-reported shoulder pain and function than tear size in patients with full-thickness rotator cuff tears. J Bone Joint Surg Am 2016;98:251-6. http://dx.doi.org/10.2106/JBJS.O.00444
35. PROMIS. Depression. A brief guide to the PROMIS© Depression instruments: <http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Depression_Scoring_Manual_02222017.pdf>; 2017, accessed March 1, 2018.