Quality of Life Research (2019) 28:2839–2850
https://doi.org/10.1007/s11136-019-02223-8

Prediction of Oswestry Disability Index (ODI) using PROMIS-29 in a national sample of lumbar spine surgery patients

Jacquelyn S. Pennings[1,2] · Clinton J. Devin[1,3] · Inamullah Khan[1] · Mohamad Bydon[4] · Anthony L. Asher[5] · Kristin R. Archer[1,2,6]

Accepted: 29 May 2019 / Published online: 6 June 2019

Abstract

Purpose The primary purpose was to examine the measurement properties of the PROMIS-29 to better understand its use in patients undergoing spine surgery. A secondary objective was to calculate a predictive equation between PROMIS-29 and ODI, to allow clinicians and researchers to determine a predicted ODI score based on PROMIS short form scores.

Methods 719 patients with PROMIS v2.0 and ODI responses were queried from the quality outcomes database. Validity was assessed using coefficient omega, ceiling/floor effects, and confirmatory factor analysis. Multivariable regression predicting ODI scores from PROMIS-29 domains was used to create a predictive equation. Predicted ODI scores were plotted against ODI scores to determine how well PROMIS-29 domains predicted ODI.

Results Results showed good reliability and validity of PROMIS-29 in patients undergoing lumbar spine surgery: convergent and discriminant validity, low floor/ceiling effects, and unidimensional domains. The conversion equation used 6 PROMIS-29 domains ($ODI_{\%} = 37.847 - 1.475*[PF_{raw}] + 1.842*[PAIN_{raw}] + 0.557*[SD_{raw}] - 0.642*[SR_{raw}] + 0.478*[PI_{raw}] + 0.295*[DEP_{raw}]$). Correlation between the predicted and actual ODI scores was $R = 0.88$, $R^2 = 0.78$, suggesting that the equation predicted ODI scores that are strongly correlated with actual ODI scores.

Conclusions Good measurement properties support the use of PROMIS-29 in spine surgery patients. Findings suggest accurate ODI scores can be derived from PROMIS-29 domains. Clinicians who want to move from ODI to PROMIS-29 can use this equation to obtain estimated ODI scores when only collecting PROMIS-29. These results support the idea that PROMIS-29 domains have the potential to replace disease-specific traditional PROMs.

Keywords PROMIS-29 · Oswestry Disability Index · Spine surgery · Patient-reported outcomes

Introduction

Patient-reported outcome measures (PROMs) are important tools for evaluating the success of spine surgery [1]. Deciding on the optimal PROMs remains an ongoing debate at the international and national level and at the local level for individual practices and surgeons. Although there are numerous reviews published on the topic of which PROMs best assess spine surgery outcomes, consensus still does not exist [2–9]. Important characteristics of PROMS include the need to cover the full spectrum of patient symptoms, sensitivity to responder and administration burden, and ability to administer effectively through multiple mediums including phone, tablet, or email. Most reviews recommend using various outcome domains including physical function, pain, and quality of life [3, 9]. The Oswestry Disability Index (ODI) is one of the most commonly collected and reported PROMs

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11136-019-02223-8) contains supplementary material, which is available to authorized users.

✉ Kristin R. Archer
kristin.archer@vumc.org

1 Department of Orthopaedic Surgery, Vanderbilt University School of Medicine, Nashville, TN, USA

2 Vanderbilt Center for Musculoskeletal Research, Vanderbilt University Medical Center, Nashville, TN, USA

3 Steamboat Orthopaedic and Spine Institute, Steamboat Springs, CO, USA

4 Department of Neurological Surgery, Mayo Clinic, Rochester, MN, USA

5 Carolina Neurosurgery & Spine Associates, Carolinas Healthcare System, Charlotte, NC, USA

6 Department of Physical Medicine and Rehabilitation, Vanderbilt University School of Medicine, Nashville, TN, USA

for assessing functional status in patients undergoing spine surgery. The ODI is one of the core measures used by various spine societies, such as Eurospine COMI [10], AANS [11], NASS [12], and ICHOM [13, 14] and also used in randomized clinical trials [4, 15, 16], everyday physician practice [17, 18], and in spine registries [19–22]. In a 2015 review of 22 spine registries representing 14 countries, the ODI was identified as the primary functional status measure in 18 registries [13].

The ODI has demonstrated responsiveness in spine surgery populations and has an established minimally clinically important difference (MCID) [18]. Responder burden is relatively low and it is simple and easy to score [9]. However, the ODI has been shown to have a large floor effect resulting in better distinction between degrees of functioning at above-average disability levels compared to patients with lower disability [23, 24]. There have been inconsistent findings regarding the unidimensionality of the ODI [25]. In a study comparing the responsiveness of ODI to a general health measure (SF-36), results demonstrated that a general health measure (SF-36) was just as responsive to change as the ODI, which is disease specific.

The Patient-Reported Outcomes Measurement Information System (PROMIS) tools offer an alternative to the ODI and other legacy measures utilized by clinicians and researchers. The National Institutes of Health (NIH) Task Force on Research Standards for Chronic Low Back Pain has recommended using several of the PROMIS domains as part of the standards for research on chronic low back pain [26]. PROMIS offers many benefits: it is easily accessible, free, includes multiple individual domains, and two modes of delivery including computer adaptive tests (CAT) and short forms which both provide t-scores that allow for easy comparisons to population averages [26, 27]. While CATs have been shown to increase precision of responses, the PROMIS-29 Health Profile (PROMIS-29) is a multidimensional scale comprised of several PROMIS short form domains. The PROMIS-29 is useful in situations where completing the measure on an internet enabled device is not feasible. PROMIS measures also allow for patients to be compared across pathologies when the same domain is collected.

Since PROMIS measures are not condition-specific, researchers and clinicians have been hesitant to incorporate them in place of common legacy measures such as the ODI. One barrier to using PROMIS is the lack of established MCIDs for spine populations, including spine surgery [28, 29]. A second barrier is that switching to PROMIS does not allow outcomes to be directly compared to previously collected ODI scores. Recent studies have focused on comparing PROMIS to ODI and other legacy measures to better understand its utility for patients undergoing spine surgery. Several studies have found strong correlations between CAT physical function and CAT pain interference and the ODI [30–33] and one proposed a conversion equation between the CAT physical function T score and the ODI [33]. However, studies have not evaluated the psychometric properties of the PROMIS short forms or correlated them with the ODI. Thus, the two aims of the current study are (1) evaluate the internal consistency, interpretability, construct validity, and structural validity according to the COSMIN guidelines [34] of the PROMIS-29 in patients undergoing spine surgery to further justify the use of the PROMIS-29 in this population, and (2) compare the short forms of the eight domains included in the PROMIS-29 to the ODI and compute a conversation equation. This equation will allow clinicians and researchers to determine a patient's likely ODI score based on their PROMIS short form score.

Methods

Data source: quality outcomes database (QOD)

The Quality Outcomes Database (QOD) is a prospective registry into which patients have been enrolled since 2012 [11, 35]. It is designed to evaluate risk-adjusted expected morbidity and 12-month outcomes with the aim of improving efficiency and quality of care for the most commonly performed spinal surgical procedures [36]. The QOD project is designated as a non-research, clinical quality improvement effort and based on existing federal guidelines exempted from IRB review. Patient demographics, history, and clinical characteristics were collected at baseline. Surgery information, postoperative complications, and readmissions were collected 3 months after surgery, while PROMs were longitudinally collected at baseline, 3-month, and 12-months via online survey or phone interview. All data are collected by an independent data coordinator not involved with clinical care [11]. Within sites, patients are enrolled either continuously or on a prespecified rotating cycle ensuring unbiased enrollment into the registry [11]. For the purposes of this analysis, we queried the QOD registry for patients who had both PROMIS and ODI responses. ODI has been collected since the registry's inception in 2012. PROMIS-29 has been recently added to QOD as an optional PROM as of 2018. Since August 2018, 14 of 75 sites added the optional PROMIS-29 survey to their battery of PROMs.

Inclusion and exclusion criteria

Patients undergoing lumbar spine surgery performed for a diagnosis of primary stenosis, spondylolisthesis, disc herniation, symptomatic mechanical disc collapse, and pseudarthrosis including recurrent same-level disc herniation and adjacent-segment disease were eligible for

inclusion. Exclusions included spinal infection, tumor, fracture, traumatic dislocation, neurological paralysis due to preexisting spinal disease or injury, less than 18 years of age, and incarceration. Patients who didn't have both ODI and PROMIS-29 at one time point were also excluded.

Measures

The PROMIS-29 (v2.0) is a collection of short forms from seven PROMIS domains including physical function, depression, anxiety, fatigue, sleep disturbance, participation in social roles and activities, and pain interference [37, 38]. Each domain in the PROMIS-29 has 4 items rated on a 5 point scale. There is also one pain intensity item rated from 0 "no pain" to 10 "worst pain imaginable." PROMIS-29 domains have demonstrated strong correlations (0.83 to 0.90) with the full PROMIS domain item banks [38–45].

The Oswestry Disability Index (ODI) assesses change in functional status of adults with low back pain [46]. The ODI contains ten pain-related questions scored from zero (no pain) to five (most severe pain). Scores are expressed as a percentage of total points, with ≤ 20% indicating minimal disability, 21–40% moderate disability, 41–60% severe disability, 61–80% crippled, and 81–100% completely bedbound [46].

Statistical analysis

Patient demographics and baseline characteristics were reported. Descriptive statistics of the ODI and PROMIS-29 domains were computed. Per PROMIS-29 scoring instructions, patients were required to have all items answered to obtain a score on that domain, resulting in 3% of PROMIS scores being excluded [38]. Patients also had to have at least 6 ODI items completed to be included in the analysis. After meeting this requirement, missing responses for included patients were treated with person-mean imputation [47]. PROMIS-29 *T* scores, derived from the raw scores and population norms, provide a reference for the current sample in comparison to the general population (more information on the T score metrics and normative populations can be found on the website (http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis). Following the standard of the PROMIS validation studies, PROMIS raw scores were used instead of T score for all correlations and regression [48, 49]. Data were evaluated for outliers and missing data. SPSS v25 was used for the descriptive statistics and prediction models. The lavaan package (v0.6-3) for R was used for the confirmatory factor analysis [50].

Internal consistency, interpretability, and construct validity

To assess the reliability of the domains, coefficient omega, a composite reliability, was used because it does not assume tau-equivalence or uncorrelated error variances [51]. Interpretability was assessed via evaluation of the ceiling and floor effects via reporting the proportion of patients who obtained the minimum and maximum scores for each PROMIS-29 domain and the ODI. Convergent and discriminant validity were used to assess construct validity. Specifically, we tested whether the ODI was more strongly correlated with domains conceptually related to it (convergent validity) and less strongly correlated with domains from which the ODI was conceptually different (discriminant validity) [52]. Using Bivariate Spearman's correlations, the ODI was expected to correlate more strongly with physical function (strong negative correlation), pain interference (strong positive correlation), and pain intensity (strong positive correlation) demonstrating convergent validity than with anxiety, depression, and fatigue (all moderate positive relationships), which would demonstrate discriminant validity [53]. The effect sizes for correlation coefficients were small = 0.1 to < 0.3, moderate = 0.3 to <0.50, and large ≥ 0.5 [54].

Structural validity

Confirmatory factor analysis (CFA) was used to confirm structural validity. DWLS estimation was used to test the goodness of fit between the hypothesized model and the data set. The hypothesized model included each of the four items making up each domain loading only on their respective domain. Goodness of fit was assessed with Standardized Root Mean Square Residual (SRMR), Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and adjusted chi square (χ2/df). The maximum cutoff values for the SRMR and RMSEA are 0.08 and 0.06, respectively, and the minimum cutoff values for the NNFI and the CFI are 0.95 to conclude a good fit between the model and the data [55]. The critical value for the adjusted chi square is < 3.00. Average variance explained (AVE), which should be above 0.50, and maximum shared variance (MSV), which should be smaller than the composite reliability, were used to assess reliability and discriminant validity of the CFA [56]. Standardized path coefficients (λ) < 0.50 were considered for removal from the final model. Modification indices were checked to be sure no item loaded strongly on another domain or that applying modifications would not improve overall model fit.

Prediction models

Relationships between the ODI and each PROMIS-29 domain were linear, thus linear regression models were used to create predictive equations from the PROMIS regression coefficients and constants. Raw scores were used for the primary results but equations using T scores are presented in the Supplementary Material Table 1.

The first equation was derived from the prediction of ODI from only the Physical Performance PROMIS-29 score since previous studies have suggested the Physical Performance CAT can be used as an ODI proxy [33]. The second equation was derived using all the PROMIS-29 scores that significantly predicted ODI in a multivariable regression. To confirm that non-significant factors excluded from the second equation did not meaningfully contribute to the prediction of ODI score, a third multivariable regression included using all eight PROMIS-29 domain scores and the change in R^2 from the previous model was checked for statistical significance.

Residual distributions and normal probability plots were evaluated to ensure assumptions of the regressions were met. The unstandardized predicted ODI scores, derived from the regression equations, were plotted against the actual ODI scores to determine how well the PROMIS-29 domains predicted ODI scores. Scatterplots were color coded by time point and separate best fit lines added to show the predictive ability of the regression equation at each timepoint. The final relationships, produced by the curve estimations, were also checked for linearity using curve estimation. The absolute difference between the predicted and actual ODI score was computed and the percent of people within each 5 point increment of absolute difference reported.

Results

About half of the 719 patients were female (49.5%) and most were Caucasian (91%). The mean age was 61 years (SD = 14). 137 (19%) were having a revision surgery (19%) and baseline back and leg pain were 6.4 (SD = 2.5) and 6.3 (SD = 2.6), respectively. Other sample demographics are shown in Table 1.

The means and standard deviations of the ODI and PROMIS domain scores are shown for each timepoint in Table 2. The mean ODI score was significantly lower at 3 months compared to baseline ($p < 0.001$) and at 12 months compared to baseline ($p < 0.001$). There was no difference in ODI scores between 3 and 12 months ($p = 0.110$). For PROMIS domain scores, positively valanced (i.e., higher scores considered better) mean domain scores were significantly higher at 3 months compared to baseline and negatively valanced (i.e., higher score is considered worse) mean domain scores were significantly lower at 3 months compared to baseline ($p < 0.001$). Mean differences between baseline and 12 months were also statistically significant ($p < 0.001$). No PROMIS-29 domain scores were significantly different at 12 months compared to 3 months ($p > 0.05$). Based on the PROMIS-29 T scores at baseline, the current population shows moderate impairment for physical functioning and pain interference, and mild impairment in fatigue, sleep disturbance, and participation in social roles. At follow-up time points, all PROMIS domains are in or very near normal limits.

Table 1 Respondent demographic and clinical characteristics ($N = 719$)

	n (%)[a]
Female	356 (49.5%)
Age, mean (SD)	61.0 ± 14.4
Caucasian	656 (91.2%)
Insurance	
Private insurance	325 (45.2%)
Medicare/medicaid	342 (47.6%)
Other	52 (7.2%)
Education: high school or less	294 (41.1%)
Employed	296 (41.3%)
Smoker	91 (12.7%)
Diabetes	141 (19.6%)
Depression	146 (20.3%)
Revision surgery	137 (19.1%)
Baseline back pain (NRS), mean (SD)[b]	6.4 ± 2.5
Baseline leg pain (NRS), mean (SD)[b]	6.3 ± 2.6

[a]Except where noted

[b]NRS scales measured on a 0 to 10 rating scale

Internal consistency, interpretability, and construct validity

As shown in Table 3, all ODI and PROMIS-29 domains demonstrated good internal consistency (coefficient omega > 0.84). Interpretability was assessed via evaluation of the floor and ceiling effects. For the ODI, no patient had the highest possible score (ceiling effect) at any of the timepoints. Less than 1% of patients had the lowest possible score (floor effect) at baseline but 9–11% of patients had a floor effect at follow-up.

For the positively valanced PROMIS-29 domains, floor effects decreased and ceiling effects increased from baseline to 3 and 12 months, indicating that fewer patients had the lowest (worst) possible score and more patients had the highest (best) possible score over time. In the negatively valanced domains, the opposite occurred. When the domains of

Table 2 Descriptive statistics of ODI and PROMIS-29 raw and *T* scores at each time point

	Baseline ($N=320$)				3 Months ($N=151$)				12 Months ($N=89$)			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
ODI score	45.4	16.2	0	87.5	25.6	18.6	0	76	21.6	17.7	0	65
PROMIS raw scores												
Physical function	9.9	3.7	4	20	15.0	4.3	4	20	15.3	4.1	5	20
Anxiety	8.0	3.6	4	19	6.3	3.2	4	19	6.0	3.2	4	17
Depression	7.1	3.7	4	20	6.2	3.3	4	20	6.0	3.3	4	19
Fatigue	11.4	4.2	4	20	9.2	4.0	4	20	9.2	4.2	4	20
Sleep disturbance	12.5	3.7	4	20	10.2	3.9	4	20	10.1	3.9	4	19
Participation in social roles	10.3	4.0	4	20	14.3	4.5	4	20	15.2	4.1	4	20
Pain interference	15.3	3.8	4	20	9.4	4.7	4	20	9.0	4.2	4	20
Pain intensity	6.8	1.9	0	10	3.2	2.5	0	9	3.4	2.6	0	9
PROMIS T Scores												
Physical function	34.0	5.7	22.9	56.9	42.7	8.4	22.9	56.9	43.5	8.6	26.9	56.9
Anxiety	53.8	9.4	40.3	77.9	48.4	9.4	40.3	77.9	47.4	9.2	40.3	73.3
Depression	50.9	9.4	41.0	79.4	48.2	8.9	41.0	79.4	47.2	8.9	41.0	75.7
Fatigue	55.0	9.7	33.7	75.8	49.6	10.3	33.7	75.8	49.7	10.3	33.7	75.8
Sleep disturbance	55.1	7.9	32.0	73.3	49.9	9.3	32.0	73.3	49.7	8.9	32.0	68.8
Participation in social roles	42.2	7.5	29.0	64.1	50.0	9.4	29.0	64.1	51.7	8.8	29.0	64.1
Pain interference	66.3	6.6	41.6	75.6	55.8	9.6	41.6	75.6	54.9	8.8	41.6	75.6

Medians with IQR were also run for all scores due to non-normal distributions of some PROMIS domains (particularly anxiety and depression). Since median values were very similar to the means, the mean values are reported

All PROMIS domain are valanced according to the domain name such that a higher score is associated with more of what is being measured. As such, physical function and participation in social roles are positively valanced (higher scores considered better), and the rest of the scores are negatively valanced such that a higher score is considered worse

PROMIS T scores have a mean of 50 and a standard deviation of 10 in a referent population [38]. Scores 0.5–1.0 SD worse than the mean indicate mild symptoms/impairment, while scores 1 to 2 SD worse than the mean indicate moderate symptoms/impairment, and scores 2 + SD worse than the mean indicate severe symptoms/impairment (http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis)

the PROMIS-29 were considered together, no patient had a ceiling or floor effect at any timepoint.

Bivariate correlations between the ODI and the PROMIS domain scores are provided in Table 4 to assess the construct validity. Convergent and discriminant validity were supported with correlations between the ODI and physical function ($\rho = -0.82$), pain interference ($\rho = 0.83$), and pain intensity ($\rho = 0.77$) that were stronger in magnitude than the correlations between ODI and anxiety ($\rho = 0.49$), depression ($\rho = 0.51$), and fatigue ($\rho = 0.55$). The magnitude of the correlations of ODI with depression and fatigue were slightly stronger than expected (> 0.50). However, they were not as strong as the correlations between ODI and physical function and pain (≥ 0.77), which where hypothesized to be the strongest correlations. The pattern of correlations was consistent across baseline, 3 months, and 12 months with correlations at baseline generally smaller in magnitude than those at 3 and 12 months. The remaining two PROMIS-29 domains of sleep disturbance and participation in social roles were also correlated with ODI although no specific expectations were set with regards to validity. All correlation coefficients were statistically significant ($p < 0.001$).

Structural validity

Fit statistics for the CFA revealed adequate fit with the data, χ^2 (329) = 550.54, $p < 0.001$, adjusted χ^2 (χ^2/df) = 1.67, RMSEA = 0.029 (95% CI 0.025 to 0.034), SRMR = 0.029, CFI > 0.999. In addition, the standardized factor loadings (λ) were all greater than 0.76, $p < 0.001$, which is larger than our minimum cutoff of 0.50 (see Supplementary Material Table 2). Modification indices were not large (< 30) and applying them did not improve model fit. The AVE for each domain was greater than 0.62, which is larger than the minimum cutoff of 0.50 showing that an adequate amount of variation was explained by the items within each domain. Additionally, the CR was larger than the MSV for all domains demonstrating good model discrimination. Taken together, the results support excellent fit of the data to the hypothesized model.

Prediction models

To create the predicted scores, three sets of regressions were performed (Table 5). In Model 1, PROMIS-29

Table 3 Reliability and floor/ceiling ($N = 719$)

	Construct reliability (omega)				Baseline		3 months		12 months	
	Full Sample[a]	Base-line	3 months	12 months	Floor (%)	Ceiling (%)	Floor (%)	Ceiling (%)	Floor (%)	Ceiling (%)
ODI score	0.92	0.85	0.92	0.92	< 1	0	9	0	11	0
PROMIS raw domain scores										
Physical function[b]	0.90	0.85	0.89	0.90	5	1	1	17	0	22
Anxiety[c]	0.89	0.87	0.91	0.94	24	< 1	52	< 1	52	0
Depression[c]	0.84	0.91	0.89	0.94	39	< 1	54	< 1	57	< 1
Fatigue[c]	0.94	0.92	0.94	0.95	7	4	17	1	15	2
Sleep disturbance[c]	0.88	0.85	0.88	0.87	1	3	7	3	9	1
Participation in social roles[b]	0.94	0.91	0.95	0.96	10	2	3	22	2	22
Pain interference[c]	0.97	0.94	0.97	0.97	1	18	22	7	24	2
Pain intensity (1 item)[c]	–	–	–	–	< 1	6	17	0	15	< 1
Overall					< 1	0	3	0	3	0

[a]56 patients had responses at baseline and 3 months, resulting in a sample size for the full sample ($N = 775$) being higher than the number of patients ($N = 719$)

[b]Positively valanced domains

[c]Negatively valanced domains

physical function was a strong predictor of ODI (Beta = − 3.4, SE = 0.90, $p < .001$). The R value was 0.81 with an adjusted $R^2 = 0.65$. In Model 2, six of the PROMIS-29 domains significantly predicted ODI with physical function and pain intensity being the strongest predictors. Model 2 (adjusted $R^2 = 0.78$, $R = 0.88$) had a significantly higher R^2 ($\Delta R^2 = 0.13$) than Model 1, $p < 0.001$.

Results show no discernable contributions of anxiety and fatigue to the overall model ($\Delta R^2 < 0.001$). Therefore, two predictive equations were created for ODI using (1) only physical function (PF) ($ODI_{\%} = 77.931 - 3.423*[PF_{raw}]$) and (2) the six PROMIS-29 domains of physical function, pain intensity (PAIN), sleep disturbance (SD), participation in social roles (SR), pain interference (PI), and depression (DEP) ($ODI_{\%} = 37.847 - 1.475*[PF_{raw}] + 1.842*[PAIN_{raw}] + 0.557*[SD_{raw}] - 0.642*[SR_{raw}] + 0.478*[PI_{raw}] + 0.295*[DEP_{raw}]$). Predictive equations were also computed using the T scores and are shown in Supplementary Material Table 2.

The predicted ODI scores were generated and plotted against the actual ODI scores for Model 1 and Model 2 (Fig. 1). To evaluate the predictive equation at each time point, the scatterplots have a separate symbol for each timepoint and include best fit lines for baseline, 3-months, and 12-months. The linear R and R^2 values are presented in Fig. 1.

The absolute difference between the ODI score and the predicted ODI scores for Model 1 (Median = 7.8, IQR 3.9–12.8) and Model 2 (Median = 5.8, IQR 2.6–10.3) were calculated. Table 6 shows the cumulative frequency and percent of the absolute differences between predicted and actual ODI scores for Model 1 and Model 2 in five-point increments. In Model 2, 43% of patients had predicted ODI scores that were within 5 points of the actual ODI scores and 74% of patients were within 10 points on the ODI. Model 1 was somewhat worse, with 33% of patients having predicted ODI scores that were within 5 points of the actual ODI scores and 64% of patients were within 10 points on the ODI.

Discussion

The first aim was to evaluate the internal consistency, interpretability, construct validity, and structural validity of the PROMIS-29 in patients undergoing spine surgery to further justify the use of the PROMIS-29 in this population. The second aim was to compare the short forms of the 8 domains included in the PROMIS-29 to the ODI and compute a conversation equation to allow clinicians and researchers to determine a patient's likely ODI score based on their PROMIS short form score.

Table 4 Spearman's correlations between ODI and PROMIS-29 domain raw scores ($N=719$)

	ODI score			
	Full sample[a]	Baseline	3 months	12 months
Physical function	−0.82	−0.65	−0.83	−0.83
Anxiety	0.49	0.42	0.40	0.50
Depression	0.51	0.53	0.45	0.49
Fatigue	0.55	0.46	0.52	0.56
Sleep disturbance	0.59	0.48	0.59	0.62
Participation in social roles	−0.76	−0.57	−0.75	−0.77
Pain interference	0.83	0.68	0.83	0.82
Pain intensity	0.77	0.51	0.77	0.78

[a]56 patients had responses at baseline and 3 months, resulting in a sample size for the full sample ($N=775$) being higher than the number of patients ($N=719$)

All p values $<.001$

Table 5 Multivariable regression models predicting ODI scores from PROMIS-29 domains ($N=775$)

Model	Predictor	R	Adj R^2	Beta	SE	p	Partial correlation
1	Individual domain	0.806	0.649	77.931[a]			
	Physical function			−3.423	0.090	< 0.001	−0.806
2	Significant Domains	0.884	0.779	37.847[a]			
	Physical function			−1.475	0.124	< 0.001	−0.396
	Pain intensity			1.842	0.209	< 0.001	0.303
	Sleep disturbance			0.557	0.104	< 0.001	0.190
	Participation in social roles			−0.642	0.134	< 0.001	−0.171
	Pain interference			0.478	0.156	0.002	0.110
	Depression			0.295	0.109	0.007	0.097
3	All domains	0.884	0.779	38.383[a]			
	Physical function			−1.477	0.124	0.000	−0.396
	Pain intensity			1.831	0.210	0.000	0.301
	Sleep disturbance			0.583	0.109	0.000	0.190
	Participation in social roles			−0.658	0.135	0.000	−0.174
	Pain interference			0.491	0.157	0.003	0.112
	Depression			0.299	0.148	0.043	0.073
	Fatigue			−0.112	0.109	0.307	−0.037
	Anxiety			0.056	0.144	0.699	0.014

56 patients had responses at baseline and 3 months, resulting in a sample size for the full sample ($N=775$) being higher than the number of patients ($N=719$). The combined data (over all three timepoints) were used because the predictive equations need to be usable regardless of when data are collected

[a]Constant coefficient. Model 1 root mean square error (RMSE) = 11.66. Model 2 RMSE = 9.25

Aim 1

The PROMIS-29 displays good internal consistency, interpretability, construct validity, and structural validity in the current sample. Previous reviews of PROMs for lower back pain have suggested that PROMIS measures be further investigated in samples of patients with lower back pain including head-to-head comparisons of measurement properties of PROMIS and legacy measures such as the ODI [4]. Additionally, a recent review highlights the need to evaluate the validity of the PROMIS short forms, specifically, the physical function domain in patients with lower back pain (LBP) [25]. In the current study, coefficient omega values ranged from ω = 0.84 to ω = 0.97, which are considerably above the generally accepted lower reliability limit of ω = 0.70 and similar to reliability reported for the PROMIS-29 in other samples [57]. The ceiling and floor effects within domains are reasonable when accounting for the fact that each dimension has only 4 items and similar to another study reporting floor and ceiling effects for the PROMIS-29 in a kidney transplant sample [57]. There were no ceiling or floor effects when all PROMIS-29 domains were considered

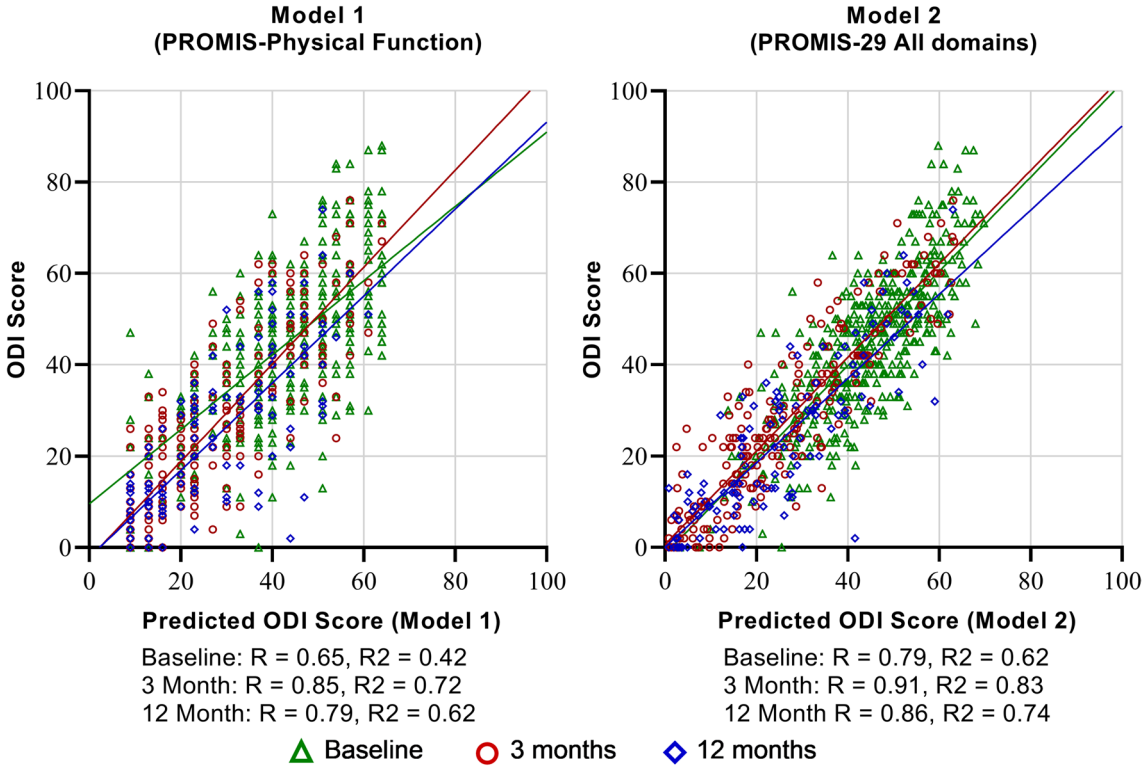

Fig. 1 Scatterplot showing relationship between the actual ODI score and the predicted ODI scores based on PROMIS-29 domains. Each marker represents the point where the predicted (*x*-axis) and actual (*y*-axis) ODI scores intersect. Points closer to a 45° angle represent more accurately predicted ODI scores with less variability. A best-fit line has been shown for the dots at each time-period individually. Best-fit lines closer to the 45° angle and points closer to the best fit line represent optimal prediction

Table 6 Cumulative frequency and percent of absolute differences between predicted and actual ODI scores for Model 1 and Model 2

	Model 1		Model 2	
	N	%	*N*	%
Within 5 pts	253	32.6	336	43.4
Within 10 pts	497	64.1	572	73.8
Within 15 pts	628	81.0	692	89.3
Within 20 pts	709	91.5	754	97.3
Within 25 pts	749	96.6	767	99.0
Within 40 pts	775	100.0	775	100.0

together meaning that no patient responded to every item on the PROMIS-29 with the worst or best possible answers at any timepoint. No ceiling or floor effects allow even patients who are less symptomatic to be distinguished over time. The floor and ceiling effects for the ODI were similar except that the ODI had about 10% of patients with floor effects at follow-up, meaning that less symptomatic patients cannot be distinguished from each other after lumbar spine surgery when using only the ODI. Similar floor effects for the ODI have been reported [58]. The PROMIS-29 domains showed good convergent ($\rho \leq 0.77$) and discriminant validity ($\rho \geq 0.55$) when compared with the ODI. Additionally, CFA showed that the data fit well to the hypothesized model and that each PROMIS-29 domain was unidimensional with factor loadings ($\lambda \geq 0.76$) well above the minimally acceptable criteria of $\lambda = 0.50$. These results are also similar to the few other PROMIS-29 studies reporting these statistics in other populations although no other studies were found that use the ODI for PROMIS-29 validity [57, 59, 60].

These internal consistency and validity results are also in line with other studies that have evaluated the PROMIS physical function CAT in orthopaedic [31] and other diverse clinical samples [61]. Additional studies have reported successful implementation of PROMIS measures in those with diabetes [62], and have reported good reliability and validity in older adults with chronic conditions [59], hematopoietic cell transplant survivors [27], and patients with arthritis [63]. One study on the ecological validity of the PROMIS in both the CAT and short form version found that that PROMIS scales reflect the real-life symptoms experienced by patients with a wide range of disease by comparing PROMIS responses to 7-day recall diaries [64]. Although further research is needed to confirm the validation of the PROMIS-29 in spine surgery patients, the current study provides evidence that using the PROMIS-29 is justified

in a spine surgery population and is a good candidate for measuring long-term PROs. These findings contribute both to the body of literature on selecting the best PROMs for lumbar spinal surgery and the literature on validation of the PROMIS, particularly the PROMIS-29.

Aim 2

The current findings suggest that an ODI score can be derived from the PROMIS-29 domain scores with a high degree of accuracy. It has been suggested that a model that is most useful for predictive purposes will have a $R^2 > .60$ and a RMSE < 10% of the possible scores, which is 10 for the ODI (100*0.10) [65]. Using only the PROMIS physical function domain score, which is made up of four items, gives a good estimate of the ODI score ($R = 0.81$, $R^2 = 0.65$; RMSE = 11.66). The RMSE of 11.66 is slightly higher than the recommended value of 10 or less, which suggests somewhat high variability in the predicted ODI scores. When using five of the domain scores plus the single pain intensity item from the PROMIS-29, the ODI scores can be estimated with significantly higher precision ($\Delta R^2 = 0.13$, $p < .001$, $R = 0.88$, $R^2 = 0.78$, RMSE = 9.25) which meets all criteria for predictive equation.

For clinicians who want to move from the ODI to the PROMIS physical function-4 (short form) or PROMIS-29, these prediction equations will provide estimated ODI scores without having to collect the ODI in addition to the PROMIS. The correlation found between the physical function-4 raw score and the ODI closely mirrors other studies correlating the physical function-CAT and ODI, which found correlations ranging from −0.81 to −0.83 (current study $\rho = -0.82$) when timepoints were combined [30, 66] and correlations at separate time points ranging from −0.66 to −0.80 (current study $\rho = -0.65$ to −0.83) [67]. One previous study provided a formula for converting the physical function-CAT to ODI scores: ODI = 153.533 − 3.012*$PF_{CAT\text{-}Tscore}$ which is in contrast to our physical function *T*-scores: $ODI_{\%} = 109.129 - 1.920*[PF_{shortformTscore}]$ [33]. Applying this previously reported formula to our data resulted in an $R^2 = .63$, which is similar to the results using the physical function *T* score equation in the current study ($R^2 = .63$). However, using the previously reported equation also resulted in 10% of predicted ODI scores being invalid (ODI score of −18). Our formula for estimating ODI from the short form resulted in all predicted ODI scores being valid (between 0 and 100).

Using the PROMIS-29 provides several advantages over other common PROMs. Since PROMIS is not disease specific, scores can be compared across many populations and standardized scores allow for patients to be evaluated in relation to general population scores. The current study demonstrates that the PROMIS domain scores correlate strongly with one of the most common back pain specific measures (ODI), suggesting that a non-disease specific PROM can be collected as a primary outcome for patients undergoing lumbar spine surgery. As previously suggested, these results support the idea that PROMIS domains have the potential to replace disease-specific traditional PROMs, and not only using the PROMIS CATs but also the PROMIS short forms such as those in the PROMIS-29 [4].

Limitations and future research

The predictive equations show limited coverage for patients with ODI scores over 80. These patients are generally classified as being bed bound [46]. They represent both a very small portion of the population of those receiving lumbar spine surgery and correspondingly a low percentage of our registry sample. Further research on patients who have the most severe lower back pain issues is needed. Another limitation is that the PROMIS-29 is newly added to the registry providing a somewhat limited follow-up sample. Particularly, data are not yet available to evaluate the responsiveness of the PROMIS-29 domains over time after lumbar spine surgery but that is a next step in evaluating its predictive validity. Future research is needed to validate the current predictive equations in an independent sample of patients undergoing lumbar spine surgery.

Conclusions

The current study demonstrated good internal consistency, interpretability, construct validity, and structural validity of the PROMIS-29 in patients undergoing spine surgery. The PROMIS-29, when considering domains together, demonstrated no ceiling or floor effects. Each PROMIS-29 domain was found to be unidimensional and demonstrated ceiling and floor effects similar to findings in other populations. ODI scores can be derived from the PROMIS-29 with a high degree of accuracy. The highest precision was found with a predictive equation that utilized the PROMIS-29 domains of physical function, sleep disturbance, participation in social roles, pain interference, and depression, and the single pain intensity item. Together, these findings further justify the use of the PROMIS-29 in patients undergoing spine surgery.

Acknowledgements The study was funded by NeuroPoint Alliance, Inc.

References

1. DeVine, J., Norvell, D. C., Ecker, E., Fourney, D. R., Vaccaro, A., Wang, J., et al. (2011). Evaluating the correlation and responsiveness of patient-reported pain with function and quality-of-life outcomes after spine surgery. *Spine, 36*(21 Suppl), S69–S74. https://doi.org/10.1097/brs.0b013e31822ef6de.
2. Chiarotto, A., Ostelo, R. W., Turk, D. C., Buchbinder, R., & Boers, M. (2017). Core outcome sets for research and clinical practice. *Brazilian Journal of Physical Therapy, 21*(2), 77–84. https://doi.org/10.1016/j.bjpt.2017.03.001.
3. Chiarotto, A., Deyo, R. A., Terwee, C. B., Boers, M., Buchbinder, R., Corbin, T. P., et al. (2015). Core outcome domains for clinical trials in non-specific low back pain. *European Spine Journal, 24*(6), 1127–1142. https://doi.org/10.1007/s00586-015-3892-3.
4. Chiarotto, A., Terwee, C. B., & Ostelo, R. W. (2016). Choosing the right outcome measurement instruments for patients with low back pain. *Best Practice & Research Clinical Rheumatology, 30*(6), 1003–1020. https://doi.org/10.1016/j.berh.2017.07.001.
5. Davidson, M., & Keating, J. L. (2002). A comparison of five low back disability questionnaires: reliability and responsiveness. *Physical Therapy, 82*(1), 8–24.
6. Goldsmith, E. S., Taylor, B. C., Greer, N., Murdoch, M., MacDonald, R., McKenzie, L., et al. (2018). Focused evidence review: Psychometric properties of patient-reported outcome measures for chronic musculoskeletal Pain. *Journal of General Internal Medicine, 33*(Suppl 1), 61–70. https://doi.org/10.1007/s11606-018-4327-8.
7. Grotle, M., Brox, J. I., & Vollestad, N. K. (2005). Functional status and disability questionnaires: What do they assess? A systematic review of back-specific outcome questionnaires. *Spine, 30*(1), 130–140. https://doi.org/10.1097/01.brs.0000149184.16509.73.
8. Stokes, O. M., Cole, A. A., Breakwell, L. M., Lloyd, A. J., Leonard, C. M., & Grevitt, M. (2017). Do we have the right PROMs for measuring outcomes in lumbar spinal surgery? *European Spine Journal, 26*(3), 816–824. https://doi.org/10.1007/s00586-016-4938-x.
9. McCormick, J. D., Werner, B. C., & Shimer, A. L. (2013). Patient-reported outcome measures in spine surgery. *Journal of the American Academy of Orthopaedic Surgeons, 21*(2), 99–107. https://doi.org/10.5435/JAAOS-21-02-99.
10. Ferrer, M., Pellise, F., Escudero, O., Alvarez, L., Pont, A., Alonso, J., et al. (2006). Validation of a minimum outcome core set in the evaluation of patients with back pain. *Spine, 31*(12), 1372–1379. https://doi.org/10.1097/01.brs.0000218477.53318.bc. **discussion 1380**.
11. McGirt, M. J., Speroff, T., Dittus, R. S., Harrell, F. E., Jr., & Asher, A. L. (2013). The National Neurosurgery Quality and Outcomes Database (N2QOD): General overview and pilot-year project description. *Neurosurgical Focus, 34*(1), E6. https://doi.org/10.3171/2012.10.FOCUS12297.
12. Daltroy, L. H., Cats-Baril, W. L., Katz, J. N., Fossel, A. H., & Liang, M. H. (1996). The North American spine society lumbar spine outcome assessment instrument: Reliability and validity tests. *Spine, 21*(6), 741–749.
13. van Hooff, M. L., Jacobs, W. C., Willems, P. C., Wouters, M. W., de Kleuver, M., Peul, W. C., et al. (2015). Evidence and practice in spine registries. *Acta Orthopaedica, 86*(5), 534–544. https://doi.org/10.3109/17453674.2015.1043174.
14. Clement, R. C., Welander, A., Stowell, C., Cha, T. D., Chen, J. L., Davies, M., et al. (2015). A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta Orthopaedica, 86*(5), 523–533. https://doi.org/10.3109/17453674.2015.1036696.
15. Forsth, P., Michaelsson, K., & Sanden, B. (2013). Does fusion improve the outcome after decompressive surgery for lumbar spinal stenosis? A two-year follow-up study involving 5390 patients. *The Bone & Joint Journal, 95*(7), 960–965. https://doi.org/10.1302/0301-620x.95b7.30776.
16. Cleland, J. A., Whitman, J. M., Houser, J. L., Wainner, R. S., & Childs, J. D. (2012). Psychometric properties of selected tests in patients with lumbar spinal stenosis. *The Spine Journal, 12*(10), 921–931. https://doi.org/10.1016/j.spinee.2012.05.004.
17. Hill, J. C., Dunn, K. M., Lewis, M., Mullis, R., Main, C. J., Foster, N. E., et al. (2008). A primary care back pain screening tool: Identifying patient subgroups for initial treatment. *Arthritis Care & Research, 59*(5), 632–641. https://doi.org/10.1002/art.23563.
18. Copay, A. G., Glassman, S. D., Subach, B. R., Berven, S., Schuler, T. C., & Carreon, L. Y. (2008). Minimum clinically important difference in lumbar spine surgery patients: A choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. *The Spine Journal, 8*(6), 968–974. https://doi.org/10.1016/j.spinee.2007.11.006.
19. van Hooff, M. L., Mannion, A. F., Staub, L. P., Ostelo, R. W., & Fairbank, J. C. (2016). Determination of the Oswestry Disability Index score equivalent to a "satisfactory symptom state" in patients undergoing surgery for degenerative disorders of the lumbar spine-a Spine Tango registry-based study. *The Spine Journal, 16*(10), 1221–1230. https://doi.org/10.1016/j.spinee.2016.06.010.
20. Niskanen, R. O. (2002). The oswestry low back pain disability questionnaire. A two-year follow-up of spine surgery patients. *Scandinavian Journal of Surgery, 91*(2), 208–211. https://doi.org/10.1177/145749690209100214.
21. McGirt, M. J., Bydon, M., Archer, K. R., Devin, C. J., Chotai, S., Parker, S. L., et al. (2017). An analysis from the Quality Outcomes Database, Part 1. Disability, quality of life, and pain outcomes following lumbar spine surgery: Predicting likely individual patient outcomes for shared decision-making. *Journal of Neurosurgery, 27*(4), 357–369. https://doi.org/10.3171/2016.11.spine16526.
22. Gabel, C. P., Cuesta-Vargas, A., Qian, M., Vengust, R., Berlemann, U., Aghayev, E., et al. (2017). The Oswestry Disability Index, confirmatory factor analysis in a sample of 35,263 verifies a one-factor structure but practicality issues remain. *European Spine Journal, 26*(8), 2007–2013. https://doi.org/10.1007/s00586-017-5179-3.
23. Saltychev, M., Mattie, R., McCormick, Z., Barlund, E., & Laimi, K. (2017). Psychometric properties of the Oswestry Disability Index. *International Journal of Rehabilitation Research, 40*(3), 202–208. https://doi.org/10.1097/MRR.0000000000000226.
24. Walsh, T. L., Hanscom, B., Lurie, J. D., & Weinstein, J. N. (2003). Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. *Spine, 28*(6), 607–615. https://doi.org/10.1097/01.brs.0000050654.97387.df.
25. Chiarotto, A., Ostelo, R. W., Boers, M., & Terwee, C. B. (2018). A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. *Journal of Clinical Epidemiology, 95,* 73–93. https://doi.org/10.1016/j.jclinepi.2017.11.005.
26. Deyo, R. A., Dworkin, S. F., Amtmann, D., Andersson, G., Borenstein, D., Carragee, E., et al. (2014). Report of the NIH Task Force on research standards for chronic low back pain. *Physical Therapy, 15*(8), 1249–1267. https://doi.org/10.1111/pme.12538.
27. Shaw, B. E., Syrjala, K. L., Onstad, L. E., Chow, E. J., Flowers, M. E., Jim, H., et al. (2018). PROMIS measures can be used to assess symptoms and function in long-term hematopoietic cell

transplantation survivors. *Cancer, 124*(4), 841–849. https://doi.org/10.1002/cncr.31089.
28. Hung, M., Saltzman, C. L., Kendall, R., Bounsanga, J., Voss, M. W., Lawrence, B., et al. (2018). What are the MCIDs for PROMIS, NDI, and ODI instruments among patients with spinal conditions? *Clinical Orthopaedics and Related Research, 476*(10), 2027–2036. https://doi.org/10.1097/CORR.0000000000000419.
29. Amtmann, D., Kim, J., Chung, H., Askew, R. L., Park, R., & Cook, K. F. (2016). Minimally important differences for patient reported outcomes measurement information system pain interference for individuals with back pain. *Journal of Pain Research, 9,* 251–255. https://doi.org/10.2147/JPR.S93391.
30. Papuga, M. O., Mesfin, A., Molinari, R., & Rubery, P. T. (2016). Correlation of PROMIS physical function and pain CAT instruments with Oswestry Disability Index and neck disability index in spine patients. *Spine, 41*(14), 1153–1159. https://doi.org/10.1097/brs.0000000000001518.
31. Hung, M., Baumhauer, J. F., Latt, L. D., Saltzman, C. L., SooHoo, N. F., Hunt, K. J., et al. (2013). Validation of PROMIS (R) Physical Function computerized adaptive tests for orthopaedic foot and ankle outcome research. *Clinical Orthopaedics and Related Research, 471*(11), 3466–3474. https://doi.org/10.1007/s11999-013-3097-1.
32. Haws, B. E., Khechen, B., Guntin, J. A., Cardinal, K. L., Bohl, D. D., & Singh, K. (2018). Validity of PROMIS in minimally invasive transforaminal lumbar interbody fusion: a preliminary evaluation. *Journal of Neurosurgery, 29*(1), 28–33. https://doi.org/10.3171/2017.11.SPINE17989.
33. Brodke, D. S., Lawrence, B. D., Ryan Spiker, W., Neese, A., & Hung, M. (2014). Converting ODI or SF-36 physical function domain scores to a PROMIS PF score. *The Spine Journal, 14*(11), S50. https://doi.org/10.1016/j.spinee.2014.08.131.
34. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research, 19*(4), 539–549. https://doi.org/10.1007/s11136-010-9606-8.
35. Asher, A. L., McCormick, P. C., Selden, N. R., Ghogawala, Z., & McGirt, M. J. (2013). The National Neurosurgery Quality and Outcomes Database and NeuroPoint Alliance: rationale, development, and implementation. *Neurosurgical Focus, 34*(1), E2. https://doi.org/10.3171/2012.10.FOCUS12311.
36. Asher, A. L., Speroff, T., Dittus, R. S., Parker, S. L., Davies, J. M., Selden, N., et al. (2014). The National Neurosurgery Quality and Outcomes Database (N2QOD): a collaborative North American outcomes registry to advance value-based spine care. *Spine, 39*(22 Suppl 1), S106–116. https://doi.org/10.1097/brs.0000000000000579.
37. Bevans, M., Ross, A., & Cella, D. (2014). Patient-Reported Outcomes Measurement Information System (PROMIS): Efficient, standardized tools to measure self-reported health and quality of life. *Nursing Outlook, 62*(5), 339–345. https://doi.org/10.1016/j.outlook.2014.05.009.
38. PROMIS. (2018). PROMIS adult profile scoring manual.
39. PROMIS. (2018). PROMIS ability to participate in social roles and activities scoring manual.
40. PROMIS. (2018). PROMIS depression scoring manual.
41. PROMIS. (2018). PROMIS fatigue scoring manual.
42. PROMIS. (2018). PROMIS physical function scoring manual.
43. PROMIS. (2018). PROMIS sleep disturbance scoring manual.
44. PROMIS. (2018). PROMIS pain interference scoring manual.
45. PROMIS. (2018). PROMIS anxiety scoring manual.
46. Fairbank, J. C., & Pynsent, P. B. (2000). The Oswestry Disability Index. *Spine, 25*(22), 2940–2952. **discussion 2952**.
47. Bono, C., Ried, L. D., Kimberlin, C., & Vogel, B. (2007). Missing data on the Center for Epidemiologic Studies Depression Scale: a comparison of 4 imputation techniques. *Research in Social and Administrative Pharmacy, 3*(1), 1–27. https://doi.org/10.1016/j.sapharm.2006.04.001.
48. Choi, S. W., Schalet, B., Cook, K. F., & Cella, D. (2014). Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. *Psychological Assessment, 26*(2), 513–527. https://doi.org/10.1037/a0035768.
49. Schalet, B. D., Revicki, D. A., Cook, K. F., Krishnan, E., Fries, J. F., & Cella, D. (2015). Establishing a common metric for physical function: Linking the HAQ-DI and SF-36 PF subscale to PROMIS((R)) physical function. *Journal of General Internal Medicine, 30*(10), 1517–1523. https://doi.org/10.1007/s11606-015-3360-0.
50. Yves, R. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.
51. Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74*(1), 137–143. https://doi.org/10.1007/s11336-008-9100-1.
52. John, O. P., & Benet-Martinez, V. (2000). Measurement: Reliability, construct validation, and scale construction. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 339–370). Cambridge, UK: Cambridge University Press.
53. Taber, K. S. (2017). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*. https://doi.org/10.1007/s11165-016-9602-2.
54. Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.
55. Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.
56. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.2307/3151312.
57. Tang, E., Ekundayo, O., Peipert, J. D., Edwards, N., Bansal, A., Richardson, C., et al. (2019). Validation of the patient-reported outcomes measurement information system (PROMIS)-57 and -29 item short forms among kidney transplant recipients. *Quality of Life Research, 28*(3), 815–827.
58. Fairbank, J. C. T. (2007). Use and abuse of Oswestry Disability Index. *Spine, 32*(25), 2787–2789. https://doi.org/10.1097/BRS.0b013e31815b9e9d.
59. Rose, A. J., Bayliss, E., Huang, W., Baseman, L., Butcher, E., Garcia, R. E., et al. (2018). Evaluating the PROMIS-29 v2.0 for use among older adults with multiple chronic conditions. *Quality of Life Research, 27*(11), 2935–2944. https://doi.org/10.1007/s11136-018-1958-5.
60. Hinchcliff, M. E., Beaumont, J. L., Carns, M. A., Podlusky, S., Thavarajah, K., Varga, J., et al. (2015). Longitudinal evaluation of PROMIS-29 and FACIT-dyspnea short forms in systemic sclerosis. *The Journal of Rheumatology, 42*(1), 64–72. https://doi.org/10.3899/jrheum.140143.
61. Schalet, B. D., Hays, R. D., Jensen, S. E., Beaumont, J. L., Fries, J. F., & Cella, D. (2016). Validity of PROMIS physical function measured in diverse clinical samples. *Journal of Clinical Epidemiology, 73,* 112–118. https://doi.org/10.1016/j.jclinepi.2015.08.039.
62. Scholle, S. H., Morton, S., Homco, J., Rodriguez, K., Anderson, D., Hahn, E., et al. (2018). Implementation of the PROMIS-29 in routine care for people with diabetes: Challenges and

opportunities. *The Journal of Ambulatory Care Management, 41*(4), 274–287. https://doi.org/10.1097/JAC.0000000000000248.
63. Fries, J. F., Cella, D., Rose, M., Krishnan, E., & Bruce, B. (2009). Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. *The Journal of Rheumatology, 36*(9), 2061–2066. https://doi.org/10.3899/jrheum.090358.
64. Stone, A. A., Broderick, J. E., Junghaenel, D. U., Schneider, S., & Schwartz, J. E. (2016). PROMIS fatigue, pain intensity, pain interference, pain behavior, physical function, depression, anxiety, and anger scales demonstrate ecological validity. *Journal of Clinical Epidemiology, 74,* 194–206. https://doi.org/10.1016/j.jclinepi.2015.08.029.
65. Alexander, D. L., Tropsha, A., & Winkler, D. A. (2015). Beware of R(2): Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR Models. *Journal of Chemical Information and Modeling, 55*(7), 1316–1322. https://doi.org/10.1021/acs.jcim.5b00206.
66. Brodke, D. S., Goz, V., Voss, M. W., Lawrence, B. D., Spiker, W. R., & Hung, M. (2017). PROMIS PF CAT outperforms the ODI and SF-36 physical function domain in spine patients. *Spine, 42*(12), 921–929. https://doi.org/10.1097/brs.0000000000001965.
67. Hung, M., Saltzman, C. L., Voss, M. W., Bounsanga, J., Kendall, R., Spiker, R., et al. (2017). Responsiveness of the Patient-reported outcomes measurement information system (PROMIS), neck disability index (NDI) and Oswestry Disability Index (ODI) instruments in patients with spinal disorders. *Spine J, 224*(3), 245–254.