

Examining the Patient-Reported Outcomes Measurement Information System versus the Scoliosis Research Society–22r in adult spinal deformity

Michael P. Kelly, MD, MSc,¹ Michael A. Kallen, PhD, MPH,² Christopher I. Shaffrey, MD,³ Justin S. Smith, MD, PhD,³ Douglas C. Burton, MD,⁴ Christopher P. Ames, MD,⁵ Virginie Lafage, PhD,⁶ Frank J. Schwab, MD,⁶ Han Jo Kim, MD,⁶ Eric O. Klineberg, MD,⁷ Shay Bess, MD,⁸ and the International Spine Study Group

¹Department of Orthopedic Surgery, Washington University School of Medicine, St. Louis, Missouri; ²Northwestern University, Feinberg School of Medicine, Chicago, Illinois; ³Department of Neurological Surgery, University of Virginia, Charlottesville, Virginia; ⁴Department of Orthopedic Surgery, University of Kansas School of Medicine, Wichita, Kansas; ⁵Department of Neurological Surgery, University of California, San Francisco, California; ⁶Hospital for Special Surgery, New York, New York; ⁷Department of Orthopedic Surgery, University of California, Davis, Sacramento, California; and ⁸Department of Spine Surgery, Denver International Spine Clinic, Presbyterian St. Luke's/Rocky Mountain Hospital for Children, Denver, Colorado

OBJECTIVE After using PROsetta Stone crosswalk tables to calculate Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function (PF) and Pain Interference (PI) scores, the authors sought to examine 1) correlations with Scoliosis Research Society–22r (SRS-22r) scores, 2) responsiveness to change, and 3) the relationship between baseline scores and 2-year follow-up scores in adult spinal deformity (ASD).

METHODS PROsetta Stone crosswalk tables were used to convert SF-36 scores to PROMIS scores for pain and physical function in a cohort of ASD patients with 2-year follow-up. Spearman correlations were used to evaluate the relationship of PROMIS scores with SRS-22r scores. Effect size (ES) and adjusted standardized response mean (aSRM) were used to assess responsiveness to change. Linear regression was used to evaluate the association between baseline scores and 2-year follow-up scores.

RESULTS In total, 425 (425/625, 68%) patients met inclusion criteria. Strong correlations (all $|r| > 0.7$, $p < 0.001$) were found between baseline and 2-year PROMIS values and corresponding SRS-22r domain scores. PROMIS-PI showed a large ES (1.09) and aSRM (0.88), indicating good responsiveness to change. PROMIS-PF showed a moderate ES (0.52) and moderate aSRM (0.69), indicating a moderate responsiveness to change. Patients with greater baseline pain complaints were associated with greater pain improvement at 2 years for both SRS-22r Pain ($B = 0.39$, $p < 0.001$) and PROMIS-PI ($B = 0.45$, $p < 0.001$). Higher functional scores at baseline were associated with greater average improvements in both SRS-22r Activity ($B = 0.62$, $p < 0.001$) and PROMIS-PF ($B = 0.40$, $p < 0.001$).

CONCLUSIONS The authors found strong correlations between the SRS-22r Pain and Activity domains with corresponding PROMIS-PI and -PF scores. Pain measurements showed similar and strong ES and aSRM while the function measurements showed similar, moderate ES and aSRM at 2-year follow-up. These data support further exploration of the use of PROMIS–computer adaptive test instruments in ASD.

<https://thejns.org/doi/abs/10.3171/2018.11.SPINE181014>

KEYWORDS adult deformity; scoliosis; spine; PROMIS; outcomes

MEASUREMENT of health-related quality of life (HRQOL) outcomes in medicine is a necessary endeavor, as it makes possible comparisons of intervention effectiveness and evaluations of quality of care.^{14,15} HRQOL measurements also aid healthcare-relat-

ed decision making. In adult spinal deformity (ASD) surgery, for example, HRQOL measurements support patient and surgeon decision making with respect to both appropriate timing for surgery and the outcome expectations of patients and surgeons alike. HRQOL instruments may be

ABBREVIATIONS ASD = adult spinal deformity; aSRM = adjusted standardized response mean; BP = Bodily Pain; CAT = computer adaptive test; ES = effect size; HRQOL = health-related quality of life; ODI = Oswestry Disability Index; PCS = Physical Component Summary; PF = Physical Function; PI = Pain Interference; PROMIS = Patient-Reported Outcomes Measurement Information System; SF-36 = Short Form–36; SRM = standardized response mean; SRS-22r = Scoliosis Research Society–22r. **SUBMITTED** August 18, 2018. **ACCEPTED** November 6, 2018.

INCLUDE WHEN CITING Published online February 22, 2019; DOI: 10.3171/2018.11.SPINE181014.

generic measures of health, such as the Short Form–36 (SF-36), or disease-specific measures, such as the Scoliosis Research Society–22r (SRS-22r) questionnaire for ASD.^{1,13}

Generic measures of health can be said to be universally applicable. They provide a picture of a patient's function or well-being regardless of that patient's disease status or comorbid conditions. The generic SF-36 comprises multiple subdomains and assesses both physical and mental health status. This tool has undergone extensive psychometric testing to ensure its internal and external validity as well as its internal consistency and test-retest reliability.^{9,10} The Patient-Reported Outcomes Measurement Information System (PROMIS), funded by the National Institutes of Health (NIH), is a system of generic health measures and is available to clinicians and researchers alike for computer adaptive test (CAT) administration, a highly effective and efficient method of measure administration, based on detailed item performance parameters derived from supporting item response theory (IRT) analyses. In short, CATs allow for the collection of precise patient-centered outcome scores while minimizing response burden and overall measure administration time. PROMIS-CAT scores are reported on a T-score metric and are typically centered so that a score of 50 represents the average score (i.e., mean) of the US general population for the particular domain being measured; T-score standard deviations are in 10-point increments. Thus, the PROMIS generic measurement system, delivered as CATs with available population-based score interpretations, allows for the rapid evaluation of patient HRQOL, including disability and pain complaints, across various diseases. The use of a common outcomes instrument across clinical practices may facilitate economic and cost-effectiveness analyses between distinct diseases, comparing, for example, the disease burden and relative cost-effectiveness of treatment for ASD and hip osteoarthritis.

The SRS-22r is a disease-specific instrument for spinal deformity and is valid and responsive to change in ASD.² Both the SF-36 and PROMIS, which are generic in orientation, have domains that may correlate with SRS-22r domains: specifically, SF-36 Physical Function (PF) and PROMIS-PF with SRS-22r Activity; SF-36 Bodily Pain (BP) and PROMIS-Pain Interference (PI) with SRS-22r Pain. The PROsetta Stone Project, a legacy-to-PROMIS score-linking project, has provided both a score-linking methodology for use in future measure linking studies and a set of ready-to-apply score conversions or crosswalks. Currently, a PROsetta Stone crosswalk exists for converting SF-36 domain scores to corresponding PROMIS domain scores (i.e., to place SF-36 scores on the PROMIS T-score metric).⁴ The PROsetta Stone library of crosswalks was developed using three different “competing” state-of-the-science psychometric linking methods to identify a single best approach for estimating the corresponding PROMIS score of interest. One major purpose of the creation of the PROsetta Stone score-linking crosswalk products was to allow researchers who had used legacy instruments in their studies to convert their measurements to the PROMIS metric for some or all of their analyses, thus enabling the reporting of findings on the widely comparable PROMIS score metric.

The purpose of this study was to use PROsetta Stone crosswalks to convert SF-36 BP and PF scores from an existing ASD cohort to their corresponding PROMIS scores. We examined the responsiveness to change of these converted PROMIS-PI and PROMIS-PF scores after a minimum of 2 years' follow-up after ASD surgery and compared these findings with legacy measurement responsiveness to change. We hypothesized that significant improvements in patient PROMIS scores would occur and that PROMIS scores would strongly correlate with SRS-22r instrument legacy scores. Use of a generic health measure in ASD may allow for immediate comparison to population norms and provide a reference against which to evaluate change. This could be of considerable benefit, as prior work in spine surgery has shown inconsistent measurement of change and disability using disease-specific measures.⁶ Furthermore, positive results from this study may help support the greater integration of PROMIS into ASD practices, including help guiding health domain measurement choices, as practitioners move to PROMIS for HRQOL assessment.

Methods

A multicenter, prospective, observational cohort of ASD patients was reviewed to identify patients who had enrolled from January 2008 to August 2014. Institutional Review Board approval was obtained at all sites, and all patients consented to enrollment in the prospective cohort. Patients were eligible for analysis if their baseline and 2-year postoperative SF-36 scores were available. All patients had been treated via surgery, with their surgery having been chosen at the discretion of the attending surgeon. Crosswalk tables created by the PROsetta Stone Project were used to estimate PROMIS-PI scores from SF-36 BP scores, and PROMIS-PF scores from SF-36 PF scores.⁴ Median scores and ranges were obtained for all HRQOL measurement data. Baseline data were compared between those included in the analysis and those lost to follow-up. No corrections for multiple comparisons were made.

Spearman's rho correlation coefficients, which are non-parametric and appropriate for the study sample's non-normally distributed SRS-22r data, were calculated between PROMIS-PI and SRS-22r Pain scores and between PROMIS-PF and SRS-22r Activity scores for all patients with corresponding data. Evans' criteria were used to evaluate the strength of obtained correlations, where a correlation coefficient of 0.00 to 0.19 is “very weak,” 0.20 to 0.39 is “weak,” 0.40 to 0.59 is “moderate,” 0.60 to 0.79 is “strong,” and 0.80 to 1.00 is “very strong.”⁷

Effect sizes (ESs) and standardized response means (SRMs) were calculated for PROMIS-PI, PROMIS-PF, SF-36 Physical Component Summary (PCS), SRS-22r Pain, and SRS-22r Activity to estimate responsiveness to change. ESs were calculated as mean change divided by the standard deviation of baseline scores. Thus, ES here represents a measure of the amount of change observed, in terms of the observed baseline score standard deviation. ESs were evaluated by the method proposed by Cohen, where a large effect size is 0.80 or greater, a moderate effect is from 0.50 to 0.79, and a small effect is from 0.20 to

TABLE 1. Demographic and descriptive data of adult spinal deformity cohort

Preop Descriptive Variable	Value
Age in yrs	58.0 (14.9)
Sex	
Male	93 (22%)
Female	332 (78%)
CCI score	1 (0–8)
Prior spine surgery	
No	217 (51%)
Yes	208 (49%)
Prior spine fusion	
No	372 (87%)
Yes	53 (13%)
Thoracic coronal Cobb angle in degrees	30.4 (20.5)
Lumbar coronal Cobb angle in degrees	24.1 (12.8)
C7 sagittal vertical axis in mm	63 (74)
Pelvic tilt in degrees	23.8 (10.9)
PI-LL mismatch in degrees	15.8 (21.2)
ODI score	44 (0–92)
SRS-22r total score	2.8 (1.1–4.6)

CCI = Charlson Comorbidity Index; LL = lumbar lordosis; PI = pelvic incidence. Values are presented as mean (SD), number of patients (%), or median (range).

0.49.⁵ SRMs were calculated as mean change divided by the standard deviation of the mean change. SRMs were adjusted according to the method of Middel and van Sonderen to avoid over- or underestimation of responsiveness to change and interpreted by Cohen's method, as previously described.¹¹ Thus, adjusted SRMs (aSRMs) were used to account for the fact that baseline score status influences the amount of change possible for any given patient (e.g., those in the middle of the score range at baseline measurement are able to improve more than those already at the top of the score range).

We employed linear regression to evaluate the impact of baseline scores on final follow-up scores. Changes in scores were plotted against baseline scores to investigate improvement that had occurred, dependent on score starting value. For these analyses, statistical significance was set at the $p < 0.05$ level. All linear regression analyses were considered as exploratory in nature, and no corrections for multiple comparisons were made. Analyses were conducted using IBM SPSS v23.0 (IBM Corp.).

Funding

DePuy Synthes Spine provides direct research support to the International Spine Study Group Foundation.

Results

In total, 625 ASD patients were enrolled in our study, and 425 (68%) had both baseline and 2-year follow-up SF-36 data, allowing for calculations of PROMIS measures and inclusion in planned analyses. Sample demographic and descriptive characteristics are presented in Table 1.

TABLE 2. Scores at baseline and 2-year follow-up

Domain	Baseline (n = 425)	2 Yrs Postop (n = 425)
Measures of pain		
SF-36 BP	8 (2–11)	6 (2–11)
SRS-22r Pain	2.4 (1.0–5.0)	3.4 (1–5)
PROMIS-PI	63.5 (37.8–76.0)	56.6 (37.8–76)
Measures of function		
SF-36 PF	17 (10–30)	21 (10–30)
SRS-22r Activity	3.0 (1.0–5.0)	3.6 (1.2–5)
PROMIS-PF	37.2 (24.5–61.7)	41.8 (24.5–61.7)

Values are presented as median (range).

The sample comprised mostly women (78%); there was a wide range of observed radiographic deformity magnitudes and disability, as measured by the SRS-22r. Those patients lost to follow-up had higher baseline Oswestry Disability Index (ODI) scores (mean difference 3.6, 95% CI 0.3–6.9, $p = 0.03$). No other baseline data were significantly different.

PROsetta Stone crosswalks provided the basis for calculating PROMIS-PI and PROMIS-PF scores at baseline and 2 years postoperatively (Table 2). Strong correlations (all $r \geq 0.7$, $p < 0.001$) were found between baseline and 2-year PROMIS scores and corresponding SRS-22r domain scores (Table 3). The median change in PROMIS-PI was 7.6 points (range –19.4 to –38.2). The median change in PROMIS-PF was –4.5 points (range –28.3 to –33.4). PROMIS-PI showed a large ES (1.09) and a large aSRM (0.88), indicating overall good responsiveness to change. PROMIS-PF showed a moderate ES (0.52) and a moderate adjusted SRM (0.69), indicating overall moderate responsiveness to change (Table 4). ESs and SRMs were similar for PROMIS-PI and SRS-22r Pain. ESs and SRMs were also similar for PROMIS-PF, SRS-22r Activity, and SF-36 PCS, though they were lower than the pain measurements, indicating less responsiveness to change within the ASD cohort. All function-domain measurements exhibited moderate ES (range 0.52–0.75).

Greater baseline pain complaints (higher PROMIS-PI) or functional limitations (lower PROMIS-PF) were associated with larger improvements in PROMIS-PI and PROMIS-PF scores at 2-year follow-up (Fig. 1). Patients

TABLE 3. Spearman rho correlations between corresponding baseline and 2-year PROMIS and SRS-22r domains

	SRS-22r Pain		SRS-22r Activity	
	Baseline	2 Yrs	Baseline	2 Yrs
PROMIS-PI				
Baseline		–0.73*		
2 yrs				–0.85*
PROMIS-PF				
Baseline			0.75*	
2 yrs				0.81*

* $p < 0.001$.

who presented without pain complaints (PROMIS-PI ≤ 50) were rare ($n = 27$, 6%). Median change in PROMIS-PI for this group showed an increase in pain (median -0.3 ; range -15 to 12.1). Patients with greater baseline pain complaints were associated with greater pain improvement at 2-year follow-up for both SRS-22r Pain ($B = 0.39$, $p < 0.001$) and PROMIS-PI ($B = 0.45$, $p < 0.001$). Patients with no physical dysfunction (PROMIS-PF ≥ 50) were rare ($n = 28$, 7%). These patients experienced a worsening of PROMIS-PF at 2-year follow-up (median 2.5 ; range -9.7 to 33.4). Patients reporting severe PROMIS-PF complaints (PROMIS-PF ≤ 30) experienced the greatest improvement in function scores (median -7.6 ; range -27.5 to -3.8). Worse functional status at baseline was associated with greater average improvements in both SRS-22r Activity ($B = 0.62$, $p < 0.001$) and PROMIS-PF ($B = 0.40$, $p < 0.001$).

Discussion

PROMIS is a generic HRQOL measurement instrument system designed to be applicable across a wide range of ages and diseases. A CAT method of administration is available, which allows for faster collection of HRQOL scores, minimizing response burden for the patient while maintaining score precision standards. As PROMIS measures are new outcomes instruments for ASD, their responsiveness to change and correlation to legacy measures have not been examined in the ASD context. The PROsetta Stone Project's crosswalk products allow for the conversion of specific legacy measure scores, including SF-36 BP and PF scores, to corresponding PROMIS domain scores.⁴ The purpose of this paper was to examine the responsiveness to change of PROMIS-PI and PROMIS-PF scores among ASD patients and to determine the potential suitability of these measures for collection as measures of health in ASD. SF-36 BP and PF scores were converted to PROMIS-PI and PROMIS-PF scores, respectively, using PROsetta Stone linking tables. The properties of those linked scores were examined, relative to legacy measure scores from SF-36 PCS and SRS-22r Pain and Activity in a cohort of surgically treated ASD patients.

We found strong correlations (all $|r| \geq 0.7$) between crosswalked PROMIS-PI and SRS-22r Pain scores and PROMIS-PF and SRS-22r Activity scores. These strong correlations indicate that PROMIS-PI and PROMIS-PF may assess pain and physical function similarly to the SRS-22r instrument. The responsiveness of corresponding PROMIS and SRS-22r domains to change was also similar. The pain domain tools showed excellent responsiveness to change, with $ES > 1$ for both and $aSRM > 0.8$. The activity domains did not perform as well, with lower ES for both PROMIS-PF and SRS-22r Activity. The SF-36 PCS exhibited the highest ES and $aSRM$, indicating this instrument was the most sensitive to change. The greater responsiveness to change of the SF-36 PCS may be a reflection of the increased question burden, which PROMIS-CATs are designed to minimize. The overall poor activity domain performances may be reflections of the surgeries themselves and not the outcome measure instruments, as pain may be more favorably affected by surgery than function, due to the nature of these long-fusion procedures. This theory is

TABLE 4. ES, SRM, and aSRM for HRQOL measurement instruments

	ES	SRM	aSRM*
PROMIS-PI	1.09	0.92	0.88
PROMIS-PF	0.52	0.60	0.69
SF-36 PCS	0.75	0.79	0.85
SRS-22r Pain	1.1	0.9	0.82
SRS-22r Activity	0.67	0.75	0.86

* According to Middel and van Sonderen.

supported by the decline in function observed for the few patients who did not report poor function at baseline.

Those patients with the most disability, in terms of both pain and function, received the greatest benefits from surgery. As seen in Fig. 1, lower baseline scores were associated with the greatest improvements at the 2-year follow-up. Similarly, if we consider a score of 50 to be normal, then those patients who had normal or better status were unlikely to receive a benefit from ASD surgery. These data are important, as they serve as a starting point for using PROMIS HRQOL measures in the shared decision-making process for ASD treatment. One must consider that the overall disease state is a combination of the various measured domains and that some tradeoff might be considered worthwhile to patients. For example, a patient with severe pain but well-preserved function may appreciate the improvement in pain that comes with a loss of function. This is seen in the decrease in PROMIS-PF scores observed in those starting with higher PROMIS-PF. Thus, surgery cannot be deemed a failure (or a success) based solely on the results observed in any single health status domain. Knowing precisely what the expected outcomes of surgery are, and which health status domains are expected to be impacted, will greatly assist with decision making and help set realistic expectations.

While no comparisons between PROMIS-PF and PROMIS-PI CATs and the SRS-22r instrument currently exist, our sample-observed correlations are consistent with prior work in spinal diseases.^{1,3,8,12} PROMIS-PF and PROMIS-PI have exhibited moderate to strong correlations with the ODI, SF-36 PF, and the Neck Disability Index (NDI).^{1,12} That both PROMIS domains are strongly correlated with the ODI and NDI emphasizes the inherent strengths and weaknesses of disease-specific HRQOL measures. The ODI and NDI are unlikely to discriminate between pain and function complaints, rather offering a composite of a patient's health. The PROMIS measures allow for assessment of individual complaints, accomplished using a similar number of administered items and with a more readily interpreted score output. The PROMIS-PF item bank has been shown to be unidimensional in spine disease, meaning that it examines function as an individual trait.⁸ The PROMIS-PF item bank and its CATs are reliable and valid, with small ceiling or floor effects. The latter is important when one considers postoperative results. If patients cluster around a particular value after surgery, then the result is more likely attributable to limits of the procedure itself rather than an inability of the instrument to detect differences between patients. Among the implications of this

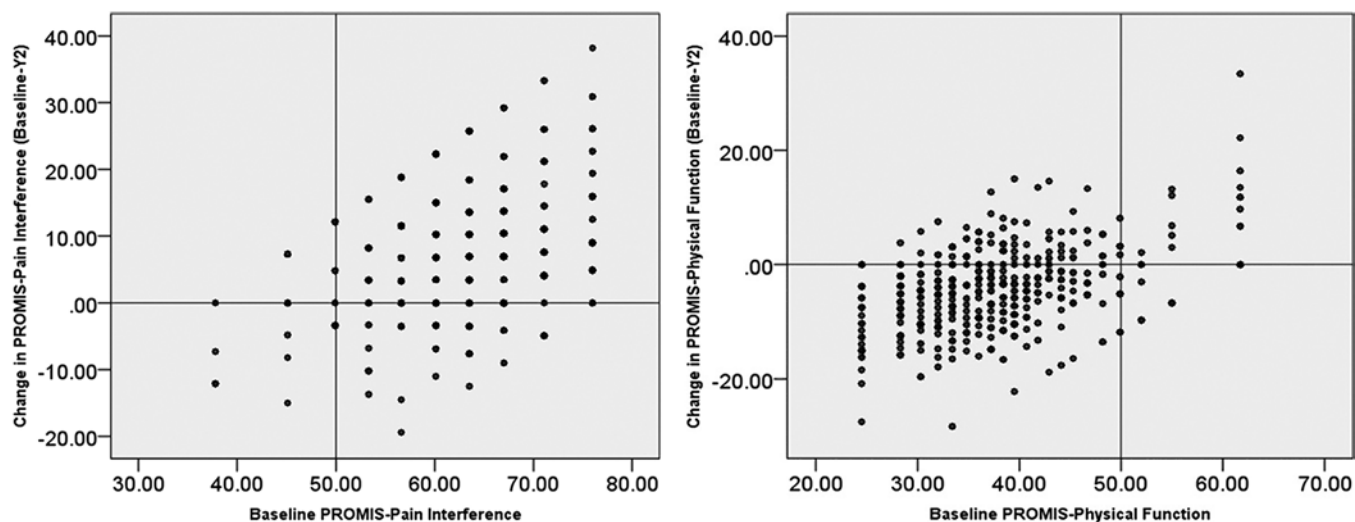


FIG. 1. Left: Baseline crosswalked PROMIS-PI versus change in PROMIS-PI at 2 years postoperatively. Higher scores indicate greater disability. **Right:** Baseline crosswalked PROMIS-PF versus change in PROMIS-PF at 2 years postoperatively. Lower scores indicate worse disability.

is that ASD expectations may be set during preoperative counseling.

Our study is limited primarily by employing indirectly obtained PROMIS-PF and PROMIS-PI scores, i.e., using crosswalked SF-36 PF and BP to PROMIS-PF and PROMIS-PI scores, respectively. Crosswalked scores are ideal for group-based comparisons, yet they may not meet score precision standards for examining individual-level patient score differences. Thus, future work in the ASD population will use directly obtained PROMIS scores. However, our study confirms the great potential of using such PROMIS generic measures with ASD patients and provides evidence that the collection of PROMIS-PF and PROMIS-PI measurement data could enhance ASD-related care and decision making. As previously described, the PROSetta Stone Project uses state-of-the-science item response theory methods to create score-linking crosswalks. The SF-36 instrument is a general measure of health and is valid across multiple disease states. ASD is a heterogeneous disease, and there may be subsets of ASD that perform better or worse according to PROMIS measurement (e.g., revision deformities requiring osteotomies versus less complex, primary deformities). Regardless, as a general measure of health the PROMIS domains, and ES/SRM, should not be affected by this. Finally, the SRS-22r comprises more than just pain and function measurements. It also includes measures of self-image and mental health domains. Some measurement of these through a PROMIS-CAT will be required if PROMIS is to supplant the SRS-22r legacy measure. The SRS-22r mental health domain is nonspecific for psychiatric disease, and PROMIS measures include CATs for both anxiety and depression. Thus, further work is needed to understand the most relevant mental health measures for ASD decision making.

Conclusions

In conclusion, we present the first assessment of PROMIS-PI and PROMIS-PF in an ASD population using

PROMIS scores obtained from a crosswalk developed from SF-36-to-PROMIS linking analyses. We have found strong correlations between the SRS-22r Pain and Activity domains with corresponding PROMIS-PI and -PF scores. Furthermore, the pain assessments showed similar and strong ESs and SRMs, while the function assessments had similar, moderate ESs and SRMs at 2-year follow-up. These data support further exploration of the PROMIS-CAT instruments in ASD, as they may offer a reliable method of HRQOL assessment with a minimum response burden. However, development of specific spinal deformity measures may be needed for self-image and disease-relevant mental health assessments.

References

1. Boody BS, Bhatt S, Mazmudar AS, Hsu WK, Rothrock NE, Patel AA: Validation of Patient-Reported Outcomes Measurement Information System (PROMIS) computerized adaptive tests in cervical spine surgery. *J Neurosurg Spine* **28**:268–279, 2018
2. Bridwell KH, Cats-Baril W, Harrast J, Berven S, Glassman S, Farcy JP, et al: The validity of the SRS-22 instrument in an adult spinal deformity population compared with the Oswestry and SF-12: a study of response distribution, concurrent validity, internal consistency, and reliability. *Spine (Phila Pa 1976)* **30**:455–461, 2005
3. Brodke DS, Goz V, Voss MW, Lawrence BD, Spiker WR, Hung M: PROMIS PF CAT outperforms the ODI and SF-36 Physical Function domain in spine patients. *Spine (Phila Pa 1976)* **42**:921–929, 2017
4. Choi SW, Podrabsky T, McKinney N, Schalet BD, Cook KF, Cella D: **PROSetta Stone (TM) Analysis Report: A Rosetta Stone for Patient Reported Outcomes, Vol 1**. Chicago, IL: Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, 2012
5. Cohen J: **Statistical Power Analysis for the Behavioral Sciences, ed 2**. Hillsdale, NJ: L Erlbaum Assoc, 1988
6. Copay AG, Martin MM, Subach BR, Carreon LY, Glassman SD, Schuler TC, et al: Assessment of spine surgery outcomes: inconsistency of change amongst outcome measurements. *Spine J* **10**:291–296, 2010

7. Evans JD: **Straightforward Statistics for the Behavioral Sciences**. Pacific Grove, CA: Brooks/Cole Publishing Co, 1996
8. Hung M, Hon SD, Franklin JD, Kendall RW, Lawrence BD, Neese A, et al: Psychometric properties of the PROMIS physical function item bank in patients with spinal disorders. **Spine (Phila Pa 1976)** **39**:158–163, 2014
9. McHorney CA, Ware JE Jr, Raczek AE: The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. **Med Care** **31**:247–263, 1993
10. McHorney CA, Ware JE Jr, Rogers W, Raczek AE, Lu JF: The validity and relative precision of MOS short- and long-form health status scales and Dartmouth COOP charts. Results from the Medical Outcomes Study. **Med Care** **30** (5 Suppl):MS253–MS265, 1992
11. Middel B, van Sonderen E: Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. **Int J Integr Care** **2**:e15, 2002
12. Patel AA, Dodwad SM, Boody BS, Bhatt S, Savage JW, Hsu WK, et al: Validation of Patient Reported Outcomes Measurement Information System (PROMIS) Computer Adaptive Tests (CATs) in the surgical treatment of lumbar spinal stenosis. **Spine (Phila Pa 1976)** **43**:1521–1528, 2018
13. Patel AA, Donegan D, Albert T: The 36-Item Short Form. **J Am Acad Orthop Surg** **15**:126–134, 2007
14. Porter ME: What is value in health care? **N Engl J Med** **363**:2477–2481, 2010
15. Porter ME, Lee TH: From volume to value in health care: the work begins. **JAMA** **316**:1047–1048, 2016

Disclosures

Dr. Kelly reports receiving clinical or research support for the study described (includes equipment or material) from DePuy Synthes Spine. Dr. Shaffrey reports being a consultant for Medtronic, NuVasive, and EOS; direct stock ownership in NuVasive; and being a patent holder for Medtronic, NuVasive, and Zimmer Biomet. Dr. Smith reports being a consultant for Zimmer Biomet, NuVasive, K2M, AlloSource, and Cerapedics; receiving royalties from Zimmer Biomet; and receiving fellowship funding from NREF and AOSpine. Dr. Burton reports receiving clinical or research support for the study described (includes equipment or material) from DePuy Synthes Spine, Pfizer, and Bioventus;

and being a consultant for DePuy Synthes Spine and AlloSource. Dr. Ames reports being an employee of UCSF; being a consultant for DePuy Synthes Spine, Medtronic, Stryker, Medtronic, K2M, and Biomet Zimmer; receiving royalties from Stryker, Biomet Zimmer, DePuy Synthes Spine, NuVasive, Next Orthosurgical, K2M, and Medtronic; performing research for Biomet Zimmer, DePuy Synthes Spine, and ISSG; serving on the editorial board of *Operative Neurosurgery*; receiving grant funding from SRS; serving on the Executive Committee of ISSG; and serving as director of Global Spine Analytics. Dr. Lafage reports direct stock ownership in Nemaris, Inc.; oversight of support of non-study-related clinical or research effort for DePuy Synthes Spine, NuVasive, K2M, and Stryker and of grants from NASS and SRS; and having speaking/teaching arrangements with AOSpine, DePuy Synthes Spine, and K2M. Dr. Schwab reports direct stock ownership in Nemaris, Inc.; oversight of support of non-study-related clinical or research effort for DePuy Synthes Spine, NuVasive, K2M, and Stryker and grants paid through ISSGF; being a consultant for Zimmer Biomet, K2M, MSD, NuVasive, and Medtronic; and having speaking/teaching arrangements with Zimmer Biomet, K2M, MSD, and NuVasive. Dr. Klineberg reports being a consultant for DePuy Synthes Spine, Stryker, K2M, Springer, Trevana, AlloSource, and AOSpine; and receiving a grant from AOSpine. Dr. Bess reports being a consultant for EOS, K2M, Misonix, and AlloSource; being a patent holder for K2M; receiving clinical or research support for the study described (includes equipment or material) from the International Spine Study Group Foundation; and oversight of support of non-study-related clinical or research effort for the International Spine Study Group Foundation. DePuy Synthes Spine provides direct research support to the International Spine Study Group Foundation.

Author Contributions

Conception and design: Kelly. Acquisition of data: Kelly, Shaffrey, Smith, Burton, Ames, Lafage, Schwab, Kim, Klineberg, Bess. Analysis and interpretation of data: Kelly, Kallen. Drafting the article: Kelly, Kallen. Critically revising the article: Kelly, Shaffrey, Smith, Ames, Bess. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Kelly. Statistical analysis: Kelly. Administrative/technical/material support: Lafage, Bess.

Correspondence

Michael P. Kelly: Washington University School of Medicine, St. Louis, MO. kellymi@wustl.edu.