Validation of Patient Reported Outcomes Measurement Information System (PROMIS) Computer Adaptive Tests (CATs) in the Surgical Treatment of Lumbar Spinal Stenosis

Alpesh A. Patel, MD, FACS[*], **Shah-Nawaz M. Dodwad, MD**[†], **Barrett S. Boody, MD**[*], **Surabhi Bhatt, BS**[*], **Jason W. Savage, MD**[‡], **Wellington K. Hsu, MD**[*], **Nan E. Rothrock, PhD**[§]

[*]Department of Orthopaedic Surgery, Northwestern University Feinberg School of Medicine, Chicago, Illinois;

[†]Department of Orthopaedic Surgery, University of Texas – Houston, Houston, Texas;

[‡]Cleveland Clinic, Center for Spine Health, Cleveland, Ohio;

[§]Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

Abstract

Study Design.—Prospective, cohort study.

Objective.—Demonstrate validity of Patient reported outcomes measurement information system (PROMIS) physical function, pain interference, and pain behavior computer adaptive tests (CATs) in surgically treated lumbar stenosis patients.

Summary of Background Data.—There has been increasing attention given to patient reported outcomes associated with spinal interventions. Historical patient outcome measures have inadequate validation, demonstrate floor/ceiling effects, and infrequently used due to time constraints. PROMIS is an adaptive, responsive National Institutes of Health (NIH) assessment tool that measures patient-reported health status.

Methods.—Ninety-eight consecutive patients were surgically treated for lumbar spinal stenosis and were assessed using PROMIS CATs, Oswestry disability index (ODI), Zurich Claudication Questionnaire (ZCQ), and Short-Form 12 (SF-12). Prior lumbar surgery, history of scoliosis, cancer, trauma, or infection were excluded. Completion time, preoperative assessment, 6 weeks and 3 months postoperative scores were collected.

Results.—At baseline, 49%, 79%, and 81% of patients had PROMIS pain behavior (PB), pain interference (PI), and physical function (PF) scores greater than 1 standard deviation (SD) worse than the general population. 50.6% were categorized as severely disabled, crippled, or bed bound by ODI. PROMIS CATs demonstrated convergent validity through moderate to high correlations with legacy measures ($r = 0.35$–0.73). PROMIS CATs demonstrated known groups validity when stratified by ODI levels of disability. ODI improvements of at least 10 points on average had

Address correspondence and reprint requests to Alpesh A. Patel, MD, FACS, 676 N. St. Clair St., #1350, Chicago, IL 60611; alpesh.patel@nm.org.

changes in PROMIS scores in the expected direction (PI = −12.98, PB = −9.74, PF = −7.53). PROMIS CATs demonstrated comparable responsiveness to change when evaluated against legacy measures. PROMIS PB and PI decreased 6.66 and 9.62 and PROMIS PF increased 6.8 points between baseline and 3-months post-op ($P < 0.001$). Completion time for the PROMIS CATs (2.6 min) compares favorably to ODI, ZCQ, and SF-12 scores (3.1, 3.6, and 3.0 min).

Conclusion.—PROMIS CATs demonstrate convergent validity, known groups validity, and responsiveness for surgically treated patients with lumbar stenosis to detect change over time and are more efficient than legacy instruments.

Keywords

computer adaptive tests; lumbar spinal stenosis; Oswestry Disability Index; pain; patient reported outcomes; physical function; PROMIS; short-form 12; Zurich claudication questionnaire

Lumbar spinal stenosis (LSS) is defined as a narrowing of the lumbar spinal canal that can lead to pain and disability (Figure 1). The disease usually occurs beyond the 5th decade of life and the incidence increases with age with a prevalence of 1.7% to 10%.[1–4] LSS causes low back pain and neurogenic claudication with bilateral lower extremity pain, numbness, tingling, and weakness with ambulation or standing. Surgical treatment is offered once conservative options, including medications, epidural steroid injections, and physical therapy, have failed. Compared with nonoperative care, surgical interventions for symptomatic lumbar stenosis have demonstrated significantly better outcomes with respect to pain and function.[5,6]

The recent focus on high quality, cost-conscious health care requires a better understanding of the effect of medical and surgical treatments on patient reported quality of life. Patient-reported outcome (PRO) instruments are used to enhance objective clinical data, capture the patients' perception of treatment efficacy, well-being, quality of life, physical function, pain, and satisfaction.[7] Traditional PROs for LSS include the Zurich Claudication Questionnaire (ZCQ), Oswestry disability index (ODI), and Short-Form 12 (SF-12). Psychometric limitations of these measures include disease bias, inefficiency, and impreciseness at the extremes of function.[8] Floor effects, or the inability to distinguish low function scores in PROs, are a significant issue for surgical LSS patients given their baseline level of disability and pain, thus hindering quantifying outcome changes.

The goal of the patient-reported outcomes measurement information system (PROMIS) is to develop a validated system of PRO measures that are universal across chronic conditions and demographic groups.[8,9] The reliability and validity of PROMIS measures are demonstrated in a variety of pathology including depression, cancer, chronic obstructive pulmonary disease, heart failure, and other pathologies.[8,10–14] PROMIS measures include computer adaptive tests (CATs). CATs offer precision and validity while using a smaller, targeted subset of questions administered from a large pool of items, thereby reducing time needed to complete PROs and potentially improving utilization.[15–20]

The validity of using PROMIS CATs in surgical treatment of LSS is unknown. The purpose of this study is to evaluate the convergent validity, known groups validity, and

responsiveness to change of PROMIS CATs in patients undergoing surgical treatment of LSS.

MATERIALS AND METHODS

After institutional approval, all consecutive patients undergoing surgery for the treatment of symptomatic LSS between 18 and 95 years were enrolled. All patients had attempted and failed nonoperative care and deemed surgical candidates by one of three fellowship trained spine surgeons (A.P., J.S., and W.H.). Patients with prior lumbar surgery, non-English speaking, a history of scoliosis, cancer, trauma, or infection were excluded. Patients completed the PRO assessment using wireless internet tablets using the Assessment Center^{SM}, as a web-based, online data collection tool used.

Assessments occurred preoperatively (visit 1) and 6 weeks (visit 2) and 3 months (visit 3) postoperatively using individual secure login. Baseline assessments were completed in clinic. Postoperative assessments were completed via telephone or internet. Patients unable to use the iPad had the study coordinator read questions out loud and enter their response. At each time point, patients were administered the PROMIS pain behavior CAT (PB), PROMIS pain interference CAT (PI), PROMIS physical function CAT (PF), ODI, ZCQ, and SF-12. Global assessment of change was captured at 6 weeks and 3 months postoperatively. Patients were additionally asked about comorbid conditions to assess the influence of comorbidities on pain, physical function, as well as a global rating of change to assess the patient's perception of change between assessments. The time for completion of each PRO was captured through the Assessment center software.

MEASURES

ZCQ is a disease-specific PRO regarding LSS. There are 12 questions and an additional six questions for those who have undergone treatment. ZCQ evaluates symptom severity, physical function, and satisfaction with treatment. The higher the score, the higher is the level of disability.

ODI is a PRO intended to evaluate the limitations of different activities of daily living. It is comprised of 10 sections, scored on a 0 to 5 scale, 5 representing the most significant disability. ODI score is calculated by dividing the summed score by the total possible score, which is then multiplied by 100 and conveyed as a percentage.

SF-12 is a 12-item measure that evaluates physical, social, and mental function. It is expressed as a physical component score (PCS) and mental component score (MCS). The SF-12 scale uses a T-score (general population mean = 50, SD = 10) with greater scores indicating improved health.

PROMIS CATs are administered using an algorithm that uses the previous question response to identify the appropriate subsequent targeted question. CATs stop questions when a specific measurement precision (standard error <3.0) or fixed number of items (12) is reached. Therefore, 4 to 12 are administered to a patient. PROMIS utilizes T-scores, where 50 points reflect the general population mean (SD = 10). The PROMIS PF CAT v1.2 is

administered from 121 items and evaluates capability for physical activities. Higher scores indicate better physical function. The PROMIS PI CAT v1.0 assesses how pain interferes with activity and has 41 items. The PROMIS PB CAT v1.0 has 39 items and evaluates verbal and nonverbal expressions of pain. For PROMIS PI and PB CATs, higher scores indicate more pain or expressions of pain.

The impactful comorbid condition question assesses the influence of comorbidities on pain and physical function. The question "Are your answers to today's questions being affected by any conditions (*i.e.*, arthritis, knee pain, heart disease, lung disease, *etc.*) other than what you are being seen for today?" is answered yes/no.

The global rating of change question evaluates perception of change between assessments to evaluate responsiveness ("How is your neck or back condition since your last visit with us?"). Responses were "much better," "slightly better," "about the same," "slightly worse," and "much worse."

Statistical Analysis

PROMIS CAT scores were exported from Assessment Center[SM]. SF-12 PCS and MCS scores were calculated using the QualityMetric Health Outcomes(Lincoln, RI USA)™ Scoring Software 4.5. ODI scores were calculated according to developers' instructions as the percentage of total possible points.

In order to test discriminant (known-groups) validity, patients were grouped by disease severity at baseline as measured by the ODI as well as by level of limitation in activity or work (SF-12 item 3a). PROMIS was compared across groups using single Student *t* tests.

In order to evaluate responsiveness, the PROMIS CAT and legacy measures were compared across time for those respondents with data from all three assessments. Changes between assessments were calculated for all measures. Mean change from baseline (visit 1) scores were compared using single Student *t* tests. Pearson correlation coefficients were also calculated using the change scores in order to evaluate responsiveness over time. The Global Assessment of Change responses were collapsed into two groups: those who reported feeling "much better" and all others. The standardized response mean (SRM = mean change/SD of change) was calculated to quantify the relative level of change within these groups. Effect sizes (mean difference divided by pooled SD) were calculated to provide standardized estimates of group differences.

Changes were assessed for reaching minimal clinic important difference (MCID) thresholds. Change scores were compared with MCID estimate for the following measures; PROMIS PI, PB, and PF 50% of SD, PROMIS PI 3.5 to 5.5, ODI 6.8 to 22.9, SF-12 PCS 2.5 to 12.6, SF-12 MCS 2.4 to 15.9.

Descriptive statistics were calculated for all scores at baseline to examine level of impairment. Floor and ceiling effects were evaluated by determining percentage of patients who had the highest and lowest possible scores for an instrument. Convergent validity was

assessed using Pearson correlation coefficients between PROMIS CATs, ZCQ, ODI, and SF-12 at baseline.

RESULTS

Of the 98 (63 female, 35 male) patients enrolled (mean age = 61.9, SD = 13.8), 82% completed baseline, 6 weeks and 3 months assessments.

At baseline, patients demonstrated impairments in physical function and pain on all measures including PROMIS PF (mean = 35.0, SD = 6.1), PROMIS PI (mean = 64.3, SD = 7.2), PROMIS PB (mean = 60.3, SD = 4.7), ODI (mean = 43.0, SD = 17.5), ZCQ total symptom severity (mean = 3.3, SD = 0.7), ZCQ PF (mean = 2.6, SD = 0.6), and SF-12 PCS (mean = 33.2, SD = 8.3). Convergent validity was supported with multiple statistically significant correlations in the expected direction at baseline between PROMIS CATs and legacy measures. Specifically, ODI scores correlated strongly with PROMIS PB, PI, and PF ($r = 0.60$, 0.73, and -0.58, respectively, all $P < 0.01$). ZCQ PF and SF-12 PCS correlated strongly with PROMIS PF ($r = -0.061$, $P < 0.01$; $r = 0.50$, $P < 0.01$, respectively).Additionally, ZCQ pain strongly correlated with PROMIS PI and PB at baseline ($r = 0.66$ and 0.59, $P < 0.01$).

Known groups validity was supported. Patients reporting ODI improvements at time 2 had expected decreases in PROMIS PI and PB (-12.98 and -9.74, respectively) and increased PROMIS PF scores (mean = 7.53; Table 1). PROMIS change scores reached statistical significance between improved and unchanged/worsened patients with the improved group reporting better outcomes (all $P < 0.001$).

While only 6% to 9% of patients exhibited baseline PROMIS scores within five points of the general population mean, by 3 months the number increased to approximately 33% to 40% of patients (Table 2). Physical function and pain improved following surgery outcome measures as expected. Observed change scores for PROMIS PB and PI demonstrated decreases of 6.66 and 9.62, respectively between baseline and 3-months ($P < 0.001$), while PROMIS PF increased 6.8 points over the same time period ($P < 0.001$) (Table 3). The other legacy measures demonstrated score changes consistent with the observed trend seen with PROMIS CATs (ODI = -19, SF-12 PCS = 8.57, MCS = 5.04, ZCQ pain = -1.31, ZCQ neuroischemic = -0.95, ZCQ total = -1.10; each $P < 0.001$; Table 3). The improvements seen with PROMIS, ODI, and SF-12 scores reach MCID thresholds.

PROMIS CATs demonstrated responsiveness to treatment between time 1 and 2 when comparing patients who reported improvement compared with all others with SRM of PROMIS PB, PI, and PF of -1.20, -1.22, and 0.80, respectively as shown in Table 4 ($P < 0.05$). PROMIS CATs also demonstrated responsiveness between time 2 and time 3 with SRM of PROMIS PB, PI, and PF of -0.19, -0.33, and 0.40 as shown in Table 5.

The three PROMIS instruments took an average of 2.6 minutes to complete together, with individual CAT completion times of 1.0 minutes for PB (SD = 0.8), 0.8 minutes for PI (SD = 0.6), and 0.8 minutes for PF (SD = 0.8). This compares favorably with the completion times

for the ODI (mean = 3.1 min, SD = 1.4), ZCQ (mean = 3.6 min, SD = 1.6), and SF-12 (mean = 3.0 min, SD = 1.3) and is reduced compared with the total time for legacy measures.

PROMIS CATs demonstrated minimal floor and ceiling effects (Table 1). This is relevant for LSS patients as a substantial number reported severe symptoms as determined by baseline ODI score (severe disability = 32.0%, crippled = 16.5%, and bed-bound = 2.1%). A reduced floor effect allows for more precise measurement of those with more impairment.

Patients' reported disability was commonly described as being unaffected by overlapping or concomitant pathology (ICC), with 70% and 59% of patients noting no concomitant painful pathology at baseline and 3-month follow up time points.

DISCUSSION

This study establishes convergent validity, known groups validity, and responsiveness, of the PROMIS PF, PI, and PB CATs in surgically treated LSS. These measures were brief and exhibited minimal floor and ceiling effects. To our knowledge, this is the first assessment of the validity of PROMIS CATs for physical function, pain interference, and pain behavior in surgically treated LSS.

PROMIS measures offer some advantages over legacy outcome instruments such as the ODI, ZCQ, and SF-12. First, PROMIS allows universal symptom assessment so scores can be compared across any other condition. Item banks enable flexibility in administration through use of CATs or fixed length short forms. PROMIS allows comparisons even if patients did not answer the same questions. Item banks can also be improved over time through the addition of new items further reducing floor and ceiling effects.

At baseline, there is severe disability and pain in LSS patients which is consistent with previously published reports.[1,5,6] Up to 40% of surveyed patients stated their answers were affected by concurrent comorbidities. This suggests that attributing a patient's state of health/disease or treatment to a singular disease entity can be misleading. Therefore, PROs such as PROMIS that evaluate overall perception of pain and function is more effective to understand overall disability than disease specific PROs.

Web-based data collection for PROMIS instruments allows for tracking completion times, time and date stamps on responses, immediate scoring, and automated tracking of missing data. Although CATs require a computer for administration, their advantage in speed and measurement precision facilitate making PROs available in real time during a clinical encounter. This information can be used by health-care providers to facilitate assessment of the patient, treatment evaluation, planning or modification. Patients can use PRO information for tracking their health and facilitating patient-provider communication.

This study has several limitations. The 3-month follow-up period was selected for assessing the validity of PROMIS CATs.[7] However, this time may not be sufficient for capturing clinically significant outcomes as such this study does not provide validation of the surgical procedures performed. Parker *et al*[21] suggested 12 months follow up, as they found 3 months ODI MCID for lumbar surgery predicted 12 months MCID thresholds with only 62.6%

specificity and 86.8% sensitivity. Longer follow-up of 1 to 2 years would be needed to investigate the sustained effectiveness of surgical treatment. In the absence of defined MCIDs for LSS, we reviewed available known MCIDs for comparable thoracolumbar spine pathologies. While there are few publications for MCIDs for PROMIS PB, PI, and PB, an acceptable but controversial estimate is 50% of the reported standard deviation (SD).[22] Amtmann *et al*[13] recently reported that a MCID of 3.5 to 5.5 points in PROMIS PI scores may be useful in low back pain patients. Some thoracolumbar spine literature reports a range of 6.8 to 14.9 point decrease in ODI as a MCID and SF-12 PCS and MCS improvement of 2.5 to 6.1 and 10.1, respectively, as a MCID.[23–27] Parker *et al* studied MCIDs for decompression following same level recurrent lumbar stenosis, reporting MCID ranges for ODI (8.2–19.9), SF-12 MCS (7.0–15.9), and SF-12 PCS (2.5–12.1). Previously reported MCIDs for extension of lumbar fusion for adjacent segment disease for various outcome measures included ODI (6.8–16.9), SF-12 PCS (6.1–12.6), and SF-12 MCS (2.4–10.8). ODI MCID for transforaminal lumbar interbody fusion for degenerative lumbar spondy-lolithesis are reported to range from 11 to 22.9. Due to variability in deriving and reporting MCID thresholds, physicians should interpret reaching MCID thresholds in isolation with caution.[28]

References

1. Kalichman L, Cole R, Kim DH, et al. Spinal stenosis prevalence and association with symptoms: the Framingham Study. Spine J 2009;9:545–50. [PubMed: 19398386]

2. Roberson GH, Llewellyn HJ, Taveras JM. The narrow lumbar spinal canal syndrome. Radiology 1973;107:89–97. [PubMed: 4689450]

3. De Villiers PD, Booysen EL. Fibrous spinal stenosis. A report on 850 myelograms with a water-soluble contrast medium. Clin Orthop Relat Res 1976;140–4. [PubMed: 1253477]

4. Ishimoto Y, Yoshimura N, Muraki S, et al. Prevalence of symptomatic lumbar spinal stenosis and its association with physical performance in a population-based cohort in Japan: the Wakayama Spine Study. Osteoarthritis Cartilage 2012;20: 1103–8. [PubMed: 22796511]

5. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonoperative treatment for lumbar spinal stenosis four-year results of the Spine Patient Outcomes Research Trial. Spine (Phila Pa 1976) 2010;35:1329–38. [PubMed: 20453723]

6. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. N Engl J Med 2008;358:794–810. [PubMed: 18287602]

7. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. J Eval Clin Pract 2006;12:559–68. [PubMed: 16987118]

8. Hung M, Hon SD, Franklin JD, et al. Psychometric properties of the PROMIS physical function item bank in patients with spinal disorders. Spine (Phila Pa 1976) 2014;39:158–63. [PubMed: 24173018]

9. Cella D, Yount S, Rothrock N, et al. The patient-reported outcomes measurement information system (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007;45 (5 Suppl 1):S3–11.

10. Jensen RE, Potosky AL, Reeve BB, et al. Validation of the PROMIS physical function measures in a diverse US population-based cohort of cancer patients. Qual Life Res 2015;24:2333–44. [PubMed: 25935353]

11. Hung M, Baumhauer JF, Latt LD, et al. Validation of PROMIS (R) physical function computerized adaptive tests for orthopaedic foot and ankle outcome research. Clin Orthop Relat Res 2013;471: 3466–74. [PubMed: 23749433]

12. Flynn KE, Dew MA, Lin L, et al. Reliability and construct validity of PROMIS(R) measures for patients with heart failure who undergo heart transplant. Qual Life Res 2015;24: 2591–9. [PubMed: 26038213]

13. Amtmann D, Kim J, Chung H, et al. Comparing CESD-10, PHQ-9, and PROMIS depression instruments in individuals with multiple sclerosis. Rehabil Psychol 2014;59:220–9. [PubMed: 24661030]

14. Irwin DE, Atwood CA Jr, Hays RD, et al. Correlation of PROMIS scales and clinical measures among chronic obstructive pulmonary disease patients with and without exacerbations. Qual Life Res 2015;24:999–1009. [PubMed: 25307510]

15. Choi SW. Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. Appl Psychol Measur 2009;33:644–5.

16. Fitzpatrick R, Davey C, Buxton MJ, et al. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2:i–v; 1–74.

17. Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. Clin Exp Rheumatol 2005;23 (5 Suppl 39):S53–7.

18. Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. Qual Life Res 1997;6:595–600. [PubMed: 9330558]

19. Weiss DJ. Computerized adaptive testing for effective and efficient measurement in counseling and education. Measure Eval Counsel Dev 2004;37:70–84.

20. Godil SS, Parker SL, Zuckerman SL, et al. Determining the quality and effectiveness of surgical spine care: patient satisfaction is not a valid proxy. Spine J 2013;13:1006–12. [PubMed: 23685216]

21. Parker SL,Asher AL,Godil SS,et al.Patient-reportedoutcomes3months after spine surgery: is it an accurate predictor of 12-month outcome in real-world registry platforms? Neurosurg Focus 2015;39:E17.

22. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407–15. [PubMed: 2691207]

23. Parker SL, Mendenhall SK, Shau DN, et al. Minimum clinically important difference in pain, disability, and quality of life after neural decompression and fusion for same-level recurrent lumbar stenosis: understanding clinical versus statistical significance. J Neurosurg Spine 2012;16:471–8. [PubMed: 22324801]

24. Parker SL, Mendenhall SK, Shau D, et al. Determination of minimum clinically important difference in pain, disability, and quality of life after extension of fusion for adjacent-segment disease. J Neurosurg Spine 2012;16:61–7. [PubMed: 21962034]

25. Parker SL, Adogwa O, Paul AR, et al. Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. J Neurosurg Spine 2011;14:598–604. [PubMed: 21332281]

26. Parker SL, Adogwa O, Mendenhall SK, et al. Determination of minimum clinically important difference (MCID) in pain, disability, and quality of life after revision fusion for symptomatic pseudoarthrosis. Spine J 2012;12:1122–8. [PubMed: 23158968]

27. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003;41:582–92. [PubMed: 12719681]

28. Copay AG, Martin MM, Subach BR, et al. Assessment of spine surgery outcomes: inconsistency of change amongst outcome measurements. Spine J 2010;10:291–6. [PubMed: 20171937]

Key Points

- PROMIS is an adaptive, responsive assessment tool that measures patient-reported health status that is funded by the NIH.

- PROMIS CATs offer precision and validity while requiring a smaller, targeted subset of questions administered from a large collection (*i.e.*, item banks), thereby significantly reducing the time needed to complete a measure.

- PROMIS CATs demonstrate convergent validity, known groups' validity, and responsiveness for surgically treated patients with symptomatic lumbar spinal stenosis.
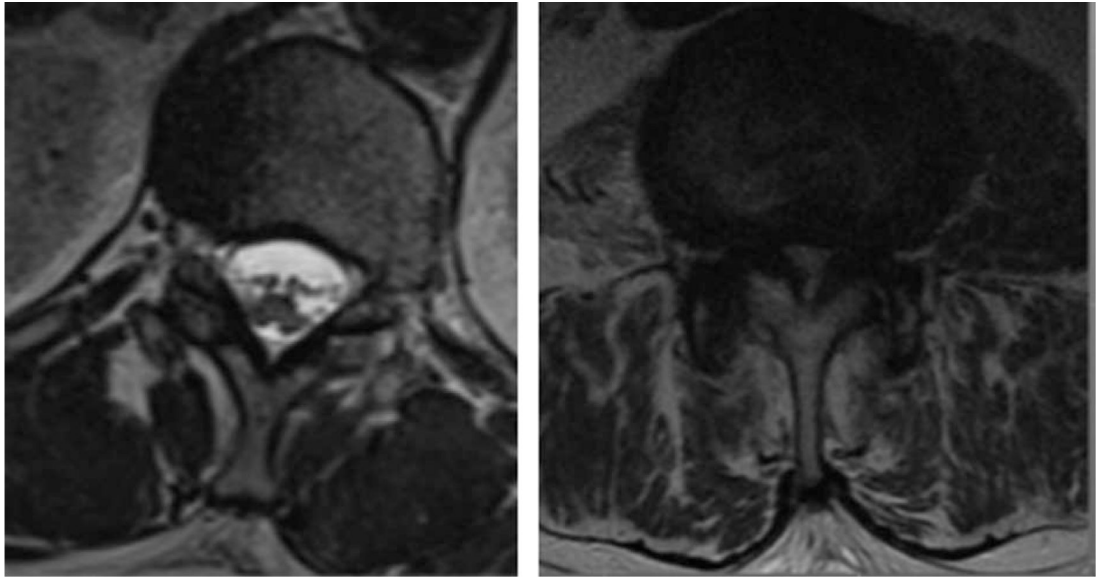
Figure 1.
Normal spinal canal cross-section on left image compared with stenotic spinal canal cross-section where the space for the white cerebrospinal fluid bathing the neural elements is significantly decreased.

Author Manuscript   Author Manuscript   Author Manuscript   Author Manuscript

TABLE 1.

Floor and Ceiling Effects

Label	Baseline			3-Month		
	N	% at Floor	% at Ceiling	N	% at Floor	% at Ceiling
PROMIS						
Pain behavior	97	1	0	87	14	0
Pain interference	97	1	0	87	17	0
Physical function	97	0	0	87	0	0
Oswestry disability index (ODI)	97	0	0	86	0	0
Zurich						
Symptom severity—pain (1–5)	98	0	5	87	9	0
Symptom severity—neuroischemic (1–5)	97	4	0	87	21	0
Symptom severity—total (1–5)	98	0	0	87	5	0
Physical function (1–4)	98	0	0	87	16	0
Satisfaction (1–4)	28	0	25	88	26	0
SF-12						
Physical component score	97	0	0	87	0	0
Mental component score	96	0	0	87	0	0

Note that Assessment center only measures instrument time to the nearest minute. Some extreme timing outliers were excluded from descriptive stats.

PROMIS indicates patient reported outcomes measurement information system; SF-12, Short-Form 12.

TABLE 2.

Percentage of Patients Within Five Points (MCID) of the Population Mean (50)

	Baseline	Time 2	Time 3	Time 4
PROMIS pain behavior				
Within five points of general population	8 (8%)	31 (37%)	35 (40%)	18 (45%)
Worse than general population	89 (92%)	53 (63%)	52 (60%)	22 (55%)
PROMIS pain interference				
Within five points of general population	9 (9%)	28 (33%)	37 (43%)	17 (42%)
Worse than general population	88 (91%)	57 (67%)	50 (57%)	23 (58%)
PROMIS physical function				
Within five points of general population	6 (6%)	21 (25%)	29 (33%)	13 (32%)
Worse than general population	91 (94%)	63 (75%)	59 (67%)	27 (68%)

MCID indicates minimal clinic important difference; PROMIS, patient reported outcomes measurement information system.

TABLE 3.

Change in Scores Between Visits

Assessment	Change in…	N	Mean	SD	Effect Size	Range	p-value[*]
2 vs 1	**PROMIS**						
	Pain Behavior T-Score	83	−5.94	7.42		−28, 6.2	<0.001
	Pain Interference T-Score	84	−7.95	9.45		−33.1, 19.2	<0.001
	Physical Function T-Score	83	3.50	8.23		−17.4, 27.8	<0.001
	Oswestry						
	Oswestry Disability Index (ODI)	79	−11.43	17.42		−53.6, 24	<0.001
	Zurich						
	Symptom Severity - Pain (1–5)	83	−1.09	0.99		−3.7,1.0	<0.001
	Symptom Severity - Neuroischemic (1–5)	83	−0.89	0.92		−3.5, 1.2	<0.001
	Symptom Severity - Total (1–5)	84	−0.98	0.82		−3.6, 0.7	<0.001
	Physical Function (1–4)	84	−0.78	0.76		−2.6, 1.2	<0.001
	SF-12						
	Physical Component Score	83	5.04	10.72		−24.8, 34.5	<0.001
	Mental Component Score	82	3.16	11.14		−42.2, 28.1	0.012
3 vs 2	**PROMIS**						
	Pain Behavior T-Score	78	−0.55	9.02		−27.2, 21.9	0.594
	Pain Interference T-Score	79	−1.00	9.97		−21.8, 29.4	0.377
	Physical Function T-Score	79	3.03	6.40		−16.3, 19.8	<0.001
	Oswestry						
	Oswestry Disability Index (ODI)	75	−7.12	14.96		−73.1, 18	<0.001
	Zurich						
	Symptom Severity - Pain (1–5)	77	−0.21	0.73		−2.3, 1.3	0.015
	Symptom Severity - Neuroischemic (1–5)	77	−0.06	0.62		−1.8, 2.0	0.358
	Symptom Severity - Total (1–5)	78	−0.12	0.47		−1.4, 0.9	0.022
	Physical Function (1–4)	78	−0.19	0.52		−1.4, 1.2	0.002
	SF-12						
	Physical Component Score	78	3.22	9.39		−27.5, 25.1	0.003
	Mental Component Score	78	2.21	10.33		−16.6, 40.6	0.062
3 vs 1	**PROMIS**						
	Pain Behavior T-Score	86	−6.66	9.49	−0.70	−34.2, 23.9	<0.001
	Pain Interference T-Score	86	−9.62	10.95	−0.88	−36.1, 15.2	<0.001
	Physical Function T-Score	87	6.80	7.06	0.96	−5.8, 24.6	<0.001
	Oswestry						
	Oswestry Disability Index (ODI)	85	−19.00	19.87	−0.96	−70, 35.6	<0.001
	Zurich						
	Symptom Severity - Pain (1–5)	87	−1.31	1.03	−1.27	−3.67, 1	<0.001

Assessment	Change in…	N	Mean	SD	Effect Size	Range	p-value[*]
	Symptom Severity - Neuroischemic (1–5)	86	−0.95	0.99	−0.96	−3.5, 1.5	<0.001
	Symptom Severity - Total (1–5)	87	−1.10	0.87	−1.26	−3.6, 0.6	<0.001
	Physical Function (1–4)	87	−0.96	0.70	−1.37	−2.6, 0.6	<0.001
	SF-12						
	Physical Component Score	86	8.57	10.84	0.79	−20.7, 36.4	<0.001
	Mental Component Score	85	5.04	9.08	0.56	−13.1, 29.1	<0.001

[*] P-value for t test of null hypotheses that mean change = 0.

TABLE 4.

Change in Scores Between Visits 1 and 2, by Patient-Rated Change Category

Change From Visit 1 to 2 in...	Much Better					Slightly Better—Much Worse					
	n	Visit 1 Mean (SD)	Visit 2 Mean (SD)	Mean Change (SD)	SRM	n	Visit 1 Mean (SD)	Visit 2 Mean (SD)	Mean Change (SD)	SRM	P-Value*
PROMIS											
Pain behavior	28	61.6 (3.9)	53.1 (7.9)	−8.5 (7.1)	−1.20	19	59.7 (3.7)	59.4 (4.6)	−0.3 (4.0)	−0.08	<0.001
Pain interference	28	65.2 (7.5)	54.9 (8.2)	−10.3 (8.4)	−1.22	19	63.1 (6.8)	61.0 (7.8)	−2.1 (5.5)	−0.38	<0.001
Physical function	28	34.7 (6.2)	41.0 (8.3)	6.4 (8.0)	0.80	19	34.8 (5.5)	35.0 (7.0)	0.2 (7.0)	0.03	0.009
SF-12											
PCS	28	33.4 (8.9)	41.4 (10.1)	8.0 (11.7)	0.69	19	36.1 (6.8)	35.2 (8.5)	−0.9 (8.1)	−0.11	0.006
MCS	27	47.5 (11.2)	49.8 (13.4)	2.3 (13.4)	0.17	19	45.6 (10.4)	46.9 (12.6)	1.3 (10.2)	0.13	0.778
ODI	26	44.3 (19.0)	26.3 (19.1)	−18.0 (16.5)	−1.09	18	39.8 (17.4)	38.3 (13.6)	−1.5 (13.9)	−0.11	0.001
Zurich											
Pain	27	3.80 (0.75)	2.31 (0.84)	−1.49 (0.95)	−1.57	20	3.65 (0.75)	3.28 (0.71)	−0.37 (0.88)	−0.42	<0.001
Neuroischemic	28	3.12 (0.99)	1.90 (0.80)	−1.21 (0.89)	−1.36	20	2.70 (1.00)	2.28 (0.93)	−0.43 (0.59)	−0.71	0.001
Symptom severity	28	3.40 (0.79)	2.07 (0.63)	−1.33 (0.74)	−1.80	20	3.11 (0.82)	2.71 (0.75)	−0.40 (0.55)	−0.73	<0.001
Physical function	28	2.72 (0.59)	1.76 (0.65)	−0.96 (0.71)	−1.35	20	2.50 (0.55)	2.23 (0.59)	−0.27 (0.62)	−0.44	0.001

SRM, standardized response mean = mean change/SD.

*
P-value is comparing mean change scores between the two groups.

MCS indicates mental component score; ODI, Oswestry disability index; PCS, physical component score; PROMIS, patient reported outcomes measurement information system; SD, standard deviation; SF-12, Short-Form 12.

TABLE 5.

Change in Scores Between Visits 2 and 3, by Patient-Rated Change Category

Change From Visit 2 to 3 in…	Much Better					Slightly Better—Much Worse					P-Value*
	n	Visit 2 Mean (SD)	Visit 3 Mean (SD)	Mean Change (SD)	SRM	n	Visit 2 Mean (SD)	Visit 3 Mean (SD)	Mean Change (SD)	SRM	
PROMIS											
Pain behavior	23	53.1 (7.0)	51.2 (8.9)	−1.9 (10.3)	−0.19	27	57.0 (6.8)	57.8 (5.6)	0.8 (7.9)	0.10	0.288
Pain interference	23	54.8 (8.4)	51.4 (7.7)	−3.4 (10.2)	−0.33	27	59.8 (7.4)	59.4 (8.0)	−0.5 (8.1)	−0.06	0.268
Physical function	23	40.8 (9.0)	44.3 (4.9)	3.5 (8.8)	0.40	27	35.6 (7.1)	38.0 (4.2)	2.4 (5.7)	0.43	0.604
SF-12											
PCS	23	41.1 (9.2)	46.2 (6.5)	5.0 (7.8)	0.65	27	35.7 (10.3)	36.9 (6.8)	1.1 (10.6)	0.11	0.149
MCS	23	48.5 (12.8)	54.4 (9.9)	5.8 (10.8)	0.54	27	48.4 (12.6)	48.8 (10.1)	0.4 (11.2)	0.04	0.091
ODI	22	26.0 (15.6)	15.8 (10.2)	−10.2 (13.4)	−0.76	26	40.1 (17.9)	32.2 (12.9)	−7.9 (18.2)	−0.44	0.629
Zurich											
Pain	23	2.23 (0.81)	2.06 (0.55)	−0.22 (0.80)	−0.28	26	3.22 (0.73)	3.03 (0.71)	−0.19 (0.69)	−0.28	0.907
Neuroischemic	23	1.91 (0.69)	1.71 (0.83)	−0.21 (0.65)	−0.32	27	2.31 (0.89)	2.41 (0.80)	0.09 (0.69)	0.13	0.124
Symptom severity	23	2.07 (0.46)	1.86 (0.51)	−0.21 (0.48)	−0.44	27	2.68 (0.72)	2.66 (0.66)	−0.02 (0.46)	−0.04	0.155
Physical function	23	1.76 (0.56)	1.49 (0.43)	−0.27 (0.52)	−0.52	27	2.21 (0.68)	1.99 (0.59)	−0.22 (0.60)	−0.37	0.770

SRM, standardized response mean = mean change/SD.

*
P-value is comparing mean change scores between the two groups.

MCS indicates mental component score; ODI, Oswestry disability index; PCS, physical component score; PROMIS, patient reported outcomes measurement information system; SD, standard deviation; SF-12, Short-Form 12.