

Rational Selection of Patient-Reported Outcomes Measures in Lumbar Spine Surgery Patients

SUSANNE H. BODEN, MD,¹ KEVIN X. FARLEY, BS,² COLLIER CAMPBELL, BS,² SCOTT D. BODEN, MD,² MICHAEL B. GOTTSCHALK, MD²

¹Oakland University William Beaumont School of Medicine, Rochester, Minnesota, ²Emory University School of Medicine, Department of Orthopaedics, Atlanta, Georgia

ABSTRACT

Background: The Short Form-12 (SF-12) was developed as a shorter version of the SF-36, yet there has been limited validation of its reliability at measuring postoperative changes. The purpose of this study was to determine if the SF-12 could safely substitute for the SF-36 in measuring postoperative change in lumbar spine surgery patients and if the condition specific (Oswestry Disability Index [ODI]) or pain (visual analog scale [VAS]) instruments, provided additional utility.

Methods: A total of 972 patients from a single center who underwent lumbar spine surgery for a predominant symptom of radiating leg pain with ($n = 237$) or without ($n = 735$) fusion and prospectively completed both SF-36 and ODI instruments before and after surgery were included. The SF-12 score was calculated from the appropriate subset of SF-36 responses. The absolute sensitivity and the intraclass correlation coefficient were calculated. Reliability of each instrument to measure preoperative to postoperative change was calculated as the standardized response mean.

Results: The SF-12 and SF-36 demonstrated a strong correlation with each other ($[0.97, P < .001]$ and $[0.93, P < .001]$, respectively) preoperatively and postoperatively. The SF-12 and SF-36 scores were moderately to strongly inversely correlated with the ODI. The ODI showed greater reliability at measuring change than the SF-12 for both fusion (0.94 versus 0.72) and nonfusion (0.81 versus 0.33) lumbar surgery patients.

Conclusions: The SF-12 was as effective as the SF-36 to measure general health status in lumbar spine surgery patients, and both were moderate to strong predictors of ODI preoperatively and postoperatively, but lack the reliability to detect change seen with the ODI or VAS after surgical intervention.

Level of Evidence: 3.

Clinical Relevance: These data suggest that the SF-12 is a valid substitute for the SF-36 to measure postoperative outcomes changes, but that the ODI should continue to be used to measure condition specific changes in function.

Lumbar Spine

Keywords: lumbar spine surgery, patient reported outcomes, Oswestry Disability Index, SF-12, SF-36, visual analog scale, quality of life, health status

INTRODUCTION

As health care economics and quality initiatives place increasing emphasis on patient outcomes, it is important to assess the reliability of these outcome measures in specific patient populations. One of the most widely used tools for evaluating general health status is the Short Form (SF-36), which assesses both mental and physical quality of life. While an abbreviated version, the SF-12, has been validated in nonspinal conditions,^{1–3} there is only 1 small study comparing it with the SF-36 specifically in patients with lumbosacral spinal disorders.⁴ Other commonly used instruments include the Oswestry

Disability Index (ODI), a condition-specific measure that assesses patient disability associated with low back pain, and the visual analog scale (VAS), a measure of pain intensity. Conflicting evidence exists as to whether these instruments correlate with one another in specific spinal pathologies.^{5–7} Clinically, the goal is to measure patient outcomes in the least burdensome way. Many spine centers have historically collected SF-36 for many years, and the question of whether it is “safe” to switch to the less burdensome SF-12 for long-term follow up has not been answered.

Only 1 study has compared the SF-36 with the SF-12 specifically in lumbar spine surgery patients,

and it demonstrated that the SF-36 physical (PCS) and mental (MCS) component scores strongly correlated with their respective component scores in the SF-12.⁴ However, the study of Lee et al⁴ was limited by sample size, including only 74 patients, and did not assess reliability at detecting changes from preoperative to postoperative states. While Ko and Chae determined a moderate correlation between the SF-36 and the ODI, this study only included 69 patients with lumbar spinal stenosis and did not assess the ability to measure postoperative surgical changes.⁷

Although there are multiple comparisons of SF-36 and SF-12 in a variety of clinical disease categories, no large sample studies have validated the SF-12 as a substitute outcome measure for SF-36 specifically in patients undergoing lumbar spinal surgery. Additionally, limited research compares the SF-12, ODI, and VAS in specific lumbosacral diagnostic groups or their reliability at measuring change from preoperative to postoperative states. The purpose of this study was (1) to determine if SF-12 scores correlate with SF-36, ODI, and VAS scores in patients undergoing surgery for specific lumbosacral pathologies, and (2) to compare the reliability of these tools to measure change between the preoperative and postoperative states. This information will help determine if patients that have SF-36 baselines can be followed up with SF-12 instead and whether the other instruments provide additional information relevant to patient outcomes after surgery.

MATERIALS AND METHODS

Our patient-reported outcomes registry identified 972 patients who underwent primary lumbar decompression spine with or without fusion surgery for a symptom of predominantly (>50%) radiating leg pain due to spinal stenosis with or without degenerative spondylolisthesis from 2010 to 2017 who had completed both the SF-36 and the ODI. Patients with tumors or infections were not included. SF-12v2 PCS and MCS scores were calculated from the subset of SF-36 responses, a previously validated approach which gives comparable results to those obtained if the surveys are administered separately.⁸ The visual analog scale (VAS) for pain was converted to a 0–100 scale. Surveys were completed on wireless tablets at the preoperative and postoperative clinic appointments in the waiting room with timestamps.

Surveys were compared preoperatively, 3 months postoperatively, and 6 months postoperatively. Descriptive statistics were performed to assess demographic and other details of the patient cohort. Analysis was performed based on diagnosis (spinal stenosis with or without degenerative spondylolisthesis) or surgery type (lumbar laminectomy or discectomy with or without fusion). T-tests or Wilcoxon rank-sum tests were used to compare parametric and nonparametric continuous data. All analyses were performed on SPSS (Version 25.0, IBM Corp., Armonk, NY).

Reliability of each instrument to detect surgical change (effect size) was measured as the standardized response mean (SRM), which was calculated as the mean change from the preoperative to postoperative scores divided by the standard deviation of the mean change (SRM = mean[postoperative – preoperative]/standard deviation). Higher values of this metric indicate greater reliability of the survey to detect a postoperative change in health status or function. As described by Cohen, an effect size of <0.20 is trivial, 0.20–0.50 is small, 0.50–0.80 is moderate, and >0.80 is large.^{8–12} The absolute sensitivity, measured as the quartile-based coefficient of variation, was calculated to determine the sensitivity of the scale to detect differing levels of disease severity, with greater scores indicating a greater dispersion to the scale across the population. The quartile-based coefficient of variation was calculated by dividing the interquartile range of survey response scores over the median of the survey response scores at a single time point.⁸ The intra-class correlation coefficient (ICC) was used to determine the degree of agreement between each survey, with a greater ICC indicating a greater level of agreement.¹ The ICC score ranges from 0 to 1, with 0–0.10 representing no agreement, 0.11–0.40 representing slight agreement, 0.41–0.60 representing fair agreement, 0.61–0.80 representing moderate agreement, and 0.81–1.00 representing substantial agreement.^{1,13} For the ICC, all diagnoses and surgery types were grouped together, and the ICC was assessed at multiple time points.¹³ Floor and ceiling effects were also evaluated by determining the proportion of participants who achieved the lowest or highest possible scores, respectively. Floor and ceiling effects were considered present if more than 15% of individuals scored the lowest or highest possible total score on any of the outcomes measurement instruments.

Table 1. Correlation of Short Form (SF) physical (PCS) and mental (MCS) component scores with Oswestry Disability Index (ODI) at multiple perioperative time points for patients undergoing lumbar surgery.

	SF-36 (PCS) versus ODI		SF-12 (PCS) versus ODI		SF-36 (MCS) versus ODI		SF-12 (MCS) versus ODI	
	ICC (95% CI)	P Value						
Preoperative	-0.429 (-0.482, -0.373)	<.001	-0.425 (-0.478, -0.369)	<.001	-0.433 (-0.486, 0.378)	<.001	-0.406 (-0.460, 0.349)	<.001
3 months	-0.532 (-0.594, -0.463)	<.001	-0.481 (-0.548, 0.408)	<.001	-0.517 (-0.580, -0.447)	<.001	-0.502 (-0.567, 0.430)	<.001
6 months	-0.637 (-0.718, -0.538)	<.001	-0.628 (-0.712, 0.528)	<.001	-0.698 (-0.768, -0.612)	<.001	-0.646 (-0.726, -0.549)	<.001

Abbreviations: CI, confidence interval; ICC, interclass correlation coefficient.

RESULTS

This study included a total of 972 patients (42.9% female) with an average age of 62.6 ± 12.8 years. A total of 76.3% (742) of patients self-identified as white, 14.9% (145) self-identified as black or African-American, and 8.7% (85) self-identified as another race or ethnicity. From the initial cohort with preoperative surveys completed, 238 patients (24.5%) underwent lumbar, fusion and 734 (75.5%) underwent lumbar discectomy or laminectomy without fusion. A total of 324 (33.3%) patients had a diagnosis of spinal stenosis with degenerative spondylolisthesis, and 648 (66.7%) had spinal stenosis without degenerative spondylolisthesis. The mean time from preoperative survey completion to surgery was 15.7 ± 9.8 days, indicating that the baseline surveys were obtained shortly before the surgical intervention.

When analyzing changes in the various health-related quality of life (HRQoL) instruments after surgery, several substantial correlations were identified. Both MCS and PCS components of the SF-36 and SF-12 scores were strongly correlated preoperatively (0.967 [*P* < .001] and 0.934 [*P* < .001], respectively) and remained significant for both postoperative time points. Correlation of SF-36 and SF-12 scores to ODI scores prior to surgery were less strong than with each other. The PCS components of the SF-36 and SF-12 showed preoperative correlations of -0.429 (*P* < .001) and -0.425 (*P* < .001), respectively, with preoperative ODI scores, and the strength of these correlations

increased postoperatively (Table 1). Similarly, the preoperative MCS components of the SF-36 and SF-12 scores showed moderate inverse correlations with ODI of -0.433 (*P* < .001) and -0.406 (*P* < .001), respectively. These correlations also increased across both postoperative time points. When comparing SF-36, SF-12, and ODI with VAS scores, only a weak correlation appeared preoperatively (correlation coefficients: -0.098-0.121, all *P* < .05). No significant correlation appeared between VAS and the other HRQoL measures at any postoperative time point (all *P* > .05; Table 2).

Absolute sensitivity, ranging from 0 to 1.0, or poor to excellent, of each survey was determined to assess the relative ability to detect differences in levels of disease severity. All the instruments showed similar results (ranging from 0.36 to 0.50). VAS showed the best absolute sensitivity for patients undergoing lumbar laminectomy or discectomy without fusion (0.50), while for patients undergoing a fusion procedure, ODI and SF-36 (PCS) were slightly better (0.46; Table 3).

The reliability of each HRQoL instrument to detect change after surgery was compared in patients undergoing lumbar surgery with and without fusion. In both cohorts, the VAS pain score and ODI showed greater reliability at detecting postoperative change than the SF-12 and SF-36 at all time points, except at 3 months in the lumbar fusion cohort, when ODI and SF-12 (PCS) were equivalent. VAS showed superiority detecting change at all times, except the 6 month survey for

Table 2. Correlation of Short Form-12 (SF-12) and Oswestry Disability Index (ODI) with visual analog scale (VAS)-pain at multiple perioperative time points for patients undergoing lumbar surgery.

	SF-12 (PCS) versus VAS		SF-12 (MCS) versus VAS		ODI versus VAS	
	ICC (95% CI)	P Value	ICC (95% CI)	P Value	ICC (95% CI)	P Value
Preoperative	-0.098 (-0.174, -0.021)	.006	-0.086 (-0.163, -0.009)	.015	0.121 (0.044, 0.197)	.001
3 months	-0.171 (-0.354, 0.024)	.053	-0.109 (-0.298, 0.087)	.137	0.159 (-0.036, 0.343)	.055
6 months	-0.211 (-0.452, 0.057)	.061	-0.172 (-0.418, 0.098)	.105	0.152 (-0.119, 0.401)	.134

Abbreviations: CI, confidence interval; ICC, interclass correlation coefficient; MCS, mental component score; PCS, physical component score.

Table 3. Absolute sensitivity of preoperative Short Form-12 (SF-12), Short Form-36 (SF-36), Oswestry Disability Index (ODI), and visual analog scale (VAS)-pain.^a

	SF-12 (PCS)	SF-36 (PCS)	SF-12 (MCS)	SF-36 (MCS)	ODI	VAS-Pain
Lumbar fusion (n = 232)	0.43	0.46	0.4	0.44	0.46	0.43
Lumbar laminectomy or discectomy without fusion (n = 716)	0.37	0.39	0.36	0.36	0.46	0.50

Abbreviations: MCS, mental component score; PCS, physical component score.

^aMeasured as the coefficient of variation: interquartile range/median. Higher values indicate greater sensitivity of the scale to distinguish between different levels of disease severity.

the nonfusion patients, in which ODI was slightly superior (Table 4).

Reliability at measuring postoperative change for each of these HRQoL measures was also assessed by diagnosis type: spinal stenosis with or without degenerative spondylolisthesis. In both groups, the VAS pain score and ODI demonstrated significantly higher reliability at measuring postoperative change than the MCS and PCS components of SF-36 and SF-12 at both postoperative time points. However, no advantage of VAS or ODI existed when sensitivity to change was compared between the 2 postoperative time points (Table 5).

Floor and ceiling effects were assessed on the various outcomes instruments studied. No significant floor or ceiling affects were seen for any of the

outcomes measured (Table 6). VAS had the highest amount of floor or ceiling affects with 2.5% of patients reporting the lowest possible score and 8.3% reporting the highest possible score at the preoperative appointment. Likewise, no ceiling affects were present for VAS at 6 months, but 11.5% of patients reported the lowest possible score.

Finally, we analyzed the time to complete each survey collected from our institutional registry, including an additional cohort that collected the SF-12 directly instead of the SF-36. The average time spent completing the SF-12 was significantly less than that of the SF-36, 3.1 minutes versus 7.4 minutes ($P < .001$). Additionally, patients spent a statistically significant, but likely clinically irrelevant, decreased time completing the SF-12 as

Table 4. Comparison of the properties of the physical (PCS) and mental (MCS) component scores of the Short Form-36 (SF-36), Short Form-12 (SF-12), and the Oswestry Disability Index (ODI) scales.

	Preoperative, Mean (SD)	Postoperative, Mean (SD)	Change, Mean (SD)	Sensitivity to Change
1-3 months				
Lumbar fusion (n = 96)				
SF-12 (PCS)	29.19 (9.62)	29.45 (9.02)	0.26 (10.33)	0.03
SF-36 (PCS)	28.79 (9.04)	30.96 (9.14)	2.17 (9.31)	0.23
SF-12 (MCS)	48.54 (13.02)	46.82 (12.72)	-1.72 (13.55)	0.10
SF-36 (MCS)	47.57 (13.48)	45.44 (13.28)	-2.13 (13.64)	0.16
ODI	45.64 (15.70)	41.66 (18.53)	4.02 (17.02)	0.23
VAS-pain (n = 39)	6.72 (1.99)	3.95 (2.14)	2.7 (2.08)	1.32
Lumbar laminectomy or discectomy (without fusion; n = 394)				
SF-12 (PCS)	31.13 (8.34)	33.37 (9.77)	2.26 (10.65)	0.21
SF-36 (PCS)	31.11 (8.49)	34.67 (9.36)	3.54 (10.57)	0.33
SF-12 (MCS)	48.64 (11.44)	48.38 (12.30)	-0.26 (13.48)	0.02
SF-36 (MCS)	47.41 (12.12)	46.87 (13.02)	-0.54 (13.76)	0.04
ODI	46.41 (14.68)	34.17 (18.97)	12.23 (18.76)	0.65
VAS-pain (n = 43)	5.67 (2.50)	3.58 (2.55)	2.09 (3.32)	0.63
3-6 months				
Lumbar fusion (n = 45)				
SF-12 (PCS)	27.68 (8.00)	35.06 (10.80)	7.37 (10.22)	0.72
SF-36 (PCS)	27.93 (7.16)	37.11 (10.61)	9.18 (10.43)	0.88
SF-12 (MCS)	46.47 (12.54)	51.95 (11.17)	5.48 (11.28)	0.48
SF-36 (MCS)	45.05 (13.35)	50.09 (11.97)	5.04 (11.67)	0.43
ODI	50.33 (14.33)	32.42 (19.03)	17.91 (19.11)	0.94
VAS-pain (n = 23)	6.96 (2.01)	4.13 (2.36)	2.83 (2.54)	1.15
Lumbar laminectomy or discectomy (without fusion; n = 117)				
SF-12 (PCS)	31.13 (8.20)	34.81 (10.43)	3.67 (10.93)	0.33
SF-36 (PCS)	30.83 (7.74)	36.04 (9.81)	5.21 (10.57)	0.49
SF-12 (MCS)	47.10 (12.32)	48.15 (12.58)	1.04 (13.89)	0.07
SF-36 (MCS)	46.11 (13.29)	47.19 (13.63)	1.07 (14.04)	0.07
ODI	48.69 (14.34)	33.14 (18.46)	-15.37 (19)	0.81
VAS-pain (n = 24)	6.50 (2.04)	4.08 (2.32)	2.42 (2.79)	0.86

Abbreviation: SD, standard deviation.

Table 5. Sensitivity to change^a by diagnosis type.

	<3 Months			3-6 Months			Between Postoperative Surveys		
	Δ Change	SD	Sensitivity to Change	Δ Change	SD	Sensitivity to Change	Δ Change	SD	Sensitivity to Change
Spinal stenosis without degenerative spondylolisthesis									
SF-12 (PCS)	2.82	10.78	0.26	4.10	10.37	0.40	3.06	9.93	0.31
SF-36 (PCS)	4.07	10.56	0.39	5.31	9.63	0.55	3.40	8.43	0.40
SF-12 (MCS)	-0.33	13.49	-0.02	2.32	12.78	0.18	4.29	10.89	0.39
SF-36 (MCS)	-0.48	13.83	-0.03	2.79	13.39	0.21	4.66	11.83	0.39
ODI	11.71	18.70	0.63	15.64	18.43	0.85	7.52	15.21	0.49
VAS-pain	1.96	2.84	0.69	2.27	2.68	0.85	-0.05	1.80	0.03
Spinal stenosis with degenerative spondylolisthesis									
SF-12 (PCS)	0.07	9.97	0.01	4.91	11.03	0.45	5.65	9.85	0.57
SF-36 (PCS)	1.75	9.75	0.18	6.83	11.35	0.60	5.58	8.02	0.62
SF-12 (MCS)	-1.02	13.35	-0.08	2.80	14.48	0.19	3.52	11.67	0.30
SF-36 (MCS)	-1.58	13.38	-0.12	2.44	14.50	0.17	3.61	12.54	0.29
ODI	8.52	18.50	0.46	16.93	19.41	0.87	8.59	14.32	0.60
VAS-pain	4.00	2.00	2.00	3.43	2.47	1.39	0.80	1.48	0.54

Abbreviation: MCS, mental component score; ODI, Oswestry Disability Index; PCS, physical component score; SD, standard deviation; SF-12, Short Form-12; SF-36, Short Form-36; VAS, visual analog scale.

^aResponsiveness to change; higher values indicate greater sensitivity of the scale to detect changes following surgical intervention. VAS-pain not available for all patients.

compared with the ODI, 3.1 minutes versus 3.3 minutes ($P < .001$).

DISCUSSION

The SF-12 is an increasingly used instrument for collecting health outcome data for patients with spinal disorders; however, there remains a paucity of literature assessing the reliability of the SF-12 compared with the SF-36 specifically in patients undergoing lumbar surgery. Furthermore, limited evidence exists regarding the correlation between the SF-12 and other commonly used tools, such as the ODI and the VAS following lumbar surgery. This study revealed a strong correlation between the SF-12 and SF-36 for patients undergoing lumbar surgery (both fusion and nonfusion) and demonstrated that the SF-12 and SF-36 are only moderate predictors of the ODI.

The SF-12, SF-36, ODI, and VAS scores were compared at the preoperative, 3-month postoperative, and 6-month postoperative time periods. The finding that the PCS and MCS scores of the SF-12 strongly correlated to those of the SF-36 is

consistent with observations in various nonspine orthopaedic patient populations, including those with osteoarthritis and diabetic foot disease.^{1,2,14} These results are also consistent with the only other study that assessed patients with lumbosacral spinal disorders.⁴ However, Lee et al⁴ only analyzed the preoperative survey data and did not include any postoperative data collection. Thus, ours is the first large study to show that the SF-12 is a valid alternative to the SF-36 for preoperative and postoperative assessments of health status in patients with lumbar surgical disorders undergoing surgery for predominantly radiating leg pain.

A challenge in measuring health outcomes is balancing the need for sufficient data points to achieve reliable measurements with resource utilization (patient time) and ease of completing the tool, which will affect patient compliance. The SF-12 took approximately half the patient completion time compared with the SF-36, and the average time to complete the ODI was similar. While the SF-36 had slightly better correlative values with the ODI, they were not high enough to recommend continuing to

Table 6. Proportion of patients reporting floor^a or ceiling^b effects for each outcome at different time points.

	SF-12 PCS		SF-12 MCS		SF-36 PCS		SF-36 MCS		ODI		VAS	
	Floor	Ceiling	Floor	Ceiling	Floor	Ceiling	Floor	Ceiling	Floor	Ceiling	Floor	Ceiling
Preoperative	0	0	0	0	0	0	0	0	0.21	0	2.5	8.3
3 months	0	0	0	0	0	0	0	0	5.1	0	5.1	0.9
6 months	0	0	0	0	0	0	0	0	3.8	0	11.5	0

^aFloor: Percent of patients with the lowest possible score.

^bCeiling: Percent of patients with the highest possible score.

use the longer SF-36 over the SF-12. Ideally, relieving this time burden on the patients by replacing the SF-36 with the SF-12 should increase patient compliance.

When examining the correlation between the PCS and MCS scores of the SF-12 and SF-36 with the ODI, this study revealed a moderate inverse correlation preoperatively and early postoperatively, which increased to a strongly inverse correlation by 6 months postoperatively. These results strengthen the findings in the literature in patients undergoing lumbar surgery, as Ko and Chae found moderate correlation in their 69 patients at 1 year.⁷ Our moderate inverse correlations found both preoperatively and early postoperatively reinforce the recommendation by Ko and Chae that the ODI and SF-36 provide complementary information.

Our data demonstrated that general health status and condition-specific questionnaires correlate poorly with pain as measured by VAS in patients undergoing lumbar surgery. Very weak or no correlation existed between the SF-12, SF-36, or ODI and VAS preoperatively or postoperatively. These results are comparable to a systematic review which showed little correlation between the SF-36 and VAS after spinal surgery.⁵ These findings support the continued administration of the VAS together with a general health status instrument when measurement of changes in pain is a critical part of the outcome of an intervention. These findings also highlight our lack of understanding of pain, how it is generated by the body, and how variably it can affect patient function.

When optimizing the use of these various instruments, clinicians need to take into account their reliability at measuring change postoperatively. The reliability at measuring postprocedure change in this study was relatively similar for SF-12, SF-36, ODI, and VAS, irrespective of surgical procedure. At early follow up, the sensitivities of all the instruments, except for VAS, were relatively low, especially in the lumbar fusion group. This can likely be explained by the fact that surgeons typically limit patient function in the acute postoperative period to prevent injury. In addition, patients may limit their function due to anxiety before being seen and cleared in follow-up visits. As such, their functional level is often limited by physician or patient-imposed restrictions rather than inherent function. At the 6-month follow up,

the sensitivities of all of the instruments improved compared with their values at the 3-month follow up. VAS remained similar between the 3-month and 6-month follow up, indicating it may be the most useful tool to detect change in the immediate postoperative period. The instruments had a higher reliability at measuring change in the lumbar fusion group than in the nonfusion group. The ODI had a higher reliability at measuring change than the SF-12 or SF-36 and demonstrated a larger effect size at the 6-month postoperative time period. These same patterns in the instruments' reliability at measuring change were also observed when the groups were analyzed by diagnosis (spinal stenosis with degenerative spondylolisthesis or spinal stenosis without degenerative spondylolisthesis) rather than by surgery type. This observation is consistent with that of previous studies comparing the responsiveness to change of these instruments after spine surgery.^{5,6}

Murphy et al⁶ found a significant correlation in the change in ODI compared with the change in the SF-12 PCS scores in their patient population, but concluded that ODI scores were not applicable for evaluating a patient's quality of life, as not all domains improved equally after surgery. DeVine et al⁵ did not find a strong correlation between the instruments and concluded that the 3 tests should be administered together. Our findings demonstrate that both the SF-12 and SF-36, general health status measures, are moderate to strong predictors of ODI preoperatively and postoperatively. As predicted, the ODI, which is a disease-specific instrument, is more sensitive to postoperative change than the SF-12 and SF-36 for patients undergoing lumbar spine surgery. Given that the SF-12 has a strong correlation with the SF-36 and that the ODI is the most reliable at measuring change postoperatively, we recommend using the SF-12 in combination with the ODI to fully assess patient outcomes after lumbar spine surgery.

Several primary limitations of this study exist. First, the study was limited to patients undergoing elective lumbar spinal surgery. Therefore, these results may not be generalizable to patients treated nonoperatively or patients with emergent surgical problems. Second, while the follow-up period analyzed for the various instruments was only 6 months, the largest differences in patient response compared with baseline were observed

during the first 6 postoperative months and then tended to diminish over time, making the early time points the most critical for comparison of sensitivity to change between various instruments. Third, the VAS pain score was not available for the majority of the patients, so the conclusions related to VAS may be less reliable and thus were not a primary focus of this study. Lastly, the National Institutes of Health-supported Patient-Reported Outcomes Measurement Information System (PROMIS) was not readily available at the time this registry began, and thus no comparisons with this newer computer-adaptive testing system exist.¹⁵ PROMIS has recently been validated against the SF-12v2 to estimate health utility index values for patients presenting for lumbar spine surgery, but has not been validated with respect to sensitivity or reliability at measuring changes after lumbar spine surgery, nor has it become universally used yet.¹⁶ Thus, the findings from the present study remain relevant when following patients that previously had SF-36 recorded as their baseline and in situations where the computerized adaptive testing version of PROMIS is not readily available.

CONCLUSIONS

In conclusion, this is the first study to verify that the SF-12 is a valid substitute for the SF-36 for preoperative and postoperative general health status assessments in patients with predominant leg pain undergoing lumbar spine surgery. Both the MCS and PCS components of the SF-12 and SF-36 demonstrated a strong correlation with each other preoperatively and up to 6 months postoperatively. Both the SF-12 and SF-36 were moderate to strong predictors of ODI preoperatively and postoperatively, but lacked the sensitivity to detect postoperative change compared with the ODI. Thus, the SF-12 can substitute for the SF-36 in the follow up of lumbar spine surgery patients, but the ODI should continue to be used to measure disease-specific changes in function in these patients. Additionally, a VAS may be the most useful predictor in detecting early postoperative changes in this patient cohort. This study provides a data-driven basis to rationalize the selection of patient outcomes surveys in this surgical population and lessen patient burden without substantially compromising accuracy or sensitivity.

REFERENCES

1. Wukich DK, Sambenedetto TL, Mota NM, Suder NC, Rosario BL. Correlation of SF-36 and SF-12 component scores in patients with diabetic foot disease. *J Foot Ankle Surg.* 2016;55(4):693–696.
2. Webster KE, Feller JA. Comparison of the short form-12 (SF-12) health status questionnaire with the SF-36 in patients with knee osteoarthritis who have replacement surgery. *Knee Surg Sports Traumatol Arthrosc.* 2016;24(8):2620–2626.
3. Jenkinson C, Layte R, Jenkinson D, et al. A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies? *J Public Health Med.* 1997;19(2):179–186.
4. Lee CE, Browell LM, Jones DL. Measuring health in patients with cervical and lumbosacral spinal disorders: is the 12-item short-form health survey a valid alternative for the 36-item short-form health survey? *Arch Phys Med Rehabil.* 2008;89(5):829–833.
5. DeVine J, Norvell DC, Ecker E, et al. Evaluating the correlation and responsiveness of patient-reported pain with function and quality-of-life outcomes after spine surgery. *Spine (Phila Pa 1976).* 2011;36(21 Suppl):S69–S74.
6. Murphy HA, Warnick E, McEntee R, et al. Which domains of the ODI best predict change in physical function in patients after surgery for degenerative lumbar spondylolisthesis? *Spine (Phila Pa 1976).* 2018;43(11):805–812.
7. Ko S, Chae S. Correlations between the SF-36, the Oswestry-Disability Index and Rolland-Morris Disability Questionnaire in patients undergoing lumbar decompression according to types of spine origin pain. *Clin Spine Surg.* 2017;30(6):E804–E808.
8. Singh A, Gnanalingham K, Casey A, Crockard A. Quality of life assessment using the Short Form-12 (SF-12) questionnaire in patients with cervical spondylotic myelopathy: comparison with SF-36. *Spine (Phila Pa 1976).* 2006;31(6):639–643.
9. Imaeda T, Uchiyama S, Wada T, et al. Reliability, validity, and responsiveness of the Japanese version of the Patient-Rated Wrist Evaluation. *J Orthop Sci.* 2010;15(4):509–517.
10. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care.* 1990;28(7):632–642.
11. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis.* 1987;40(2):171–178.
12. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: L. Erlbaum Associates; 1988.
13. Shrout PE. Measurement reliability and agreement in psychiatry. *Stat Methods Med Res.* 1998;7(3):301–317.
14. Gandhi SK, Salmon JW, Zhao SZ, Lambert BL, Gore PR, Conrad K. Psychometric evaluation of the 12-item short-form health survey (SF-12) in osteoarthritis and rheumatoid arthritis clinical trials. *Clin Ther.* 2001;23(7):1080–1098.
15. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol.* 2010;63(11):1179–1194.
16. Raad M, Neuman BJ, Kebaish KM, Riley LH 3rd,

Skolasky RL. Estimating health utility in patients presenting for spine surgery using Patient-Reported Outcomes Measurement Information System (PROMIS) health domains. *Spine (Phila Pa 1976)*. 2019;44(13):908–914.

Disclosures and COI: No funds were received in support of this study. Emory Institutional Review Board approval was received for this study.

Corresponding Author: Scott D. Boden, MD, Emory University School of Medicine, Department of Orthopaedics, 59 Executive Park South,

Suite 3047, Atlanta, GA 30329. Phone: (404) 778-7143; Fax: (404) 778-7117; Email: sboden@emory.edu.

Published 30 June 2020

This manuscript is generously published free of charge by ISASS, the International Society for the Advancement of Spine Surgery. Copyright © 2020 ISASS. To see more or order reprints or permissions, see <http://ijssurgery.com>.