JAMA Surgery | Original Investigation

Development and Validation of a Prediction Model for Pain and Functional Outcomes After Lumbar Spine Surgery

Sara Khor, MASc; Danielle Lavallee, PharmD, PhD; Amy M. Cizik, PhD, MPH; Carlo Bellabarba, MD; Jens R. Chapman, MD; Christopher R. Howe, MD; Dawei Lu, MD; A. Alex Mohit, MD; Rod J. Oskouian, MD; Jeffrey R. Roh, MD, MBA; Neal Shonnard, MD; Armagan Dagal, MD; David R. Flum, MD, MPH

IMPORTANCE Functional impairment and pain are common indications for the initiation of lumbar spine surgery, but information about expected improvement in these patient-reported outcome (PRO) domains is not readily available to most patients and clinicians considering this type of surgery.

OBJECTIVE To assess population-level PRO response after lumbar spine surgery, and develop/validate a prediction tool for PRO improvement.

DESIGN, SETTING, AND PARTICIPANTS This statewide multicenter cohort was based at 15 Washington state hospitals representing approximately 75% of the state's spine fusion procedures. The Spine Surgical Care and Outcomes Assessment Program and the survey center at the Comparative Effectiveness Translational Network prospectively collected clinical and PRO data from adult candidates for lumbar surgery, preoperatively and postoperatively, between 2012 and 2016. Prediction models were derived for PRO improvement 1 year after lumbar fusion surgeries on a random sample of 85% of the data and were validated in the remaining 15%. Surgical candidates from 2012 through 2015 were included; follow-up surveying continued until December 31, 2016, and data analysis was completed from July 2016 to April 2017.

MAIN OUTCOMES AND MEASURES Functional improvement, defined as a reduction in Oswestry Disability Index score of 15 points or more; and back pain and leg pain improvement, defined a reduction in Numeric Rating Scale score of 2 points or more.

RESULTS A total of 1965 adult lumbar surgical candidates (mean [SD] age, 61.3 [12.5] years; 944 [59.6%] female) completed baseline surveys before surgery and at least 1 postoperative follow-up survey within 3 years. Of these, 1583 (80.6%) underwent elective lumbar fusion procedures; 1223 (77.3%) had stenosis, and 1033 (65.3%) had spondylolisthesis. Twelve-month follow-up participation rates for each outcome were between 66% and 70%. Improvements were reported in function, back pain, and leg pain at 12 months by 306 of 528 surgical patients (58.0%), 616 of 899 patients (68.5%), and 355 of 464 patients (76.5%), respectively, whose baseline scores indicated moderate to severe symptoms. Among nonoperative patients, 35 (43.8%), 47 (53.4%), and 53 (63.9%) reported improvements in function, back pain, and leg pain, respectively. Demographic and clinical characteristics included in the final prediction models were age, sex, race, insurance status, American Society of Anesthesiologists score, smoking status, diagnoses, prior surgery, prescription opioid use, asthma, and baseline PRO scores. The models had good predictive performance in the validation cohort (concordance statistic, 0.66-0.79) and were incorporated into a patient-facing, web-based interactive tool (https://becertain.shinyapps.io/lumbar_fusion_calculator).

CONCLUSIONS AND RELEVANCE The PRO response prediction tool, informed by population-level data, explained most of the variability in pain reduction and functional improvement after surgery. Giving patients accurate information about their likelihood of outcomes may be a helpful component in surgery decision making.

JAMA Surg. 2018;153(7):634-642. doi:10.1001/jamasurg.2018.0072
Published online March 7, 2018.

Invited Commentary
page 643

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: David R. Flum, MD, MPH, Department of Surgery, University of Washington School of Medicine, 1959 NE Pacific St, Box 356410, Seattle, WA 98195 (daveflum@uw.edu).

Lumbar spine fusion surgery is an increasingly performed and expensive procedure that has raised controversy about its appropriate utilization. The results of randomized clinical trials have been split on the efficacy of spine fusion for back pain,[1-4] with certain subgroups of patients experiencing less than 20% chance of improvement.[5] To maximize value and reduce cost in spine care, there is an increasing focus on better patient selection and improved informed decision making about surgery. While functional impairment and pain are the most common symptoms that lead to lumbar spine fusion, information about expected improvement in these patient-reported outcome (PRO) domains is not readily available to most patients and clinicians when they are deciding on operating. In 2014, Washington state's Dr Robert Bree Collaborative developed a set of standards that have been incorporated into insurance contracts that require routine PRO measurements in patients undergoing lumbar fusion and public release of PRO data to improve shared decision making.[6,7] Improvement in PRO levels after spine surgery has not been assessed on a population level, and there is limited information about factors associated with PRO response that can be used to inform shared decision making.

Patient characteristics, modifiable risk factors, and surgical factors have been shown to impact infection rates and other adverse events,[8-10] but much less is known about PRO response after spine surgery. The Bree recommendations for PRO gathering[6,7] have aimed to provide patients and clinicians with information to improve treatment decisions or determine modifiable variables that could improve PRO responses after surgery. Recent reports have explored PRO responses in a single center or using a registry of voluntarily reporting surgeons,[11-13] but these may not be adequately representative of patients undergoing surgery in the community at large. To date, to our knowledge, there have not been any statewide, population-based analyses of PRO responses among patients having lumbar spine fusion surgery.

The purpose of this study was to describe PRO responses among patients who did and did not undergo lumbar fusion surgery, using data drawn from a prospective quality improvement program in Washington state (Spine Surgical Care and Outcomes Assessment Program [Spine SCOAP]), and to develop and validate a prediction model for improvement in function, back pain, and leg pain 1 year after surgery. The study aimed to create an interactive shared decision-making tool that can help identify the patients who are most likely to benefit from lumbar fusion surgery.

Key Points

Question Which patients are most likely to have improvement in function, back pain, and leg pain after lumbar fusion surgery?

Findings Using statewide prospective data from 15 hospitals and 1965 adult surgical candidates, 3 prediction tools were generated to predict the likelihood of improvements in function, back pain, and leg pain after lumbar fusion surgery. The predictive ability and calibration of these predictive tools were confirmed in a validation cohort.

Meaning These predictive tools could be incorporated into decision-making activities in the clinic and may be helpful in managing expectations for patients considering lumbar fusion surgeries.

Methods

Study Population

The Spine SCOAP registry includes medical record data on patient characteristics, perioperative medical and surgical details, and clinical outcomes for spine surgery patients from hospitals across Washington. The registry represents approximately 75% of eligible elective spine fusion procedures in the state.[14] Spine SCOAP excludes patients who have procedures on more than 5 spinal levels and those whose procedures are done for diagnoses of trauma, tumor, or infection. The Comparative Effectiveness Translational Network (CERTAIN) at the University of Washington partners with spine clinics to collect baseline and follow-up PRO scores from candidates for spine surgery through the CERTAIN Hub, which is a web-based portal, and a survey center.[15] The CERTAIN linked these PROs to the Spine SCOAP records, producing the deidentified data set for this study. Adult patients who had a clinic visit to consider a lumbar fusion procedure between January 1, 2012, and December 31, 2015, at 15 hospitals were included. Baseline clinical characteristics for the nonsurgical group were limited because these individuals did not have a procedure and were not part of the Spine SCOAP data collection process. The final date of follow-up data collection for this study was December 31, 2016.

This research project used deidentified patient data. As a result, it was determined by the human subject division of the institutional review board at the University of Washington to be exempt.

Primary Outcomes

The 3 outcomes of interest were function, back pain, and leg pain improvement at 1 year. Function was measured using the Oswestry Disability Index (ODI), which is a widely accepted standard for assessing functional status of patients with back pain.[16] The ODI questionnaire consists of 10 sections designed to assess limitations of various activities of daily living. Each section is scored on a 0 to 5 scale, where a score of 5 indicates the greatest disability. The result is a composite index on a 100-point scale, with a score of 100 reflecting the maximum possible low back disability at the time of survey. The Numerical Rating Scale (NRS),[17] on which patients are asked to rate their average pain intensity in the past week on an 11-point scale (where 0 indicates no pain and 10 indicates worst possible pain), has been widely used as an instrument to measure back and leg pain. On the NRS, back and leg pain were assessed using these questions: "How would you rate your average back pain in the past week?" and "How would you rate your average leg pain in the past week?" For our study, functional improvement was defined as a reduction of 15 points or more, from baseline in the ODI. Back pain and leg pain improvement was defined by a reduction of 2 points or more from

baseline in the NRS.[17,18] These are considered the minimal clinically important difference (MCID).[16,19] The MCID levels represent the threshold value in these PRO scores that patients and clinicians perceive as clinically meaningful. They are recommended by current guidelines for assessment of lumbar fusion outcomes.[20,21]

The PRO surveys were administered preoperatively (0-60 days before surgery) and at 2 months, 6 months, 12 months, 18 months, 24 months, 30 months, and 36 months postsurgery. Postsurgical surveys were administered in a multimodal approach (mail, email, telephone, or text message), based on patient preferences. Patients were included in this study if they had at least 1 baseline (presurgery) score and 1 postsurgery score (up to 36 months after surgery) on function, back pain, or leg pain.

Statistical Analysis

Model Development

Among 783 patients with at least 1 baseline ODI score and 1 postsurgery ODI score, 238 (30.4%) did not complete the 12-month ODI questionnaire because they had withdrawn from the study, could not be reached, or chose not to respond. Similarly, 533 of 1466 respondents (36.4%) and 218 of 726 (30.0%) did not complete the NRS back pain and NRS leg pain assessments, respectively, at 1 year. To account for missing data among the predictors and follow-up data, which were assumed to be missing at random, a multiple imputations procedure with 40 imputation iterations was performed (eAppendix in the Supplement).

The training data sets used for the model development for each outcome were obtained by randomly selecting 85% of the data. Since the primary outcomes of interest were MCID improvements from baseline (a binary variable), only patients with the possibility for improvement were included. The ODI model included only patients with a baseline ODI score greater than or equal to 15, and the NRS models included only patients with baseline NRS levels greater than or equal to 2. Three binary logistic regression models with the dependent variables ODI improvement, NRS-measured back pain improvement, and NRS-measured leg pain improvement at 1 year were generated using the imputed training data sets. The results were pooled to yield estimates, confidence intervals, and *P* values that incorporated missing-data uncertainty. Baseline patient characteristics and surgical parameters listed in **Table 1** were assessed as independent variables for the models. These parameters included patient demographics and risk factors (7 variables), comorbid conditions (7 variables), diagnosis (11 variables), opiate use (1 variable), surgical (2 variables), and hospital characteristics (2 variables). Predictors were included in the multivariate analyses if they yielded a *P* value of less than .05 in the univariate analyses, or had a *P* value greater than .05 but were considered to be clinically relevant by author consensus. **Table 2** shows the list of variables that remained in the final models.

Model Performance

To assess model calibration, calibration plots were generated comparing the predicted with the observed probabilities, and

Table 1. Patient Characteristics

Characteristics	No. (%) (N = 1583)
Age, y, mean (SD)	61.3 (12.5)
Female	944 (59.6
Obese[a]	727 (46.0)
Race	
White	1422 (89.8)
African American	41 (2.6)
Asian	29 (1.8)
Other	91 (5.8)
Insurance	
Private	1103 (69.7)
Medicare	189 (11.9)
Medicaid	131 (8.3)
Workers' compensation	107 (6.8)
Other	53 (3.4)
Smoking status	
Never	721 (45.6)
Current	205 (13)
Previous	607 (38.3)
Unknown	50 (3.2)
Low albumin[b]	65 (4)
ASA score ≥3	510 (32.3)
Comorbid conditions	
Rheumatoid arthritis	31 (2)
Hypertension	919 (58)
Diabetes	285 (18)
Asthma	219 (14)
Sleep apnea	245 (15)
Coronary artery disease	148 (9)
Prior spine surgery	395 (25)
Diagnosis	
Degenerative disc	473 (30)
Disc herniation	220 (14)
Postlaminectomy/failed back syndrome	238 (15)
Instability[c]	274 (17)
Spondylosis	281 (18)
Spondylolisthesis	1033 (65)
Stenosis	1223 (77)
Pseudarthrosis	75 (5)
Other spine problems	264 (17)
Radiculopathy	1461 (92)
Myelopathy	55 (3)
Medications	
Prescription opiate use	889 (56)
Operative Characteristics	
Invasiveness index, mean (SD)[d]	9.6 (4.6)
Surgeon type	
Neurosurgeon	625 (39.7)
Orthopedic	945 (60)
Other	5 (3.2)

(continued)

Table 1. Patient Characteristics (continued)

Characteristics	No. (%) (N = 1583)
Hospital Characteristics	
Hospital type	
Nonteaching hospital	582 (36.8)
Teaching hospital with spine surgical residency	287 (18.1)
Teaching hospital without spine surgical residency	714 (45.1)
Large hospital[e]	1543 (97)

Abbreviation: ASA, American Society of Anesthesiologists.

[a] Obesity was defined as a body mass index, calculated as weight in kilograms divided by height in meters squared, greater than or equal to 30.

[b] The cutoff level for low albumin was defined as less than or equal to 35 g/L; to convert albumin into grams per liter, multiply by 10.

[c] Instability refers to excessive motion (≥4 mm translation, and/or 10° angular motion) that is readily documented by radiographic studies and results in pain, deformity, and/or neurological deficit.

[d] An invasiveness index was calculated based on the type of intervention at each vertebral level (anterior/posterior decompression, anterior/posterior fusion, and anterior/posterior instrumentation) and the number of operated levels. Each intervention at each level was scored 1 point, and the scores were summed to produce the invasiveness index. This index has been shown to correlate with clinical outcomes and complications in spine surgery patients.[22,23]

[e] A large hospital was defined as any having more than 200 licensed beds.

the mean observed to expected ratios of probabilities were calculated. To assess the models' abilities to discriminate between those with or without the outcome of interest, the concordance (C-statistics) from each of the imputation iterations were averaged to create single C-statistics.

Model Validation

The validation data set for each outcome variable consisted of the remaining 15% of the full data after the training data set was randomly selected. Each prespecified model was fitted with the patient characteristics of the validation cohort to predict the outcome of interest. The predictions were then compared with the actual outcome values, and C-statistics and calibration plots were generated.

Probability Estimates

The regression coefficients from the final models (eTable 1 in the Supplement) were used to estimate the probability of achieving MCID improvement at 12 months in ODI scores, NRS back pain scores, and NRS leg pain scores for 5 hypothetical patients undergoing lumbar fusion, based on their unique personal and surgical characteristics at surgical consultation.

All analyses were performed using Stata version 14 (StataCorp). All tests were 2-sided and *P* values less than .05 were considered to be significant. All mean values were reported with standard deviations.

Results

A total of 1583 patients (mean [SD] age, 61.3 [12.5] years; 944 [59.6%] female) were included in this study. Table 1 summarizes baseline characteristics. The most common indications for surgery were stenosis (n = 1223; 77.9%) and spondylolisthesis (n = 1033; 65.3%). A total of 1461 cases (92.3%) were identified to have radiculopathy. The majority of cases were performed in teaching hospitals (n = 1001; 63.2%), but only 287 (18.1%) were performed in hospitals with a spine surgical residency program.

The **Figure** shows the unadjusted PRO scores at each point. The median baseline ODI, NRS back pain, and NRS leg pain scores were 46 of 100, 6 of 10, and 6 of 10, respectively. On average, PRO scores improved over time after surgery, with most of the improvements observed in the first 2 months. The PRO scores were significantly improved at 12 months compared with baseline scores at a population level, but large variations were observed at the individual level (**Table 3**).

At 12 months, 109 of 545 patients (20.0%) still experienced severe disability, and 31 (5.7%) had extremely severe symptoms or were bedbound. Of 528 patients who had a baseline score of 15 or greater on the ODI, 306 (58.0%) experienced an MCID improvement at 12 months, while 86 (15.8%) reported no change or worse function compared with their baseline scores.

Similarly, at 12 months, 251 of 933 patients (26.9%) still had moderate pain, and 117 (12.5%) reported severe pain. A total of 616 of 899 patients who had a baseline NRS score for back pain of 2 or greater (68.5%) achieved a 2-point MCID improvement in back pain, and 180 (19.2%) had no improvement or a higher (worse) score.

Leg pain response rate was the highest, with 355 of 464 patients who had started with a baseline NRS score for leg pain of 2 or greater (76.5%) achieving at least an MCID at 12 months. However, severe leg pain was still experienced 12 months after surgery by 59 of 508 patients (12.7%).

A total of 382 individuals (mean [SD] age, 62 [13] years; 202 [53.0%] female) had an office visit with a spine surgeon but did not have surgery. A total of 241 individuals (63.1%) visited a surgeon for presurgical purposes, 55 (14.4%) for consultation, and 86 (22.5%) for a purpose that was not described. There was no difference in ODI scores, NRS-measured back pain, and NRS-measured leg pain at baseline or over time between groups who visited the clinic for these different purposes. There was also no significant difference in baseline ODI scores, NRS back pain scores, and NRS leg pain scores between the nonsurgical group and the surgical group. Of the individuals who did not have surgery, 124 of 426 (29.1%) and 35 of 80 (43.8%) experienced an MCID improvement in ODI scores at 2 and 12 months, respectively. In this group, 205 of 435 patients (47.1%) and 47 of 88 patients (53.4%) achieved an MCID in NRS back pain scores at 2 and 12 months, respectively. Similar to the surgical group, leg pain improvement rate was the highest, with 198 of 410 patients (48.3%) achieving MCID at 2 months, and 53 of 83 patients (63.9%) achieving this level of improvement in 12 months. At both 2 and 12 months, patients who had undergone operations had a significantly higher rate of response compared with those who had not had an operation.

Model Development and Validation

Variables included in the final models were age, sex, race, insurance status, American Society of Anesthesiology (ASA)

Table 2. Model Odd Ratios

Characteristics	ODI		NRS Back Pain		NRS Leg Pain	
	Odds Ratio (95%CI)	*P* Value	Odds Ratio (95%CI)	*P* Value	Odds Ratio (95%CI)	*P* Value
Age	1.00 (0.98-1.02)	.75	1.02 (1.00-1.03)	.03	0.99 (0.97-1.02)	.59
Male	0.92 (0.64-1.33)	.67	0.92 (0.64-1.31)	.63	0.80 (0.48-1.34)	.40
Insurance[a]						
Medicaid	0.38 (0.14-1.02)	.06	0.41 (0.24-0.69)	<.001	0.75 (0.27-2.07)	.58
Workers' compensation	0.20 (0.07-0.53)	<.001	0.52 (0.27-1.02)	.06	0.48 (0.19-1.2)	.12
Other	0.74 (0.46-1.19)	.22	0.74 (0.45-1.21)	.23	1.44 (0.76-2.73)	.27
Race/ethnicity nonwhite[a]	0.97 (0.55-1.69)	.91	0.89 (0.51-1.54)	.67	0.58 (0.27-1.28)	.18
ASA score ≥3	0.84 (0.55-1.27)	.41	0.79 (0.57-1.08)	.14	0.66 (0.41-1.04)	.07
Smoking status[a]						
Current	0.43 (0.22-0.84)	.01	0.58 (0.35-0.96)	.03	0.64 (0.25-1.64)	.35
Previous	0.66 (0.44-0.99)	.05	0.81 (0.60-1.09)	.17	0.76 (0.48-1.2)	.23
Prior surgery	0.61 (0.35-1.06)	.08	0.83 (0.55-1.26)	.39	0.98 (0.56-1.69)	.93
Spondylolisthesis	1.74 (0.93-3.27)	.08	1.63 (1.19-2.22)	<.001	1.3 (0.71-2.35)	.40
Disc herniation	1.64 (0.96-2.82)	.07	1.12 (0.73-1.73)	.61	1.61 (0.72-3.59)	.24
Postlaminectomy/failed back syndrome	0.92 (0.48-1.76)	.81	0.94 (0.63-1.40)	.75	0.44 (0.25-0.77)	<.001
Stenosis	1.13 (0.67-1.91)	.64	1.07 (0.74-1.56)	.70	1.17 (0.63-2.18)	.61
Pseudarthrosis	0.35 (0.11-1.10)	.07	0.47 (0.22-1.02)	.06	0.6 (0.2-1.79)	.36
Radiculopathy	0.63 (0.31-1.27)	.20	0.97 (0.54-1.74)	.91	0.38 (0.12-1.19)	.10
Prescription opiate use	1.05 (0.74-1.49)	.77	0.65 (0.50-0.86)	<.001	0.72 (0.48-1.09)	.13
Asthma	0.54 (0.30-0.98)	.04	0.86 (0.55-1.32)	.48	0.87 (0.45-1.68)	.67
Baseline						
ODI score	1.05 (1.03-1.07)	<.001				
NRS back pain score			1.53 (1.44-1.64)	<.001		
NRS leg pain score					0.80 (0.48-1.34)	<.001

Abbreviations: ASA, American Society of Anesthesiologists; NRS, Numerical Rating Scale; ODI, Oswestry Disability Index.

[a] Reference values were private insurance, white race/ethnicity, and never smoked.

score, smoking status, diagnoses, prior surgery status, prescription opioid use, asthma, and baseline PRO scores (Table 2). (The ASA is a measure of general physical health status at the time of the operation; this risk score is used to describe preoperative patient risk. A patient with ASA score of 3 or more is considered to have severe or extreme systemic disease.) According to the final logistic regression model for ODI scores, factors that were significantly associated with lower odds of improvement were nonprivate insurance (workers' compensation odds ratio [OR], 0.20; 95% CI, 0.07-0.53), current smoking (OR, 0.43; 95% CI, 0.22-0.84) or previous smoking (OR, 0.66; 95% CI, 0.44-0.99), asthma (OR, 0.54; 95% CI, 0.30-0.98), and a lower baseline score (OR, 1.05; 95% CI, 1.03-1.07; Table 2). In the back pain model, factors that were associated with lower odds of NRS back pain improvement included younger age (OR, 1.02; 95% CI, 1.00-1.03), nonprivate insurance (OR, 0.41; 95% CI, 0.24-0.69), current smoking (OR, 0.58; 95% CI, 0.35-0.96), current spondylolisthesis (OR, 1.63; 95% CI, 1.19-2.22), use of opiate prescription (OR, 0.65; 95% CI, 0.50-0.86), and a low baseline NRS back pain score (OR, 1.53; 95% CI, 1.44-1.64). There were fewer significant predictors in the leg pain model, including only postlaminectomy or failed back syndrome (OR, 0.44; 95% CI, 0.25-0.77) and a low baseline NRS leg score (OR, 0.80; 95% CI, 0.48-1.34).

All 3 models had good discrimination (C-statistic, 0.73-0.75) and were well calibrated (observed to expected ratios, 1.00) in the development cohort. There was also good discrimination in the validation cohort (C-statistics, 0.66 for ODI, 0.79 for back pain, and 0.69 for leg pain). Observed to expected ratios were between 0.94 and 1.02, with confidence intervals spanning 1 for all 3 models. This indicated no significant differences between the observed and model predictions. (Details of model performance are in eTables 1 and 2 and eFigures 1 and 2 in the Supplement.)

Applying the Model for Future Predictions

Table 4 shows 5 hypothetical lumbar fusion candidates with different baseline characteristics and baseline pain and function scores (Table 4). Patients B and C are identical except that patient B is a current smoker and patient C has stopped smoking. Through this sole difference, this hypothetical patient has an increased likelihood of improvement in function, back pain, and leg pain by 10%, 7%, and 5%, respectively.

Patients D and E are identical except that patient D has severe disability and pain before surgery and patient E has moderate disability and pain. The likelihood of improvement in function, back pain, and leg pain were 47%, 37%, and 33% lower, respectively, in the patient with moderate disability and pain, highlighting the importance of informed decision making and expectation management for patients with moderate disability and pain. A user-friendly version of the predictive

calculator has been made available online at https://becertain.shinyapps.io/lumbar_fusion_calculator.

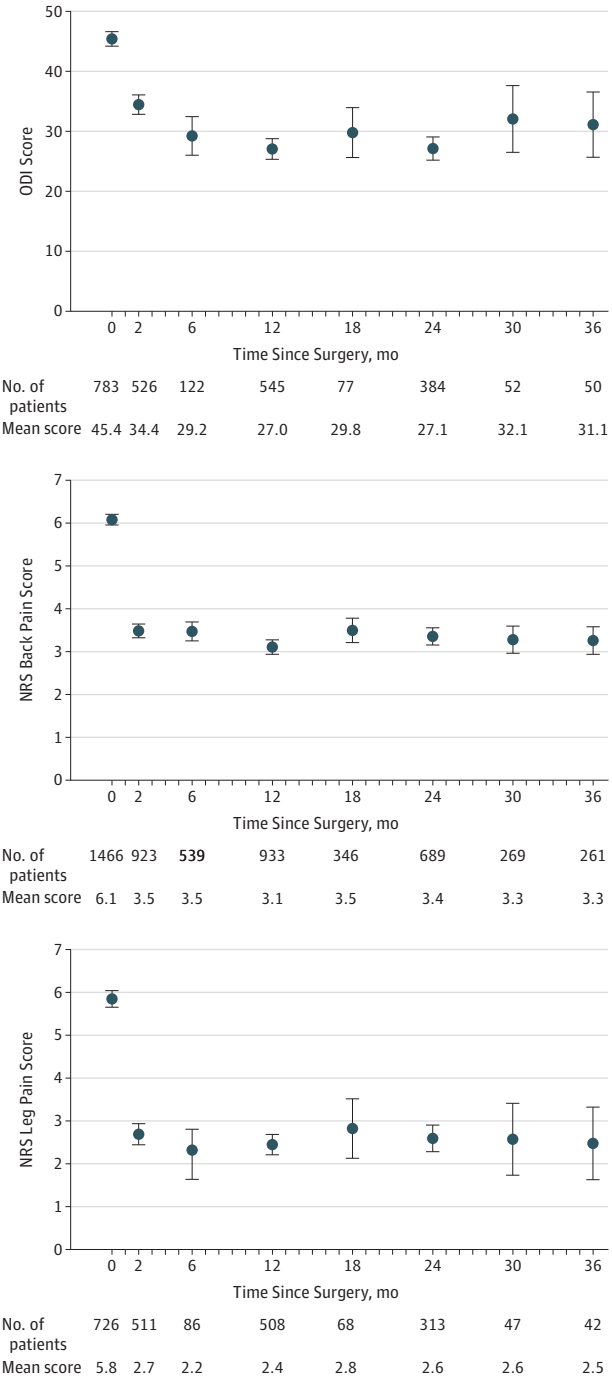
Discussion

As we move toward value-based care and shared decision making, there is an increasing need to collect and use PRO scores not just in research settings, but also in routine clinical care or quality improvement activities. To our knowledge, this is the first statewide assessment of PRO scores concerning elective lumbar fusion surgery. We found that most patients had an improvement in PRO scores at 12 months: 306 surgical patients (58.0%) achieved an MCID improvement in ODI score, 616 patients (68.5%) in NRS-measured back pain, and 355 patients (76.5%) in NRS-measured leg pain, respectively. More importantly, we found significant variability in outcomes based on patient characteristics. We developed a user-friendly tool to allow personalized PRO prediction (specifically in improvement in function, back pain, and leg pain at 12 months after surgery) based on patient characteristics, diagnoses, and baseline PRO scores. Our final models had good calibration and predictive performance in a validation cohort (C-statistics, 0.66-0.79), demonstrating that they can accurately predict potential outcomes in new populations with similar characteristics.

Subgroups that were less likely to improve with lumbar fusion surgery included those with Medicaid or workers' compensation as their primary insurance source, current and previous smokers, and those with low (better) preoperative disability and pain scores. This finding correlates with other spine surgery literature, which has suggested workers' compensation, Medicaid, and preoperative scores as determinants of clinical outcomes.[11,12,24-27] In our models, smoking status was a key predictor of potentially worse outcome. Current smokers have a significantly lower likelihood of improvement in both function (OR, 0.43; 95% CI, 0.22-0.84) and back pain (OR, 0.58; 95% CI, 0.35-0.96) when compared with people who have never smoked. Patients who quit smoking have increased but still lower odds of improvement in function (OR, 0.66; 95% CI, 0.44-0.99; P = .045) and back pain (OR, 0.81; 95% CI, 0.60-1.09; P = .17) compared with those who have never smoked, although the difference in the odds of back pain improvement was insignificant. Smoking has been shown in other studies to contribute to poor clinical outcomes, higher nonunion rates, higher infection rates,[28-30] and a much reduced rate in ODI improvement compared with the outcomes in nonsmokers.[29] Our findings highlight the importance of preoperative tobacco cessation efforts. Future studies should examine the sensitivity of PRO scores to the duration of smoking cessation to better inform tobacco cessation program designs. Interestingly, asthma was found to be significantly associated with lower odds of ODI improvement, a factor not reported in previous studies. However, asthma may be a marker for other characteristics, and to further explore the relationship between asthma and PRO scores after lumbar spine surgery, we are currently conducting a prospective study funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (clinicaltrials.gov identifier NCT02611479) on severity, duration, and symptoms of asthma.

Invasiveness of surgery, surgical techniques, surgeon type (orthopedic surgeon vs neurosurgeon), and hospital type (large

Figure. Unadjusted Scores of Functional Disability, Back Pain, and Leg Pain at Baseline and Multiple Points After Surgery



Error bars show 95% confidence intervals. NRS, Numeric Rating Scale; ODI, Oswestry Disability Index.

Table 3. Unadjusted PRO Scores and Improvement at 12 Months

	No. (%)	
	At Baseline	At 12 Months
Disability		
No. of patients	783	545
ODI score/100, median (range)	46 (2-100)	24 (0-90)
0 to 20 (Minimal)	55 (7.02)	248 (45.5)
21 to40 (Moderate)	266 (34.0)	157 (28.8)
41 to 60 (Severe)	307 (39.2)	109 (20.0)
61-100 (Extremely severe/bedbound)	155 (19.8)	31 (5.7)
Back Pain		
No. of patients	1466	933
NRS back pain score/10, median (range)	6 (0-10)	3 (0-10)
0 to2 (Minimal)	229 (15.6)	565 (60.6)
3 to 6 (Moderate)	516 (35.2)	251 (26.9)
7 to 10 (Severe)	712 (49.2)	117 (12.5)
Leg Pain		
No. of patients	726	508
NRS leg pain score/10, median (range)	6 (0-10)	1 (0-10)
0 to 2 (Minimal)	143 (19.7)	345 (67.9)
3 to 6 (Moderate)	254 (35.0)	104 (20.5)
7 to 10 (Severe)	329 (45.3)	59 (11.6)
Disability Improvement		
MCID in ODI at 12 mo, No. (%)[a]	NA	306 (58.0)
30% Improvement in score at 12 mo, No. (%)[a]	NA	316 (59.9)
Any improvement at 12 mo, No. (%)	NA	459 (84.2)
Back Pain Improvement		
MCID in ODI at 12 mo, No. (%)[a]	NA	616 (68.5)
30% Improvement in score at 12 mo, No. (%)[a]	NA	586 (65.2)
Any improvement at 12 mo, No. (%)	NA	753 (80.8)
Leg Pain Improvement		
MCID in ODI at 12 mo, No. (%)[a]	NA	355 (76.5)
30% Improvement in score at 12 mo, No. (%)[a]	NA	344 (74.1)
Any improvement at 12 mo, No. (%)	NA	415 (81.9)

Abbreviations: MCID, minimal clinically important difference; NA, not applicable; NRS, numerical rating scale; ODI, Oswestry Disability Index; PRO, patient-reported outcome.

[a] These fields include only those with baseline score of 15 or more for ODI (n = 528), baseline NRS score of 2 or more for back pain (n = 899), or baseline NRS score of 2 or more for leg pain (n = 464).

vs small, teaching hospital vs nonteaching hospital) did not appear to have an effect on disability or pain improvement and were dropped from the models, a recognition that patient variables were the main factors associated with PRO improvement. This study also showed that patients who did not have surgery experienced a mean improvement of 24% in ODI score after 12 months. Although the framework for data assessment in this study did not allow for a direct comparison between the operative group and the nonoperative group, this rate of improvement in nonoperated patients is within the published range.[2-4,31]

As PRO scores become more comprehensive and sophisticated, it is important that we develop translation models to appropriately manage the results. Prior prediction models in lumbar surgery examined clinical outcomes such as postoperative adverse events, but did not look at PRO scores or did so in a limited population.[8,10-13] McGirt et al[11] published an ODI prediction model using data from a single institution. Their model explained about 50% of the data. They used a continuous outcome instead of a dichotomous outcome, making it difficult to directly compare performance of their model with ours. More recently, the same group developed prediction models for ODI, back pain, and leg pain using multicenter data from the Quality Outcomes Database.[13] Their study included both patients undergoing fusion procedures and patients who were not undergoing fusion procedures, and they treated the 12-month scores as ordinal rather than dichotomous outcomes. While they included multiple centers from different states, they only included voluntarily participating surgeon practices, and that may not be as representative as the approach of Spine SCOAP, which included almost all practitioners in the state. Despite the differences in population and outcomes, there was general agreement between our findings in terms of the predictors for a positive outcome, although it is difficult to compare the weight of each predictor given the different outcome measures.

Limitations

There are several limitations to our study. Adverse events were not included in our models because the objective was to develop a tool for presurgical prediction. Moreover, our models were built on data from Washington state, where 90% of the patients were white and 70% had private insurance. Our results may not be generalizable to other state populations.

Another limitation was the incomplete survey rates for outcomes at 12 months, which were 30% to 36%. However, these missing data were accounted for through the use of a multiple imputation procedure.[32]

Further, there might be important predictors of outcomes that are not available in our data set, such as symptom durations, measures of anxiety, depression, widespread body pain, and pain interference. Other studies have shown an association of depression with poor function outcomes after spine surgery.[33,34] Widespread pain has also been shown to be associated with increased opioid use and poor spine and anthroplasty outcomes.[35-37] We plan to address the effects of these factors in our ongoing study funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases. Moreover, preoperative diagnoses used in the model were determined by the treating physician and were not validated independently. We also acknowledge that continuous outcomes, such as an exact postoperative ODI score, might be easier to interpret than predicting the odds of having an improvement. However, MCID measures are well accepted in the spine research field as the standard outcomes and are recommended by current guidelines for spine outcomes evaluations.[20] Lastly, our outcome cutoffs were chosen based on commonly used MCID measures for spine procedures (2 points for the NRS score and 15 points for the ODI score), but we acknowledge that a single agreed-on MCID has yet to be established.[38] To accommodate the different preferences in the way outcomes are presented (eg, MCID, percentage of change, minimal pain or

Table 4. Results of the Models on Hypothetical Patients

Characteristic	Hypothetical Patient A	B	C	D	E
Age, y	65	25	25	55	55
Sex	Male	Male	Male	Female	Female
Insurance type	Medicaid	Medicaid	Medicaid	Workers' compensation	Workers' compensation
Race/ethnicity	White	White	White	White	White
ASA score	I	III	III	I	I
Smoking status	Never	Current	Previous	Previous	Previous
Prior surgery	No	Yes	Yes	No	No
Diagnosis	Spondylolisthesis	Stenosis with radiculopathy	Stenosis with radiculopathy	Spondylolisthesis with radiculopathy	Spondylolisthesis with radiculopathy
Prescription opiate use	No	Yes	Yes	Yes	Yes
Asthma	No	Yes	Yes	No	No
Baseline scores (scale)					
ODI disability score (100)	80	80	80	90	46
NRS back pain (10)	8	8	8	9	6
NRS leg pain (10)	6	6	6	9	6
Likelihood of improvement of ≥1 MCID, %[a]					
Function	89	33	43	77	30
Back pain	83	26	33	82	46
Leg pain	90	51	56	85	52

Abbreviation: MCID, minimal clinically important difference.

[a] Minimal clinically important differences were defined as a score increase of 15 or more on the Oswestry Disability Index (on a scale of 100) for function and 2 on a Numerical Rating Scale score (on a scale of 10) for back pain and leg pain.

function), we included these other outcomes in the user-friendly online model.

Conclusions

In 2014, Washington state encouraged public reporting of PRO data in spine surgery and the use of these data to improve shared decision making. This information was unknown at the time of the recommendation, but this study demonstrates that, on a statewide level, the majority of patients had improvement in disability, back pain, and leg pain at 1 year. The evaluation also revealed subgroups within the broader population who had better-than-expected or worse-than-expected outcomes, and these could be predicted using baseline characteristics. We developed 3 clinical prediction models to determine the probabilities of improvement in function, back pain, and leg pain for lumbar fusion candidates. These models showed good accuracy in the derivation and validation cohorts and could be incorporated in a clinical setting, where a clinician and/or patient can enter the individual characteristics to predict a patient's likelihood of benefiting from a lumbar fusion procedure. This tool is available online at https://becertain.shinyapps.io/lumbar_fusion_calculator).

ARTICLE INFORMATION

Accepted for Publication: December 28, 2017.

Published Online: March 7, 2018. doi:10.1001/jamasurg.2018.0072

Author Affiliations: Surgical Outcomes Research Center, University of Washington, Seattle (Khor, Lavallee, Flum); Department of Orthopaedic and Sports Medicine, University of Washington, Seattle (Cizik, Bellabarba); Swedish Neuroscience Institute, Swedish Medical Center, Seattle, Washington (Chapman, Oskouian, Roh); Valley Medical Center, Renton, Washington (Howe); Skagit Northwest Orthopedics, Proliance Surgeons, Inc, Mount Vernon, Washington (Lu); Franciscan Neurosurgery Associates at St Joseph, CHI Franciscan Health, Tacoma, Washington (Mohit); Rainier Orthopedic Institute, Proliance Surgeons, Inc, Puyallup, Washington (Shonnard); Department of Anesthesiology and Pain Medicine, University of Washington, Seattle (Dagal); Department of Surgery, University of Washington, Seattle (Flum).

Author Contributions: Ms Khor and Dr Flum had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Khor, Lavallee, Bellabarba, Chapman, Howe, Lu, Mohit, Oskouian, Roh, Shonnard, Flum.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Khor, Flum.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Khor.
Obtained funding: Chapman, Shonnard, Flum.
Administrative, technical, or material support: Khor, Lavallee, Cizik, Shonnard, Dagal, Flum.
Study supervision: Lavallee, Flum.

Conflict of Interest Disclosures: Dr Chapman reports stock ownership in Renovis, a medical device manufacturer. Dr Roh is a consultant and receives royalties from Stryker and has equity in Advanced Biologics and Artoss. Dr Oskouian is a consultant at Globus, Depuy, and Strytker and receives royalties from Globus and Stryker. No other disclosures are reported.

Funder/Sponsor: The project described was supported by the Life Sciences Discovery Fund (grant 4593311) and the Agency for Healthcare Research and Quality (grant R01HS022959).

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: The Spine Surgical Care Outcomes Assessment Program (SCOAP) data metrics and early development was supported by grants from the Agency for Healthcare Research

Quality and the Life Science Discovery Fund to the University of Washington, Seattle. Spine SCOAP is now administered by the Foundation for Healthcare Quality. The Comparative Effectiveness Translational Network (CERTAIN) is an Agency for Healthcare Research Quality–funded research program at the University of Washington (grant R01HS022959).

REFERENCES

1. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis: four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. *J Bone Joint Surg Am*. 2009;91(6):1295-1304.

2. Fritzell P, Hägg O, Wessberg P, Nordwall A; Swedish Lumbar Spine Study Group. 2001 Volvo award winner in clinical studies: lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. *Spine (Phila Pa 1976)*. 2001;26(23):2521-2532.

3. Brox JI, Sørensen R, Friis A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine (Phila Pa 1976)*. 2003;28(17):1913-1921.

4. Fairbank J, Frost H, Wilson-MacDonald J, Yu LM, Barker K, Collins R; Spine Stabilisation Trial Group. Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. *BMJ*. 2005;330(7502):1233-1239.

5. Carreon LY, Glassman SD, Kantamneni NR, Mugavin MO, Djurasovic M. Clinical outcomes after posterolateral lumbar fusion in workers' compensation patients: a case-control study. *Spine (Phila Pa 1976)*. 2010;35(19):1812-1817.

6. The Dr Robert Bree Collaborative. Website. http://www.breecollaborative.org. Published 2016. Accessed January 19, 2018.

7. Dr Robert Bree Collaborative Accountable Payment Models Workgroup. Lumbar fusion surgery bundle. http://www.breecollaborative.org/wp-content/uploads/Lumbar-Fusion-Bundle-Final.pdf. Published September 2014. Accessed January 18, 2018.

8. Veeravagu A, Li A, Swinney C, et al. Predicting complication risk in spine surgery: a prospective analysis of a novel risk assessment tool. *J Neurosurg Spine*. 2017;27(1):81-91.

9. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg*. 2013;217(5):833-42.e1, 3.

10. Lee MJ, Cizik AM, Hamilton D, Chapman JR. Predicting medical complications after spine surgery: a validated model using a prospective surgical registry. *Spine J*. 2014;14(2):291-299.

11. McGirt MJ, Sivaganesan A, Asher AL, Devin CJ. Prediction model for outcome after low-back surgery: individualized likelihood of complication, hospital readmission, return to work, and 12-month improvement in functional disability. *Neurosurg Focus*. 2015;39(6):E13.

12. Alentado VJ, Caldwell S, Gould HP, Steinmetz MP, Benzel EC, Mroz TE. Independent predictors of a clinically significant improvement after lumbar fusion surgery. *Spine J*. 2017;17(2):236-243.

13. McGirt MJ, Bydon M, Archer KR, et al. An analysis from the Quality Outcomes Database, Part 1: disability, quality of life, and pain outcomes following lumbar spine surgery: predicting likely individual patient outcomes for shared decision-making. *J Neurosurg Spine*. 2017;27(4):357-369.

14. Lee MJ, Shonnard N, Farrokhi F, et al; Spine SCOAP-CERTAIN Collaborative. The spine surgical care and outcomes assessment program (spine SCOAP): a surgeon-led approach to quality and safety. *Spine (Phila Pa 1976)*. 2015;40(5):332-341.

15. Flum DR, Alfonso-Cristancho R, Devine EB, et al; CERTAIN Collaborative. Implementation of a "real-world" learning health care system: Washington state's Comparative Effectiveness Research Translation Network (CERTAIN). *Surgery*. 2014;155(5):860-866.

16. Fairbank JC, Pynsent PB. The Oswestry disability index. *Spine (Phila Pa 1976)*. 2000;25(22):2940-2952.

17. Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. *Spine (Phila Pa 1976)*. 2005;30(11):1331-1334.

18. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*. 2001;94(2):149-158.

19. 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(5):598-604.

20. Ghogawala Z, Resnick DK, Watters WC III, et al. Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 2: assessment of functional outcome following lumbar fusion. *J Neurosurg Spine*. 2014;21(1):7-13.

21. Chung AS, Copay AG, Olmscheid N, Campbell D, Walker JB, Chutkan N. Minimum clinically important difference current trends in the spine literature. *Spine (Phila Pa 1976)*. 2017;42(14):1096-1105.

22. Cizik AM, Lee MJ, Martin BI, et al. Using the spine surgical invasiveness index to identify risk of surgical site infection: a multivariate analysis. *J Bone Joint Surg Am*. 2012;94(4):335-342.

23. Mirza SK, Deyo RA, Heagerty PJ, et al. Development of an index to characterize the "invasiveness" of spine surgery: validation by comparison to blood loss and operative time. *Spine (Phila Pa 1976)*. 2008;33(24):2651-2661.

24. Lad SP, Huang KT, Bagley JH, et al. Disparities in the outcomes of lumbar spinal stenosis surgery based on insurance status. *Spine (Phila Pa 1976)*. 2013;38(13):1119-1127.

25. Carreon LY, Glassman SD, Djurasovic M, et al. Are preoperative health-related quality of life scores predictive of clinical outcomes after lumbar fusion? *Spine (Phila Pa 1976)*. 2009;34(7):725-730.

26. Sanderson PL, Todd BD, Holt GR, Getty CJ. Compensation, work status, and disability in low back pain patients. *Spine (Phila Pa 1976)*. 1995;20(5):554-556.

27. Martin BI, Mirza SK, Flum DR, et al. Repeat surgery after lumbar decompression for herniated disc: the quality implications of hospital and surgeon variation. *Spine J*. 2012;12(2):89-97.

28. Jackson KL II, Devine JG. The effects of smoking and smoking cessation on spine surgery: A systematic review of the literature. *Global Spine J*. 2016;6(7):695-701.

29. Pearson A, Lurie J, Tosteson T, Zhao W, Abdu W, Weinstein JN. Who should have surgery for spinal stenosis? Treatment effect predictors in SPORT. *Spine (Phila Pa 1976)*. 2012;37(21):1791-1802.

30. Gulati S, Nordseth T, Nerland US, et al. Does daily tobacco smoking affect outcomes after microdecompression for degenerative central lumbar spinal stenosis?—a multicenter observational registry-based study. *Acta Neurochir (Wien)*. 2015;157(7):1157-1164.

31. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med*. 2007;356(22):2257-2270.

32. Biering K, Hjollund NH, Frydenberg M. Using multiple imputation to deal with missing data and attrition in longitudinal studies with repeated measures of patient-reported outcomes. *Clin Epidemiol*. 2015;7:91-106.

33. Pakarinen M, Vanhanen S, Sinikallio S, et al. Depressive burden is associated with a poorer surgical outcome among lumbar spinal stenosis patients: a 5-year follow-up study. *Spine J*. 2014;14(10):2392-2396.

34. McKillop AB, Carroll LJ, Battié MC. Depression as a prognostic factor of lumbar spinal stenosis: a systematic review. *Spine J*. 2014;14(5):837-846.

35. Ablin JN, Berman M, Aloush V, et al. Effect of fibromyalgia symptoms on outcome of spinal surgery. *Pain Med*. 2017;18(4):773-780. doi:10.1093/pm/pnw232

36. Brummett CM, Urquhart AG, Hassett AL, et al. Characteristics of fibromyalgia independently predict poorer long-term analgesic outcomes following total knee and hip arthroplasty. *Arthritis Rheumatol*. 2015;67(5):1386-1394. doi:10.1002/art.39051

37. Janda AM, As-Sanie S, Rajala B, et al. Fibromyalgia survey criteria are associated with increased postoperative opioid consumption in women undergoing hysterectomy. *Anesthesiology*. 2015;122(5):1103-1111.

38. Schwind J, Learman K, O'Halloran B, Showalter C, Cook C. Different minimally important clinical difference (MCID) scores lead to different clinical prediction rules for the Oswestry disability index for the same sample of patients. *J Man Manip Ther*. 2013;21(2):71-78.