Prediction model for outcome after low-back surgery: individualized likelihood of complication, hospital readmission, return to work, and 12-month improvement in functional disability

Matthew J. McGirt, MD,¹ Ahilan Sivaganesan, MD,² Anthony L. Asher, MD,¹ and Clinton J. Devin, MD³

¹Carolina Neurosurgery & Spine Associates, University of North Carolina School of Medicine, Charlotte, North Carolina; and Departments of ²Neurological Surgery and ³Orthopaedic Surgery, Vanderbilt University, Nashville, Tennessee

OBJECT Lumbar spine surgery has been demonstrated to be efficacious for many degenerative spine conditions. However, there is wide variability in outcome after spine surgery at the individual patient level. All stakeholders in spine care will benefit from identification of the unique patient or disease subgroups that are least likely to benefit from surgery, are prone to costly complications, and have increased health care utilization. There remains a large demand for individual patient-level predictive analytics to guide decision support to optimize outcomes at the patient and population levels.

METHODS One thousand eight hundred three consecutive patients undergoing spine surgery for various degenerative lumbar diagnoses were prospectively enrolled and followed for 1 year. A comprehensive patient interview and health assessment was performed at baseline and at 3 and 12 months after surgery. All predictive covariates were selected a priori. Eighty percent of the sample was randomly selected for model development, and 20% for model validation. Linear regression was performed with Bayesian model averaging to model 12-month ODI (Oswestry Disability Index). Logistic regression with Bayesian model averaging was used to model likelihood of complications, 30-day readmission, need for inpatient rehabilitation, and return to work. Goodness-of-fit was assessed via R^2 for 12-month ODI and via the c-statistic, area under the receiver operating characteristic curve (AUC), for the categorical endpoints. Discrimination (predictive performance) was assessed, using R^2 for the ODI model and the c-statistic for the categorical endpoint models. Calibration was assessed using a plot of predicted versus observed values for the ODI model and the Hosmer-Lemeshow test for the categorical endpoint models.

RESULTS On average, all patient-reported outcomes (PROs) were improved after surgery (ODI baseline vs 12 month: 50.4 vs 29.5%, $p < 0.001$). Complications occurred in 121 patients (6.6%), 108 (5.9%) were readmitted within 30 days of surgery, 188 (10.3%) required discharge to in-patient rehabilitation, 1630 (88.9%) returned to work, and 449 (24.5%) experienced an unplanned outcome (no improvement in ODI, a complication, or readmission). There were 45 unique baseline variable inputs, derived from 39 clinical variables and 38 questionnaire items (ODI, SF-12, MSPQ, VAS-BP, VAS-LP, VAS-NP), included in each model. For prediction of 12-month ODI, R^2 was 0.51 for development and 0.47 for the validation study. For prediction of a complication, readmission, inpatient rehabilitation, and return to work, AUC values ranged 0.72–0.84 for development and 0.79–0.84 for validation study.

CONCLUSIONS A novel prediction model utilizing both clinical data and patient interview inputs explained the majority of variation in outcome observed after lumbar spine surgery and reliably predicted 12-month improvement in physical disability, return to work, major complications, readmission, and need for inpatient rehabilitation for individual patients. Application of these models may allow clinicians to offer spine surgery specifically to those who are most likely to benefit and least likely to incur complications and excess costs.

<http://thejns.org/doi/abs/10.3171/2015.8.FOCUS15338>

KEY WORDS predictive modeling; patient-reported outcomes; registry

ABBREVIATIONS ASA = American Society of Anesthesiologists; AUC = area under the ROC curve; MCID = minimum clinically important difference; MSPQ = Modified Somatic Performance Index; ODI = Oswestry Disability Index; PRO = patient-reported outcome; ROC = receiver operating characteristic; VAS = visual analog scale; VAS-BP = VAS for back pain; VAS-LP = VAS for leg pain; VAS-NP = VAS for neck pain.

SUBMITTED July 8, 2015. **ACCEPTED** August 26, 2015.

INCLUDE WHEN CITING DOI: 10.3171/2015.8.FOCUS15338.

THE Institute of Medicine estimates that 30% of health care expenditures and treatments do not improve patient health outcomes in real-world care.² Even with the adoption of evidence-based medicine, there remain tremendous differences in the safety, effectiveness, and cost of care across US health care systems. Wide variation in outcomes is observed at the individual patient level, despite patients' receiving identical treatments for the same diagnoses. Treatments that fail to yield benefits for patients contribute to the epidemic of ineffective or wasteful care.^{1,6}

Patients, physicians, hospital systems, and third-party payers all aim to identify which patients or disease subgroups are least likely to respond to surgery, are prone to costly complications, and are associated with overutilization of services. Randomized controlled trials are ideally suited to determine whether therapies, on average, provide improved outcome per disease process. However, their high costs and selective patient enrollment prohibit the comprehensive, diverse, and high-volume patient enrollment needed to analyze individual outcomes across the vast spectrum of patient and disease subtypes. In contrast, granular prospective patient-reported outcomes registries that enroll all patients allow for both the statistical power and the multitude of measured risk factors required to appropriately power patient-level prognostic models.

Degenerative spine disease is one of the most prevalent and costly disease states worldwide. In the U.S. alone, the total direct cost of treating low-back pain is estimated at \$100 billion.^{5,9,10,15,17} The rate of spine surgery has increased 40% in the past two decades.^{3,4,13,18} Surgical treatment for the most common lumbar spine diagnoses have been confirmed to be efficacious in several randomized controlled trials.^{11,16,19,20} Nonetheless, its safety and effectiveness have been found to vary widely at the patient-level, with up to 25% of patients experiencing minimal improvement in quality of life and up to 10% experiencing a major complication or hospital readmission.¹² To date, a validated prediction model for lumbar spine surgery has yet to be introduced to provide individualized estimations of the risks and benefits of surgery. These models allow clinicians to offer spine surgery specifically to those who are most likely to benefit and unlikely to incur complications and excess costs. Moreover, such models empower a surgeon to have a substantive personalized discussion with a patient. Applied on a large scale, they have the potential to increase the overall value (benefit/cost) of spine surgery by preventing costly spine care that has a minimal chance of helping the patient. The aims of our study were 1) to develop and measure the performance of novel multivariate prediction models to estimate individual patients' unique risks and benefits of elective lumbar surgery for degenerative spine disease and 2) to validate their performance on a second outcomes data set.

Methods

We conducted a prospective 12-month patient-reported outcomes study spanning 4 years and 1800 consecutive surgical lumbar spine cases performed at a single medi-

cal center (Vanderbilt University Medical Center). Study design focused on extensive preoperative data collection in domains of physical and mental function, quality of life, general health, psychological, social, occupational, and unique health history to establish comprehensive individual preoperative patient profiles.

Registry Methodology

Every consecutive patient undergoing elective lumbar spine surgery over a 4-year period at our center was enrolled into a prospective, Web-based, single-center spine registry, regardless of diagnosis or specific surgical treatment. Registry data were gathered at preoperative, 3-month postoperative, and 12-month postoperative clinic visits. Electronic medical record (EMR) review was used to collect variables such as patient comorbidities, details of the individual surgery, and the perioperative course. Perioperative morbidity, the need for inpatient rehabilitation after surgery, and mortality were assessed via EMR review and confirmed on patient interviews. Questionnaires were administered via one-on-one patient interviews to collect baseline and postoperative patient-reported outcomes. They included measures of pain (visual analog scales [VASs]), and disease-specific physical disability (the Oswestry Disability Index [ODI]). In addition, the mental component of quality of life and somatic perception and anxiety were assessed. By virtue of an institutional review board waiver, verbal consent during the interview was sufficient for data collection from patients. This waiver was granted because, although the registry does constitute a form of research, the outcome assessments do not alter standard care or treatments delivered and therefore were not deemed a significant risk to patients. All clinical data were prospectively collected and entered into the REDCap portal and stored in a HIPAA (Health Insurance Portability and Accountability Act)-secure fashion.

Model Development

Depending on the outcome measure of interest, the sample sizes used for model development ranged from 750 to 1200. This variability stems from the existence of some missing values for the outcomes of interest, which were handled using list-wise deletion (no imputation was performed). These samples were obtained by randomly selecting 80% of the full data to serve as "training" data sets. All baseline input variables were selected a priori based on recognized clinical importance. Linear regression was used to model 12-month ODI after verifying using LOESS fit that a linear relationship with various explanatory variables was reasonable. Because our goal was to optimize prediction while properly accommodating model uncertainty, the linear regression was fitted using Bayesian model averaging.

To model the other 5 categorical outcome variables (complications, readmission, inpatient rehabilitation, return to work, and a composite measure of unplanned outcome), multiple logistic regression with Bayesian model averaging was used. The goodness-of-fit for the ODI model was assessed using R^2 (explained variation). For

the models of categorical endpoints mentioned above, the c-statistic, or the area under the receiver operating characteristic (ROC) curve (AUC), was used. Multi-collinearity is always a concern when multiple regression is performed, however it is less so here because our primary goal is to make predictions on new data (as opposed to understanding the influence of any particular covariate). Multi-collinearity does not generally affect the efficacy of extrapolating a given model to external data.

Model Validation

For each outcome variable, the 20% of the full data that remained after random selection of the training data set was used as the validation data set. Our use of 80% of the data for model development and 20% for model validation is a standard practice in predictive modeling efforts. Each model was used to predict values for its corresponding outcome, using the validation data set, and those predictions were compared with the true observed values. Discrimination, which is one aspect of predictive performance, was assessed using R^2 for the ODI model and using the c-statistic for the categorical endpoint models. Calibration, another aspect of performance, was assessed using a plot of predicted versus observed values for the ODI model and using the Hosmer-Lemeshow test for the categorical endpoint models. All statistical analyses were performed using R, with a significance level of 0.05. Of note, the final performance measures were actually averaged over 10 successive rounds of 80/20 partitions of the data. This is known as 10-fold cross validation, and it has the effect of reducing the variability in the performance estimates that are reported here.

Results

Registry Cohort

A total of 1803 patients undergoing lumbar spine surgery were prospectively enrolled at the time of this study. Table 1 summarizes the baseline characteristics for this cohort, many of which serve as explanatory variables in our predictor models, as well as the incidence of major clinical outcomes, such as morbidity, hospital readmission, reoperation, and disability level at 12 months after surgery. The average age of the patients at the time of surgery was 55.92, nearly half were male, nearly 90% were Caucasian, 44% were employed before surgery, 74% could ambulate without assistance, and their average baseline ODI was 50.44%. Within 30 days of surgery, the mortality rate was 0.3%, the reoperation rate was 4.3%, and the readmission rate was 5.9%. Transfer from the hospital to inpatient rehabilitation or a skilled nursing facility was noted in 11.6% of patients.

Overall, the average ODI 1 year after surgery was significantly improved (50.4% vs 29.5%, $p < 0.05$) and 88.9% of patients returned to work. Fig. 1 displays a Kaplan-Meier curve for the number of days until return to work. Significant ($p < 0.05$) improvements in all PROs were reported 12 months after surgery for all outcome measures: back pain (12-month VAS: 3.5 ± 3.2), leg pain (12-month VAS: 2.8 ± 3.4), disability (12-month ODI: 29.6 ± 20.1) and EQ-5D (12-month QALY [quality-adjusted life year]: 0.7

TABLE 1. Registry cohort characteristics*

Variable	Value
ODI %, mean (SD)	50.44 (16.00)
EQ-5D, mean (SD)	0.54 (0.22)
12-mo ODI %, mean (SD)	29.45 (20.04)
Age in yrs, mean (SD)	55.92 (13.72)
No. of prior spine surgeries	0.79 (1.86)
Preop narcotic use in days, mean (SD)	373.20 (1001.3)
No. of levels Involved, mean (SD)	2.35 (2.12)
BMI	30 (7.38)
MJOAS score, mean (SD)	10.98 (3.28)
SF-12 PCS, mean (SD)	27.77 (9.82)
SF-12 MCS, mean (SD)	46.83 (12.47)
VAS-NP, mean (SD)	2.99 (3.33)
VAS-AP, mean (SD)	2.30 (3.28)
VAS-BP, mean (SD)	5.98 (3.15)
VAS-LP, mean (SD)	5.46 (3.57)
Total ZSDS sum score, mean (SD)	36.67 (10.10)
Total MSPQ sum score, mean (SD)	6.99 (5.13)
Primary diagnosis	
C1–2 pathology	4
Deformity/scoliosis	112
Fracture	68
Herniated disc	472
Infection	8
Pseudarthrosis	117
Spondylolisthesis	313
Spondylolysis	56
Stenosis	597
Tumor	88
Motor deficits	
No	1069
Yes	762
N/A	6
Primary vs revision surgery	
Primary	1313
Revision	521
N/A	3
Minimally invasive vs open	
Minimally invasive	40
Open	1797
Mortality (death w/in 30 days)	
Yes	5
No	1829
Sex	
Female	931
Male	906
Race	
African American	153
Asian	10
Caucasian	1655

(continued)

TABLE 1. Registry cohort characteristics* (continued)

Variable	Value
Race (continued)	
Hispanic	7
Other	12
Employed	
No	1037
Yes	800
Retired vs disabled	
Disabled	359
Other	224
Retired	454
Return to work	
No	32
Yes	758
Ambulatory	
No	33
Yes, w/ assistance	449
Yes, w/o assistance	1355
Duration of symptom(s)	
<3 mos	264
3–12 mos	683
>12 mos	890
Smoker	
Current	434
No	815
Previous	588
Insurance payer	
Medicare/Medicaid	613
Private	974
TennCare	71
Uninsured	30
VA/other gov	124
Predominant symptom	
Myelopathy	369
Neurogenic claudication	220
Cauda equina syndrome	20
Reoperation (w/in 30 days)	
No	1757
Yes	78
ASA grade	
1	46
2	577
3	1125
4	47
N/A	42
History of CAD	
No	1491
Yes	305
N/A	41

(continued)

TABLE 1. Registry cohort characteristics* (continued)

Variable	Value
History of HTN	
No	822
Yes	980
N/A	35
History of MI	
No	1720
Yes	174
N/A	43
History of AFib	
No	1724
Yes	70
N/A	43
History of CHF	
No	1748
Yes	44
N/A	45
History of COPD	
No	1754
Yes	141
N/A	42
History of arthritis	
No	687
Yes	1115
History of diabetes	
No	1414
Yes	382
N/A	41
History of osteoporosis	
No	1741
Yes	54
N/A	42
Readmission (w/in 30 days)	
No	1726
Yes	108

AFib = atrial fibrillation; CAD = coronary artery disease; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; gov = government; HTN = hypertension; MI = myocardial infarction; MJOAS = modified Japanese Orthopaedic Association Scale; N/A = not available; ZSDS = Zung Self-Rating Depression Scale.

* This table provides the mean and standard deviations for continuous variables and the counts for each level for categorical variables captured in our prospective, longitudinal spine registry.

± 0.2). Nevertheless, at the individual patient level, wide variation was seen in PROs at 12 months for all diagnoses captured in the registry, as depicted in Fig. 2. Fourteen percent of patients did not show any improvement in disability (ODI) and 24% did not achieve a minimum clinically important difference (MCID: 15% improvement).¹⁴ Moreover, 449 patients (24.5%) experienced an unplanned outcome after surgery. For purposes of this model development, unplanned outcome was defined as a surgical

FIG. 1. Kaplan-Meier curve for days until return to work. The x-axis on this plot represents the number of days until return to work. The y-axis represents the proportion of patients, among those who were employed before surgery, who had returned to work. Three-quarters of those patients returned to work within a month, and no patients returned to work more than 100 days after surgery.

complication, readmission to the hospital, or lack of any improvement in ODI by 12 months after surgery.

Model Development

For each outcome variable, Table 2 lists the beta coefficients and goodness-of-fit measures for the model development step. Because Bayesian model averaging was employed, the coefficients were weighted by the significance level of their corresponding variables. However, all variables (not just those that are significant) contribute to the overall fit and predictions—the omission of any one variable from the models would reduce predictive performance. Certain predictor variables were selected a priori for mandatory inclusion in the models based on a belief

(based on clinical experience) that they influence the outcomes of interest. Using Table 2, personalized predictions for any outcome can be computed by multiplying a patient's unique value for each explanatory variable by its corresponding coefficient, summing those terms, and then adding the intercept (if one exists). For each categorical variable (all of which are binary), a value of "1" is chosen if patients fit into the category, and a value of "0" is chosen if they do not. The result of this calculation will yield a 12-month ODI score or a predicted probability (categorical outcomes) of complication, hospital readmission, return to work, or a composite unplanned outcome. Please contact the authors for more information on these calculations. Figure 2 graphically represents predictions for 2 hypothetical patients, based on these calculations.

The R^2 goodness of fit of the model was 0.51, suggesting that over half of the observed variance could be explained by the model. No nonrandom patterns were seen in the residual plot for this model. The AUC values ranged 0.72 to 0.84 for complication, readmission, inpatient rehabilitation, and return to work. The AUC for an unplanned postoperative outcome was 0.82.

Model Validation

Table 2 also provides performance measures for the model validation step. An R^2 of 0.47 was achieved for 12-month ODI. The ROC characteristics were even higher in the validation study for complications (0.82), readmission (0.79), need for inpatient rehabilitation (0.84), return to work (0.83), and unplanned outcome (0.78). Logistic regression for each of the 5 categorical outcomes satisfied the Hosmer-Lemeshow goodness-of-fit test ($p > 0.05$). Figure 3 shows a calibration plot of predicted versus observed 12-month ODI values as a demonstration of our model performance. The correlation between predictions and observations is 0.72.

FIG. 2. Variability in patient-reported outcomes (PROs). The top 3 plots show mean values of ODI, SF-12, and EQ-5D (PROs) at baseline, 3 months, and 12 months postoperatively. There is clear improvement in the mean values over time. However, the bottom 3 plots show the individual patient-level variability in PRO change from baseline to 12 months. Each point represents 1 patient—points below the diagonal line signify improvements in ODI and worsening of SF-12 and EQ-5D. Green points represent patients who achieved the MCID, and red points represent those who did not.

TABLE 2. Model coefficients and performance measures*

Predictor	12-Mo ODI	Complication	Readmission	Rehab	RTW	Unplanned Outcome
Baseline ODI	0.17 (0.06)	-0.01 (0.01)	-4.25e-3 (0.01)	-0.01 (0.01)	-2.84e-3 (0.02)	-0.05 (0.01)
Male	0.60 (1.20)	-0.22 (0.25)	-0.28 (0.26)	-0.25 (0.24)	-0.06 (0.40)	-0.08 (0.21)
Caucasian	-2.22 (2.07)	0.32 (0.44)	0.14 (0.45)	-0.34 (0.38)	0.73 (0.58)	-0.02 (0.38)
Age	-0.03 (0.06)	-0.02 (0.01)	-0.02 (0.02)	0.04 (0.01)	-0.01 (0.02)	-0.01 (0.01)
Employed	-3.80 (1.40)	-0.21 (0.30)	-0.28 (0.27)	0.04 (0.30)	—	-0.31 (0.25)
Ambulatory	-5.09 (1.40)	-0.40 (0.27)	-0.10 (0.28)	-0.61 (0.24)	-0.08 (0.57)	0.09 (0.25)
Symptoms >1 yr	0.05 (1.28)	0.14 (0.26)	0.11 (0.28)	-0.58 (0.24)	0.46 (0.40)	0.31 (0.23)
No. of prior surgeries	0.27 (0.32)	-0.04 (0.09)	-0.03 (0.08)	-0.07 (0.05)	-0.07 (0.34)	0.08 (0.05)
Narcotic use duration (days)	2.15e-3 (6.0e-4)	-2.29e-4 (1.6e-4)	-3.3e-4 (1.8e-4)	10e-5 (9.6e-5)	-7.80e-4 (3.5e-4)	8.40e-5 (9.88e-5)
Smoker	0.09 (1.14)	0.53 (0.25)	-0.03 (0.26)	-0.28 (0.22)	-0.05 (0.37)	0.10 (0.20)
Private insurance	-0.01 (1.32)	-0.40 (0.28)	-0.12 (0.29)	-0.53 (0.26)	0.20 (0.49)	-0.12 (0.23)
Acute back pain	-18.02 (7.23)	-1.11 (1.33)	-1.01 (1.10)	-2.95 (1.12)	0.45 (2.08)	-0.97 (1.29)
Acute leg pain	-13.37 (5.55)	-1.82 (1.47)	-0.09 (1.04)	-1.46 (1.08)	0.20 (1.60)	-0.39 (1.08)
Chronic back pain	-3.27 (4.87)	0.93 (1.11)	-1.32 (0.94)	-2.67 (0.87)	1.30 (1.60)	-0.72 (0.92)
Chronic leg pain	-6.20 (4.83)	-1.54 (1.10)	-1.19 (0.93)	-1.81 (0.83)	0.24 (1.66)	-0.82 (0.90)
Myelopathy	-1.33 (3.98)	-1.10 (0.83)	0.20 (0.64)	-0.65 (0.59)	1.70 (1.41)	-0.20 (0.69)
Claudication	3.04 (1.53)	0.28 (0.32)	-0.04 (0.34)	-0.20 (0.29)	0.87 (0.53)	-0.07 (0.26)
Cauda equina	-3.93 (5.44)	-15.64 (9.9e+2)	-1.65 (1.60e+3)	-0.81 (0.96)	0.24 (1.66)	-14.50 (4.47e+2)
ASA grade >2	0.57 (1.32)	-0.23 (0.29)	0.71 (0.33)	0.05 (0.28)	-0.24 (0.41)	-0.20 (0.23)
History of CAD	-0.14 (1.70)	0.01 (0.35)	-0.29 (0.37)	-0.35 (0.29)	-0.25 (0.66)	0.22 (0.28)
History of HTN	2.07 (1.27)	0.17 (0.28)	0.35 (0.30)	0.36 (0.25)	-0.06 (0.41)	0.08 (0.22)
History of MI	-5.00 (3.30)	0.45 (0.54)	1.35 (0.54)	-0.17 (0.52)	-1.01 (0.98)	0.03 (0.54)
History of AFib	10.79 (3.16)	1.05 (0.47)	0.99 (0.47)	0.81 (0.41)	17.59 (1.5e+3)	0.14 (0.53)
History of CHF	-9.49 (4.13)	0.42 (0.61)	-0.93 (0.87)	0.10 (0.59)	19.74 (2.1e+3)	0.10 (0.70)
History of COPD	9.62 (4.39)	-0.35 (0.85)	-0.40 (0.83)	-0.36 (0.87)	-1.92 (1.47)	-0.09 (0.71)
History of arthritis	0.64 (1.31)	-0.07 (0.28)	-0.41 (-0.29)	-0.32 (0.27)	-0.77 (0.42)	-0.29 (0.23)
History of diabetes	1.94 (1.45)	0.35 (0.29)	0.58 (0.30)	0.47 (0.25)	1.28 (0.66)	0.26 (0.25)
History of osteoporosis	-0.80 (3.55)	0.12 (0.60)	1.17 (0.50)	0.53 (0.46)	17.74 (3.7e+3)	0.94 (0.55)
BMI	0.08 (0.09)	0.02 (0.02)	-0.02 (0.02)	0.02 (0.02)	-0.02 (0.03)	0.04 (0.02)
SF-12 Physical	-0.13 (0.08)	0.02 (0.02)	0.01 (0.02)	-2.45e-3 (0.01)	2.5e-3 (0.02)	-8.17e-3 (1.37e-2)
SF-12 Mental	-0.06 (0.07)	0.00 (0.00)	0.008 (0.02)	9.26e-4 (0.01)	-0.02 (0.02)	-4.40e-3 (9.53e-3)
Neck pain score	0.84 (0.26)	-0.07 (0.06)	-0.07 (0.06)	-0.12 (0.05)	-0.05 (0.09)	-0.01 (0.04)
Arm pain score	0.81 (0.28)	0.04 (0.06)	0.07 (0.06)	0.04 (0.06)	-0.02 (0.08)	0.08 (0.05)
Back pain score	0.34 (0.24)	0.05 (0.06)	0.08 (0.06)	0.10 (0.05)	-2.87e-4 (0.01)	0.13 (0.15)
Somatic perception (MSPQ)	0.61 (0.17)	4.04e-3 (0.03)	-0.01 (0.04)	-0.02 (0.03)	0.06 (0.06)	0.05 (0.03)
Motor deficit	-0.59 (1.23)	0.36 (0.26)	-0.52 (0.29)	0.18 (0.24)	0.40 (0.43)	-0.08 (0.22)
Primary (vs revision) surgery	-4.59 (1.4)	-0.03 (0.30)	-0.09 (0.32)	0.22 (0.27)	-0.31 (0.51)	-0.23 (0.24)
Spinal deformity	-3.18 (3.55)	-0.11 (0.67)	0.49 (0.68)	1.08 (0.58)	4.20 (1.94)	0.25 (0.61)
Fracture	4.68 (4.41)	-0.18 (0.76)	0.56 (0.71)	-0.13 (0.69)	0.04 (1.10)	0.99 (0.72)
Disc herniation	-9.28 (3.04)	-0.09 (0.58)	-0.07 (0.61)	-0.38 (0.57)	1.75 (0.84)	-0.55 (0.52)

(continued)

TABLE 2. Model coefficients and performance measures* (continued)

Predictor	12-Mo ODI	Complication	Readmission	Rehab	RTW	Unplanned Outcome
Pseudarthrosis	-2.46 (3.69)	0.56 (0.66)	0.54 (0.71)	0.05 (0.64)	1.07 (1.24)	0.24 (0.60)
Spondylolisthesis	-9.33 (3.12)	-0.48 (0.60)	-0.34 (0.63)	0.19 (0.52)	1.44 (0.85)	-0.62 (0.53)
Spondylosis	-1.10 (5.30)	-0.45 (1.18)	-1.56 (1.36e+3)	0.73 (0.96)	0.88 (1.34)	-1.71 (1.17)
Stenosis	-7.27 (3.00)	0.10 (0.56)	0.49	-0.21 (0.49)	1.20 (0.83)	-0.14 (0.50)
Fusion	2.08 (1.47)	0.27 (0.30)	-0.23 (0.31)	0.31 (0.28)	-0.63 (0.43)	0.34 (0.26)
Goodness of fit (development)	R ² = 0.51	AUC = 0.72	AUC = 0.74	AUC = 0.84	AUC = 0.79	AUC = 0.82
Performance (validation)	R ² = 0.47	AUC = 0.82	AUC = 0.79	AUC = 0.84	AUC = 0.83	AUC = 0.78

BMI = body mass index; rehab = rehabilitation; RTW = return to work; — = not applicable.

* The entries in the stub column (far left) are the various predictor variables in our models and the data column heads are the outcome variables. Each cell displays the mean (SD) of the beta coefficients (weighted averages via Bayesian model averaging) for the corresponding predictor and outcome variable. For a given patient, multiply the values for each predictor variable by the coefficient, sum the terms, and then add the intercept (if one exists) to calculate the predicted outcome of interest. Please contact the authors for more information on these calculations.

Discussion

Robust Predictive Models

It has been established that considerable patient-level variability exists in clinical outcomes after elective spine surgery. In light of this, we have introduced models which use a large array of baseline patient characteristics to predict ODI at 12 months, postoperative complications, hospital readmission, need for inpatient rehabilitation, and return to work after surgery. Linear regression and Bayesian model averaging were used to develop the models, based on data from a single center spine registry. The models explain approximately one-half of the variation in 12-month ODI and achieve diagnostic accuracy ranging from 72%–84% for surgical morbidity and return to work. Using standard measures of discrimination and calibration, we have affirmed the predictive performance of the models. They are a significant advancement relative to models published

by Lee and colleagues, which have lower ROC characteristics and only predict 2 outcomes—surgical-site infections and medical complications.^{7,8} Our models' performance measures are similar, however, to that of recently published American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) models for complications after spine surgery. The ACS NSQIP models are generated from a much larger national data set, but they do not address any patient-reported outcomes.

Our decision to employ Bayesian model averaging stems from the fact that traditional modeling techniques require the selection of one model from a set of models that often have similar levels of goodness-of-fit. There is a layer of statistical uncertainty inherent in that selection that is routinely ignored. Bayesian model averaging addresses this uncertainty by creating weighted-average predictions over a whole set of possible models, which has the potential to improve overall predictive performance.

FIG. 3. Calibration plot. The y-axis represents observed values for 12-month ODI in our registry, while the x-axis represents the corresponding predicted values for 12-month ODI based on our regression models. The dotted line has a 45° slope through the origin and is where vertical and horizontal values are equal.

To illustrate how models such as ours could one day be applied in a clinical setting, Fig. 4 presents the predicted outcomes for 2 hypothetical patients whom a neurosurgeon might see in clinic. The patients have differing baseline characteristics, as seen in the panel on the left, and the panel on the right shows that they have divergent predictions for every outcome of interest. A Web- or phone-based application that generates graphical predictions such as this, after querying patients about preoperative factors in the clinic waiting room, would allow neurosurgeons and any other spine care providers to create comprehensive risk profiles that are tailored to specific patients. These equations can easily be applied to a calculator or Web-based tool for application in practice. In reality, decisions should not be made solely on evidence-based historical outcomes in similar or matched patients as this model/tool provides, but rather based on the combination of experience, real-time interpersonal judgment, and decision support tools like this. It can draw the surgeon’s attention to outliers of particular risk they may have gone unappreciated in everyday practice. It is meant to supplement the knowledge of surgeon and patient in the decision-making process at the individualized level.

Influence of Preoperative Factors

In addition to generating predictions of clinical outcomes after spine surgery, our modeling efforts also reveal which baseline factors are most influential in relation to those outcomes. A higher preoperative ODI, a longer duration of narcotic use, hypertension, atrial fibrillation, more

severe extremity pain, and depression/anxiety are significantly associated with a higher 12-month ODI (less effective care). In contrast, preoperative employment, ability to walk, acute back or leg pain, better general health (SF-12 PCS), a diagnosis of disc herniation, and the absence of prior spine surgery, are associated with a lower 12-month ODI (more effective care). These findings correlate with clinical intuition—patients who have less pain, fewer comorbidities, and better functional status at baseline tend to have less disability after surgery. Patients with acute-onset pain may do better than those with chronic symptoms, because the cause of the pain is more likely due to a structural pathology that is successfully correctable with surgery. Those without prior surgeries also tend to do better because revision surgeries can be more technically difficult, have pain generators that are not surgically correctable, and involve more complications.

Our predictive models demonstrate that a history of smoking, atrial fibrillation, a higher ASA (American Society of Anesthesiologists) grade, a history of myocardial infarction, diabetes, or osteoporosis are associated with a higher risk of postoperative complications and adverse events. This is understandable, as higher ASA grades represent physical status that is less “fit for surgery,” and the other comorbidities have previously been demonstrated to raise the risk of intra- and postoperative adverse events. Osteoporosis, in particular, can impede bony stability and fusion after surgery, potentially leading to hardware failure and possible reoperations. Furthermore, we found that the ability to ambulate before surgery, private insurance,

FIG. 4. Divergent predictions for 2 hypothetical cases. Here we present 2 hypothetical patients with differing baseline characteristics, as seen in the panel on the left. The panel on the right shows that the 2 patients have divergent predictions of 12-month ODI as well as the probabilities of complication, readmission, need for inpatient rehabilitation, return to work, and unplanned outcome based on our regression models. Predictions such as these can help surgeons understand benefits and risks of surgery that are specific to an individual patient.

and acute versus chronic back or leg pain are associated with a reduced likelihood of the need for inpatient rehabilitation after surgery. In contrast, increasing age, increasing number of operated levels, a history of atrial fibrillation or arthritis, and a diagnosis of spinal deformity are associated with an increased likelihood of being discharged to inpatient rehabilitation. Deformity surgeries can be some of the most complex and extensive procedures performed in hospitals, involving considerable blood loss and longer lengths of stay, which explains the higher probability of inpatient rehabilitation. Finally, for working patients without workers' compensation claims who are planning on returning to work after surgery, a preoperative diagnosis of depression, arthritis, and prolonged preoperative opioid use significantly reduced the likelihood of returning to work. Patients with a diagnosis of spondylolisthesis or disc herniation are more likely to return to work. These findings are understandable, given that patients with mood difficulties or long-term opioid consumption may have reduced functional status and/or motivation that prevents reemployment.

Limitations of Study

There are limitations to consider when interpreting these results. First, an R^2 of 0.51 was achieved for 12-month ODI during model development. This means that our model explains only half the variability in 12-month ODI, leaving room for improvement. This is not surprising, given the challenge of modeling a continuous variable like ODI that has 100 possible values, but it highlights the fact that our model is not yet complete. Inclusion of additional predictor variables in our registry, as well as consideration of nonlinear regression strategies, will likely be necessary to increase predictive performance. The model in its current form, however, is significant in that it represents the first known attempt to predict disability 1 year after elective lumbar surgery, and in doing so provides a set of preoperative factors that are influential in the outcome.

Another limitation is that while our models perform well at predicting observed categorical outcomes when applied to an internal validation data set, their performance remains untested in an external data set from a different medical center. The models presented here are built solely on data from a single-institution registry—one that reflects a particular patient population and a particular set of surgeons and techniques. Therefore, it is unclear how applicable our models may be outside of the Vanderbilt University Medical Center, and a multi-institutional prospective study is needed to investigate this. It is quite possible a patient at another medical center, with a similar risk profile to a patient in the Vanderbilt registry, may have outcomes that differ from the predictions generated by our models. However, we expect that the relative position of the Vanderbilt patient will likely be equivalent to the relative position of the external patient when compared with their respective cohorts. In other words, if a Vanderbilt patient is predicted to be in the top quantile of outcomes in our registry, a similar patient elsewhere will likely also be in that same quantile. In this manner, our models should provide some insight into general risk stratification of patients outside Vanderbilt.

Because large longitudinal data sets that include the breadth and depth of patient-reported physical and mental status such as this are rare, it may be difficult to effectively test these models in currently existing registries or research data sets. Additionally, the data inputs that power these models are fairly involved and require the focused attention of the patient for at least 20 minutes with 45 unique baseline variable inputs derived from 39 clinical variables and 38 questionnaire items (ODI, SF-12, MSPQ, VAS-BP, VAS-LP, VAS-NP). Therefore, reproducing the predictive accuracy reported here may be too time intensive and cost prohibitive. The 39 basic clinical variables, such as age, demographics, comorbidities, diagnosis, extent of surgery, and others, can be derived from the electronic medical record. However, the ODI, SF-12, MSPQ, and pain VAS scores are patient-reported data inputs that need to be collected via 38 questionnaire items. Future efforts to create “abridged” models that achieve similar predictive performance while requiring less time-intensive data entry will be important. However, it is important to establish the potential that truly granular modeling can have in forecasting long-term outcomes.

Conclusions

The predictive models we present here may have tremendous value as real-world decision support tools for patients, providers, hospital systems, and payers alike. Patients are now empowered to sit down with their physician and have a concrete discussion about expectations after surgery that is tailored to their particular risk profile. The discussion can now center around the issues of greatest importance to the patient, such as pain-related disability, potential for return to work, and risk of a complication. Compared with current standards, this represents a higher level of shared, informed, and individualized decision making. Surgeons may also benefit significantly. The mark of a great surgeon has always been the ability to choose the right intervention, at the right time, for the right patient. However, this has largely been an “art” developed only through individual experience, rather than a reproducible or scalable process. Our models bridge this gap, enabling surgeons to make decisions about surgery in a way that is systematic, data-driven, and optimized to each individual patient. Consistent use of predictive models may also facilitate practice-based learning—surgeons will be able to improve patient satisfaction and reduce the rate of adverse events.

Hospital systems stand to benefit financially as well. They can use the models to gain insight into the probable effectiveness and costs of surgical spine care for their particular patient populations. This gives them a strong position when entering into reimbursement negotiations with payers or when navigating their risks with constructing capitated or bundled payment services. With the advent of value-based purchasing and capitated payment, hospitals will be able to design “smart” bundles based on a unique understanding of their patients. Over time, this can also cause an evolution in their relationship with payers. Whereas payers currently play a paternalistic role, using blunt clinical metrics to measure and approve surgical ser-

vices, these models allow for specialty-specific and risk-adjusted policies and payment that reward high-value care.

Here we present 5 novel models, developed and validated using a comprehensive, prospective, longitudinal spine registry, which provide predictions for a patient's disability level 1 year after lumbar spine surgery as well the probability of complications, hospital readmission, the need for inpatient rehabilitation, and return to work. These models can provide patients, physicians, payers, and hospital systems with decision support tools that, if used consistently, have the potential to increase the overall value of spine surgery. By providing insight into a patient's postoperative course, in terms that are meaningful to each particular patient, our models give true meaning to the notion of personalized medicine in spine surgery.

References

1. Chan CW, Peng P: Failed back surgery syndrome. **Pain Med** **12**:577–606, 2011
2. Committee on the Learning Health Care System in America: **Best Care at Lower Cost: The Path to Continuously Learning Health Care in America**. Washington, DC: National Academies Press, 2012
3. Dagenais S, Caro J, Haldeman S: A systematic review of low back pain cost of illness studies in the United States and internationally. **Spine J** **8**:8–20, 2008
4. Deyo RA, Gray DT, Kreuter W, Mirza S, Martin BI: United States trends in lumbar fusion surgery for degenerative conditions. **Spine (Phila Pa 1976)** **30**:1441–1447, 2005
5. Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC, Jarvik JG: Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. **JAMA** **303**:1259–1265, 2010
6. Emanuel EJ, Fuchs VR: The perfect storm of overutilization. **JAMA** **299**:2789–2791, 2008
7. Lee MJ, Cizik AM, Hamilton D, Chapman JR: Predicting medical complications after spine surgery: a validated model using a prospective surgical registry. **Spine J** **14**:291–299, 2014
8. Lee MJ, Cizik AM, Hamilton D, Chapman JR: Predicting surgical site infection after spine surgery: a validated model using a prospective surgical registry. **Spine J** **14**:2112–2117, 2014
9. Luo X, Pietrobon R, Sun SX, Liu GG, Hey L: Estimates and patterns of direct health care expenditures among individuals with back pain in the United States. **Spine (Phila Pa 1976)** **29**:79–86, 2004
10. Lurie JD, Weinstein JN: Shared decision-making and the orthopaedic workforce. **Clin Orthop Relat Res** **(385)**:68–75, 2001
11. Lurie JD, Tosteson TD, Tosteson AN, Zhao W, Morgan TS, Abdu WA, et al: Surgical versus nonoperative treatment for lumbar disc herniation: eight-year results for the spine patient outcomes research trial. **Spine (Phila Pa 1976)** **39**:3–16, 2014
12. McGirt MJ, Spermoff T, Dittus R, Harrell F, Asher A: The National Neurosurgery Quality and Outcomes Database (N²QOD): general overview and pilot-year project description. **Neurosurg Focus** **34**(1):E6, 2013
13. North RB, Kidd DH, Zahurak M, James CS, Long DM: Spinal cord stimulation for chronic, intractable pain: experience over two decades. **Neurosurgery** **32**:384–395, 1993
14. Parker SL, Adogwa O, Paul AR, Anderson WN, Aaronson O, Cheng JS, 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** **14**:598–604, 2011
15. Rihn JA, Currier BL, Phillips FM, Glassman SD, Albert TJ: Defining the value of spine care. **J Am Acad Orthop Surg** **21**:419–426, 2013
16. Tosteson AN, Tosteson TD, Lurie JD, Abdu W, Herkowitz H, Andersson G, et al: Comparative effectiveness evidence from the spine patient outcomes research trial: surgical versus nonoperative care for spinal stenosis, degenerative spondylolisthesis, and intervertebral disc herniation. **Spine (Phila Pa 1976)** **36**:2061–2068, 2011
17. Waterman BR, Belmont PJ Jr, Schoenfeld AJ: Low back pain in the United States: incidence and risk factors for presentation in the emergency setting. **Spine J** **12**:63–70, 2012
18. Weinstein JN, Lurie JD, Olson PR, Bronner KK, Fisher ES: United States' trends and regional variations in lumbar spine surgery: 1992–2003. **Spine (Phila Pa 1976)** **31**:2707–2714, 2006
19. Weinstein JN, Lurie JD, Tosteson TD, Zhao W, Blood EA, Tosteson AN, 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** **91**:1295–1304, 2009
20. Weinstein JN, Tosteson TD, Lurie JD, Tosteson A, Blood E, Herkowitz H, et al: Surgical versus nonoperative treatment for lumbar spinal stenosis four-year results of the Spine Patient Outcomes Research Trial. **Spine (Phila Pa 1976)** **35**:1329–1338, 2010

Disclosures

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

Author Contributions

Conception and design: McGirt. Acquisition of data: McGirt, Devin. Analysis and interpretation of data: Sivaganesan, Devin. Drafting the article: McGirt, Sivaganesan, Devin. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: McGirt. Statistical analysis: Sivaganesan.

Correspondence

Matthew J. McGirt, Carolina Neurosurgery & Spine Associates, University of North Carolina School of Medicine, 225 Baldwin Ave., Charlotte, NC 28204. email: matt.mcgart@cnsa.com.