

Prediction Models in Degenerative Spine Surgery: A Systematic Review

Daniel Lubelski, MD¹ , Andrew Hersh, BA¹ , Tej D. Azad, MD, MS¹, Jeff Ehresman, BS¹, Zachary Pennington, BS¹ , Kurt Lehner, MD¹, and Daniel M. Sciubba, MD¹ 

Global Spine Journal
2021, Vol. 11(1S) 79S-88S
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2192568220959037
journals.sagepub.com/home/gsj



Abstract

Study Design: Systematic review.

Objectives: To review the existing literature of prediction models in degenerative spinal surgery.

Methods: Review of PubMed/Medline and Embase databases was conducted to identify articles between January 1, 2000 and March 1, 2020 that reported prediction model performance for outcomes following elective degenerative spine surgery.

Results: Thirty-one articles were included. Twenty studies were of thoracolumbar, 5 were of cervical, and 6 included all spine patients. Five studies were externally validated. Prediction models were developed using machine learning (42%) and logistic regression (42%) as well as other techniques. Web-based calculators were included in 45% of published articles. Various outcomes were investigated, including complications, infection, length of stay, discharge disposition, reoperation, readmission, disability score, back pain, leg pain, return to work, and opioid dependence.

Conclusions: Significant heterogeneity exists in methods used to develop prediction models of postoperative outcomes after degenerative spine surgery. Most internally validate their scores, but a few have been externally validated. Areas under the curve for most models range from 0.6 to 0.9. Techniques for development are becoming increasingly sophisticated with different machine learning tools. With further external validation, these models can be deployed online for patient, physician, and administrative use, and have the potential to optimize outcomes and maximize value in spine surgery.

Keywords

degenerative, degenerative disc disease, cervical, lumbar

Introduction

Value-based care has become a manifest focus of American health care policy and is driven by efforts to improve outcomes while reducing costs. Hospital systems and policy makers continue to explore methods to reduce complications, improve patient education, and increase efficiency in perioperative and postoperative settings. Given substantial variability between surgeons in the indications and interventions used for given degenerative spinal pathologies, there is commensurate variability in outcomes.¹⁻⁴ Randomized controlled trials (RCTs) remain the gold standard for determining the efficacy of an intervention and a small number have been conducted for the management of degenerative spine pathology.⁵⁻⁸ However, RCTs of surgical interventions have inherent challenges^{9,10} and cannot be performed for every clinical question. Cost and comparative effectiveness studies have emerged as an alternative to

identify operations that are more likely to yield high value outcomes. Another burgeoning approach is the development and validation of clinical prediction models.

Predictive analytics in clinical medicine has been enabled by the rapid adoption of electronic medical records, development of national registries and prospective multicenter databases, and increased awareness of machine and statistical learning methods. Clinical prediction models have the potential to provide patient-specific risk profiles and expected outcomes.

¹ Johns Hopkins University School of Medicine, Baltimore, MD, USA

Corresponding Author:

Daniel M. Sciubba, Department of Neurosurgery, Johns Hopkins University School of Medicine, 600 North Wolfe Street, Meyer 5-185A, Baltimore, MD 21287, USA.
Email: dsciubb1@jhmi.edu



Creative Commons Non Commercial No Derivs CC BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-Non Commercial-NoDerivs 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

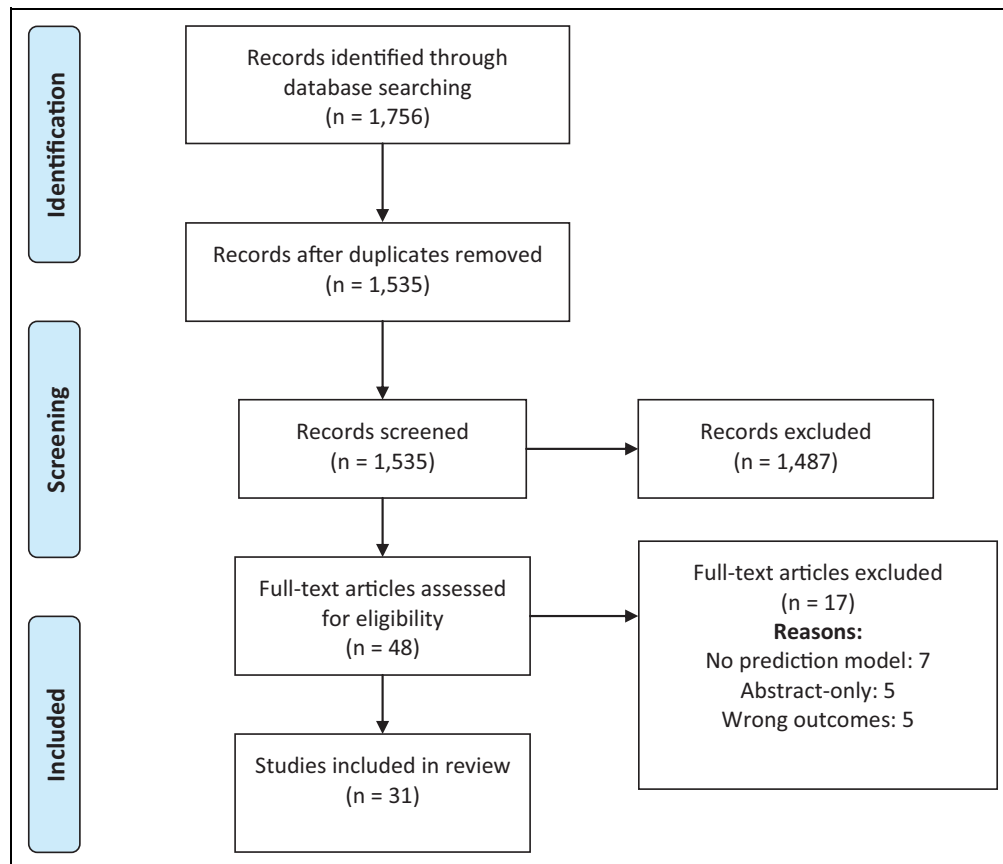


Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for articles with degenerative spine disease prediction models with 1-year outcomes after surgery.

With these tools, surgeons may be able to give a patient their expected likelihood of success for a given operation, as well as their chance for adverse outcomes and complications. On a hospital-wide and national level, these tools can help identify targets for quality improvement efforts and policy making.

Given the demonstrated variability in degenerative spinal surgery practice and outcomes, the application of more robust prediction models to this field may lead to substantial improvements in patient care. However, the studies of prediction model development for degenerative spinal surgery have been heterogeneous. These articles have focused on postoperative outcomes, length of stay (LOS), discharge disposition, and adverse events. They have also varied in terms of design, sample size, method of validation, and mode of deployment. The goal of this systematic review was to summarize the existing literature on prediction models in degenerative spinal surgery. We categorized the existing degenerative spinal surgery prediction models based on their respective outcomes and design and report the relative strengths and weaknesses of these studies to aid in interpretation and consideration for clinical deployment.

Methods

We performed a search of the English language literature using the PubMed/Medline and Embase databases to identify articles

between January 1, 2000 and March 1, 2020 that reported prediction model performance for outcomes following elective degenerative spine surgery.

Search terms included (prediction OR predictive) AND (spine OR spinal OR “spine surgery” OR “laminectomy” OR “interbody fusion” OR “discectomy” OR “discectomy” OR “spinal fusion”). We further queried the bibliographies of the included studies to identify additional relevant articles.

Inclusion criteria were English language articles involving adult patients who underwent elective spine surgery for a degenerative spinal pathology. Studies involving tumor, infection, and deformity were excluded, as were nonclinical studies. All studies were required to have a description of a model that could facilitate inputting patient-level data to predict the outcome of interest. Prediction model outcomes could include functional/disability/pain scores or more objective measures such as LOS, reoperation, readmission, and complications.

Results

We identified 1535 unique articles (Figure 1), of which 48 underwent full-text review leading to the inclusion of 31 articles in this review. Reasons for exclusion included no mention of a prediction model ($n = 7$), outcomes not fitting inclusion criteria ($n = 5$), and only abstract available ($n = 5$). Of these 31

Table 1. Studies Evaluating Complications During/After Spine Surgery.

Author, year	Institutions	Design	Time length	Sample size	Internal AUC	Calib?	Internal validation	External validation	Calc
Lee, 2014 ¹²	Single	Retrospective	2 years	1476	Overall: 0.76 Major: 0.81	Yes	Random number generator	Overall ⁵⁰ : 0.71 Major: 0.85	Yes
McGirt, 2015 ¹¹	Single	Prospective	1 year	1803	Overall: 0.72	Yes	Training/validation	No	No
Ratliff, 2016 ¹⁴	Multiple	Retrospective	30 days	279 315	Overall: 0.70 Pulmonary: 0.72	No	Training/validation	Veeravagu ⁴² :0.67	Yes
Kim, 2018 ¹³	Multiple	Retrospective	30 days	22 629	Cardiac: 0.71 VTE: 0.57 Infection: 0.61	No	Training/validation	No	No
Han, 2019 ¹⁵	Multiple	Retrospective	30 days	1 106 234	Overall: 0.70	Yes	Training/validation	No	No
Janssen, 2019 ⁴¹	Single	Retrospective	>1 year	898	Infection: 0.72	Yes	Bootstrapping	No	No

Abbreviations: VTE, venous thromboembolism; AUC, area under the curve; Calib?, calibration performed?; Calc, whether the authors reported that they developed a Web-based calculator.

Table 2. Prediction Models for Reoperation and Readmission After Spine Surgery.

Author, year	Institutions	Design	Time length	Sample size	Internal AUC	Calib?	Internal validation	External validation	Calc
McGirt, 2015 ¹¹	Single	Prospective	30 days	1803	Readmit 0.74	Yes	Training/validation	No	No
Lubelski, 2017 ¹⁷	Single	Retrospective	90 days	952	Reop 0.91 Readmit 0.78	Yes	Bootstrapping	No	Yes
Goyal, 2019 ¹⁸	Multi	Retrospective	30 days	59 145	Readmit 0.66	No	Training/validation	No	No
Hopkins, 2019 ¹⁹	Multi	Retrospective	30 days	23 264	Readmit 0.81	No	Training/validation	No	No
Siccoli, 2019 ¹⁶	Single	Retrospective	1 year	635	Reop 0.63	Yes	Training/validation	No	No

Abbreviations: Reop, reoperation; Readmit, readmission; AUC, area under the curve; Calib?, calibration performed?; Calc whether the authors reported that they developed a Web-based calculator.

articles, 5 articles (16%) included external validation. Of the 31 included studies, 20 (65%) were of thoracolumbar surgeries, 5 (16%) were cervical surgeries, and 6 (19%) were inclusive of patients undergoing any spinal surgery.

There was heterogeneity in how the prediction models were developed. Thirteen (42%) used machine learning, 13 (42%) used logistic regression, 2 (6%) used linear regression, 1 (3%) used binomial regression, 1 (3%) used both logistic and linear regression, and 1 (3%) used Cox proportional hazards regression. For internal validation, 17 (55%) used cross-validation by splitting their cohort into a training and validation sets, 9 (29%) used bootstrapping, 1 (3%) used random number generators, and 4 (13%) did not specify. Web-based calculators were included in 14 (45%) of the published articles. Various outcomes were investigated, including overall complications, infection, LOS, discharge disposition, reoperation, readmission, Oswestry Disability Index (ODI) score, back pain, leg pain, return to work, and opioid dependence.

Six articles looked at complications (Table 1), which included infection (n = 3), all-inclusive complications (n = 4), pulmonary complications (n = 2), cardiac complications (n = 2), venous thromboembolism (n = 1), and neurologic complications (n = 1).^{11-15,41} Of these, 3 were single institution studies, 1 used the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP)

database, 1 used the Truven Health Analytics MarketScan database, and 1 used both the Truven database and the Centers for Medicare and Medicaid Services (CMS) Medicare database. One was prospective while 5 were retrospective. Study follow-up ranged from 30 days to 2 years. Area under the curve (AUC) ranged from 0.57 to 0.72.

Reoperation (n = 2) and readmission (n = 4) were examined by 5 articles (Table 2).^{11,16-19} Three were single institution studies and 2 used the ACS-NSQIP database. One was prospective while 4 were retrospective. Study follow-up ranged from 30 days to 1 year. AUC ranged from 0.63 to 0.91.

Nine studies examined the LOS and discharge disposition of patients (Table 3).^{11,16,18,20-25} Of these, 2 examined discharge to a rehabilitation facility, 1 examined discharge to any facility, 5 examined nonhome discharge, and 2 examined prolonged LOS. Three were single institution, 5 used the ACS-NSQIP database, and 1 used the NeuroPoint Quality Outcomes Database (QOD) database. One was prospective while 8 were retrospective. AUC ranged from 0.75 to 0.89.

Eighteen articles examined functional outcomes (Table 4), which included quality-of-life measures (n = 11), opioid dependence (n = 3), returning to work (n = 2), patient satisfaction (n = 2), and persistent postsurgical pain (n = 1).^{11,16,17,26-40} Quality-of-life outcome measures included scores on the following validated inventories: ODI, visual analog scale for

Table 3. Prediction Models for Length of Stay and Discharge of Patients Undergoing Spine Surgery.

Author, year	Institutions	Design	Sample size	Internal AUC	Calib?	Internal validation	External validation	Calc
McGirt, 2015 ¹¹	Single	Prospective	1803	Rehab: 0.84	Yes	Training/validation	No	No
Guan, 2018 ²⁵	Multi	Retrospective	217	Nonhome disch: 0.80	Yes	N/A	No	No
Karhade, 2018 ²²	Multi	Retrospective	26 364	Nonhome disch: 0.82	Yes	Training/validation	Stopa ⁴⁵ :0.89	Yes
Goyal, 2019 ¹⁸	Multi	Retrospective	59 145	Nonhome disch: 0.87	No	Training/validation	No	No
Ogink, 2019 ²³	Multi	Retrospective	9338	Nonhome disch: 0.75	Yes	Training/validation	No	Yes
Ogink, 2019 ²⁴	Multi	Retrospective	28 600	Nonhome disch: 0.75	Yes	Training/validation	No	Yes
Siccoli, 2019 ¹⁶	Single	Retrospective	635	Prolonged LOS: 0.77	Yes	Training/validation	No	No
Harada, 2020 ²¹	Multi	Retrospective	10 453	Facility disch: 0.75	Yes	Training/validation	Harada ²¹ : 0.77	No
Lubelski, 2020 ²⁰	Single	Retrospective	257	Rehab: 0.89 Prolonged LOS: 0.89	No	Bootstrapping	No	Yes

Abbreviations: Disch, discharge; Rehab, inpatient rehabilitation; AUC, area under the curve; Calib?, calibration performed;? Calc, whether the authors reported that they developed a Web-based calculator.

leg and lower back pain, EuroQol 5-dimensions (EQ-5D), Patient Health Questionnaire-9 (PHQ-9), Pain and Disability Questionnaire (PDQ), Short Form 6-dimensions (SF-6D), and the modified Japanese Orthopedic Association (mJOA). Seven were single institution studies, 5 used the QOD database, and 6 were multi-institutional. Four were prospective while 14 were retrospective. Follow-up ranged from 90 days to 2 years. AUC ranged from 0.64 to 0.81. Specifically, AUC ranged from 0.64 to 0.81 for quality-of-life measures, 0.70 to 0.80 for opioid dependence, 0.71 to 0.81 for return to work, and 0.64 to 0.79 for patient satisfaction. AUC was 0.66 for persistent postsurgical pain.

Discussion

We identified 31 studies reporting prediction models for degenerative spinal surgery. These have mainly focused on predicting complications, readmission, reoperation, and functional/quality-of-life outcomes. We found that while almost all studies attempted to internally validate their model, external validation was rare. AUC values ranged from as low as 0.6 to as high as 0.97, and only two-thirds of papers reported calibration of their models. While most articles reported discrimination, calibration is equally important when trying to identify patients that will develop a given event versus those who will not. One should not use a model where the absolute risk estimates are not accurate. Sometimes calibration can be good in certain risk groups, but overestimates or underestimates risk in different populations. For this reason, better models are those that report both these values. Furthermore, just under half the studies reported their model in the form of a web-based calculator. Model deployment in this format greatly enhances the ability of a clinician to incorporate such a model into their clinical workflow.

Complications

Models predicting complications after degenerative spine surgery were the most commonly published; however, the types of models and the datasets used to create them varied greatly. Lee

et al¹² retrospectively evaluated 1476 patients undergoing degenerative spine surgery from a single institutional surgical registry to construct a predictive model of postoperative major complications, minor complications, surgical site infection, and durotomy. They reported an AUC of 0.76 for any complication and 0.81 for major complications and deployed their model at <http://depts.washington.edu/spinersk/>. McGirt et al¹¹ prospectively evaluated 1803 patients undergoing lumbar spine surgery at a single institution to produce a model that incorporated 45 baseline variables to predict postoperative complications with an AUC of 0.72. Most recently, Janssen et al⁴¹ reported a single institution retrospective series predicting postoperative infection with an AUC of 0.72.

The other studies that published models of complications used multi-institutional data. Ratliff et al¹⁴ retrospectively evaluated 279 315 patients from a longitudinal national claims database to construct a predictive model of complications after surgery. They produced a model with an AUC of 0.70 and deployed the algorithm in a freely available smartphone application (<http://itunes.apple.com/app/ratool/id1087663216>). The authors also externally validated this model using data from a single-institution prospective patient series (N = 246).⁴²

Kim et al¹³ retrospectively evaluated 22,629 patients using the cross-sectional NSQIP database to develop machine learning models to identify risk factors for complications after posterior lumbar spine fusion. AUCs for logistic regression and artificial neural network models both outperformed benchmark American Society of Anesthesiologists (ASA) class for predicting complications. In their logistic regression model, the AUC for predicting cardiac complications was 0.66, for predicting venous thromboembolism was 0.59, for predicting wound infection was 0.61, and for predicting mortality was 0.7. Of note, several authors including Sebastian et al⁴³ attempted to validate the previously developed NSQIP Surgical risk calculator (riskcalculator.facs.org). They found that the calculator generally had relatively poor predictive performance across all outcomes measured, including an AUC of 0.56 for reoperation, 0.61 for any complication, 0.61 for serious complications, and 0.63 for surgical site infection.

Table 4. Prediction Models for Clinical Improvement of Patients Undergoing Spine Surgery.

Author, year	Institutions	Design	Time length	Sample size	Internal AUC	Calib?	Internal validation	External validation	Calc
Spratt, 2004 ²⁷	Single	Prospective	1 year	40	AUC N/A PPV 85.7%, NPV 100%	No	N/A	No	No
Hegarty, 2012 ³⁷	Single	Prospective	90days	53	PPSP: 0.66	No	Bootstrapping	No	No
McGirt, 2015 ¹¹	Single	Prospective	1 year	1803	ODI: R ² = 0.51 Return to work: 0.79	Yes	Training/validation	No	No
Asher, 2017 ³²	Multi	Retrospective	90days	4694	Return to work: 0.71	Yes	Bootstrapping	No	Yes ^a
Lubelski, 2017 ¹⁷	Single	Retrospective	1 year	952	EQ-5D R ² = 0.43, PHQ-9 R ² = 0.35, PDQ R ² = 0.47	Yes	Bootstrapping	No	Yes
McGirt, 2017 ²⁶	Multi	Prospective	1 year	7618	ODI: 0.69, EQ-5D: 0.69 LBP: 0.67, Leg pain: 0.64	Yes	Bootstrapping	No	Yes ^a
Devin, 2018 ³³	Multi	Retrospective	90days	4689	Return to work: 0.81	Yes	Bootstrapping	No	No
Khor, 2018 ²⁸	Multi	Retrospective	1 year	1,583	ODI: 0.73, LBP: 0.75 Leg pain: 0.75	Yes	Training/validation	ODI ⁴⁷ : 0.71, LBP: 0.72, Leg pain: 0.83	Yes
Asher, 2019 ⁴⁰	Multi	Retrospective	1 year	4148	Patient satisfaction: 0.64	No	Bootstrapping	No	No
Karhade, 2019 ³⁴	Multi	Retrospective	180 days	2737	Opioid dep: 0.80	Yes	Training/validation	No	Yes
Karhade, 2019 ³⁵	Multi	Retrospective	180 days	5,413	Opioid dep: 0.80	Yes	Training/validation	No	Yes
Karhade, 2019 ³⁶	Multi	Retrospective	180 days	8,435	Opioid dep: 0.70	Yes	Training/validation	No	Yes
Meralli, 2019 ³⁹	Multi	Retrospective	2 years	539	SF-6D/mJOA: 0.7 R ² = 0.78	No	Training/Validation	No	No
Penning, 2019 ²⁹	Multi	Retrospective	N/A	719		No	N/A	No	No
Rundell, 2019 ³⁸	Multi	Retrospective	1 year	5840	Micro-disc ODI: 0.76, NRS-BP: 0.75, NRS-LP: 0.74, PSI: 0.80	Yes	Bootstrapping	No	No
					Lami ODI: 0.76, NRS-BP: 0.74, NRS-LP: 0.73, PSI: 0.81	Yes			
					Lami++Fusion ODI: 0.77, NRS-BP: 0.75, NRS-LP: 0.74, PSI: 0.79	Yes			
Siccoli, 2019 ¹⁶	Single	Retrospective	1 year	635	ODI: 0.73, LBP: 0.75, Leg pain: 0.72	Yes	Training/validation	No	No
de Silva, 2020 ³¹	Single	Retrospective	1 year	64	mJOA: 0.69	No	N/A	No	No
Staub, 2020 ³⁰	Single	Retrospective	1 year	1244	N/A	Yes	Training/validation	No	Yes

Abbreviations: N/A, not available; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; LBP, low back pain; mJOA, modified Japanese Orthopedic Association; NRS, numeric rating scale (back pain and leg pain); PSI, Patient Satisfaction Index; PPSP, persistent postsurgical pain; ODI, Oswestry Disability Index; Microdisc, microdiscectomy; Lami, Laminectomy; Calib⁴⁷, calibration performed⁴⁷; Calc, whether the authors reported that they developed a Web-based Calculator.

^a No longer available at published URL.

Han et al¹⁵ retrospectively evaluated 1 106 234 patients from the Truven MarketScan, Commercial database, the Truven MarketScan Medicare Databases, and the CMS Medicare database to develop predictive models of adverse events 30 days after spine surgery. The predictors identified included patient demographics, medical comorbidities, surgical indication, and operative characteristics and the resultant model had an AUC of 0.70 for predicting overall adverse events.

Reoperation and Readmission

Prediction models of readmission and reoperation are particularly apt for current CMS hospital quality metrics. The articles that have looked at this have been primarily retrospective, with the exception of the article by McGirt and colleagues,¹¹ who prospectively evaluated 1803 patients at a single institution to develop multiple predictive models, including one for readmission. Using 45 baseline variables, their model yielded an AUC of 0.74. They did not have external validation and the large number of baseline variables as compared with overall number of readmission events (N = 108), may potentially increase the risk of overfitting and thereby limit generalizability.

Of the models derived from retrospective analyses, Siccoli et al¹⁶ evaluated 635 patients from a prospective registry using machine learning algorithms to predict need for reoperation and patient outcomes at 12 months. Their model for reoperation had an AUC of 0.63, which is on the lower end of the spectrum. Lubelski et al¹⁷ retrospectively evaluated 952 patients from a single institution who underwent anterior or posterior cervical decompression/fusion and found that predictors of clinical outcomes included race, median income, body mass index, medical comorbidities, presenting symptoms, surgical indication, surgery type, and number of operated levels. They validated their cohort using bootstrapping and found an AUC of 0.91 for 90-day reoperation, 0.63 for 30-day emergency department visits, and 0.78 for 30-day readmission. A web-based calculator was deployed at <https://riskcalc.org/PatientsEligibleforCervicalSpineSurgery/>.

Two additional studies used the ACS-NSQIP database to generate calculators. Hopkins et al¹⁹ retrospectively evaluated 23 264 patients who underwent posterior lumbar fusion and found that predictors of 30-day readmission included medical comorbidities and whether surgery was a reoperation or index case. Despite the limitations of the NSQIP database, their model achieved an AUC of 0.81. Though not included in the original article, the authors did later report that this model was adequately calibrated.⁴⁴ In contrast, the more inclusive study by Goyal et al,¹⁸ which had cervical *and* lumbar spinal fusion patients, developed a model with poorer predictive discrimination. They evaluated 59 145 patients from the ACS-NSQIP database and produced a model with an AUC of 0.66 for unplanned admission.

The national administrative databases are readily accessible and have very large numbers, which may increase the power for statistical analysis. Predictive models that are calculated from these databases, however, may be subject to significant bias

because of how the data is collected, completeness of the included variables, and how they are categorized based on billing diagnosis and procedure codes. Models based on smaller sample sizes may potentially be superior if the data is collected prospectively and if the data collection is more nuanced and accurate. Ultimately, when evaluating different prediction models, it is important to consider how the data was collected, sample size, number of institutions included, as well as AUC, discrimination, calibration.

Length of Stay and Discharge

In addition to predicting adverse outcomes, predicting prolonged length of hospital stay and discharge disposition can improve patient experience, reduce health-facility associated complications, and reduce costs. Several authors have developed prediction models to determine expected length of stay as well as the likelihood of discharge to nonhome or inpatient rehabilitation destination.

Using their prospective data set, McGirt et al¹¹ developed a model with an AUC of 0.84 for predicting discharge to inpatient rehabilitation. Lubelski et al²⁰ retrospectively evaluated 257 patients from a single institution and published a model that had an AUC of 0.89 for likelihood of rehabilitation discharge as well as AUC of 0.89 for prolonged LOS (>7 days). The authors deployed this model as a web-based calculator at <https://jhuspine1.shinyapps.io/RehabLOS/>. Similarly, Siccoli et al¹⁶ retrospectively evaluated a prospective registry of 635 patients undergoing lumbar decompression surgery using machine learning algorithms to predict extended length of stay (>28 hours) with an AUC of 0.77.

Guan and colleagues²⁵ used the Quality Outcomes Database (QOD), a multicenter prospective registry, to develop a prediction score of discharge needs for patients undergoing lumbar fusion. With an AUC of 0.81, their model could place a patient into the low- or high-score category, which would determine the likelihood of needing additional home services or acute rehabilitation.

The other publications on predictors of rehabilitation discharge all used the ACS-NSQIP database to generate prediction models. Harada et al²¹ evaluated 10,453 patients from the ACS-NSQIP database who underwent open lumbar fusion (AUC of 0.75), and then externally validated the model using their institutional dataset (AUC of 0.77). Similarly, Karhade et al²² evaluated 26 364 ACS-NSQIP patients who underwent lumbar surgery for degenerative disc disorders to generate a model with an AUC of 0.82. Their model was then externally validated by Stopa and colleagues⁴⁵ and the authors of the original article deployed a web-based calculator at <https://sorg-apps.shinyapps.io/discdisposition/>.

Ogink and colleagues²³ then published an evaluation of 9338 patients in the ACS-NSQIP database who underwent surgery for degenerative spondylolisthesis and found that their model predicted nonhome discharge with an AUC of 0.75 (<https://sorg-apps.shinyapps.io/spondydisposition/>). Then in a parallel publication, the same group²⁴ evaluated 28 600 patients

in the ACS-NSQIP database who underwent surgery for lumbar spinal stenosis and generated a model predicting nonhome discharge with an AUC of 0.75 (<https://sorg-apps.shinyapps.io/stenosisdisposition/>). Last, analyzing 59 145 ACS-NSQIP patients who underwent either cervical or lumbar spinal fusion, Goyal et al¹⁸ produced a model predicting nonhome discharge with an AUC of 0.87.

Pain, Disability, and Quality of Life

Functional and quality-of-life outcomes are critical to delivering patient-centered spine care. Therefore, these outcome metrics have also been the focus of clinical prediction models. The ODI is a widely used and extensively validated method for quantifying low back pain-associated disability and has been used by multiple prediction studies as an outcome.⁴⁶

McGirt et al²⁶ prospectively evaluated a larger cohort of 7618 patients from the NeuroPoint QOD one year after elective lumbar spine surgery and found that predictors of patient-reported outcomes (PROs) included employment status, baseline back pain, psychological distress, baseline ODI, level of education, workers' compensation status, symptom duration, race, baseline leg pain, ASA score, age, primary symptom, smoking status, and insurance status. Internal validation yielded modest AUCs of 0.69 for ODI, 0.67 for numeric rating scale (NRS) for back pain, and 0.64 for NRS for leg pain. Siccoli et al¹⁶ achieved comparable discriminative ability for these outcomes among patients undergoing single- or multi-level decompression for lumbar spinal stenosis, with data collected from retrospective review of a prospective registry. Khor et al²⁸ collected prospective, multi-institution registry data (N = 1583) for patients undergoing elective lumbar surgery and developed predictive models that achieved AUCs of 0.73 for ODI, 0.75 for NRS back pain, and 0.75 for NRS leg pain. A web-based calculator was deployed at https://becertain.shinyapps.io/lumbar_fusion_calculator. Importantly, these models were independently, externally validated by Quddusi et al.⁴⁷

An often-underestimated aspect in the development of clinical prediction models is variable selection. In an effort to address this, Rundell et al³⁸ retrospectively evaluated 5840 patients from multiple institutions to develop prognostic models of 1-year outcomes. The key finding of this study was that ODI at 3-months postsurgery was the strongest predictor of 12-month outcomes.³⁸ Future predictive studies should think carefully about variable selection and consider feature engineering, a term in machine learning that describes using domain knowledge to create variables that may drive improved predictive performance.

While the majority of predictive models in degenerative spine surgery have focused on lumbar spine surgery, early efforts in modeling quality-of-life outcomes for cervical spine surgery patients are emerging. In addition to predicting reoperation and readmission rates, Lubelski et al¹⁷ used their single-institution cohort of patients undergoing cervical spine surgery to develop nomograms for quality-of-life outcomes (EuroQOL, EQ-5D; PHQ-9, PDQ). These nomograms

predicted quality-of-life outcomes to varying degrees, with R^2 values of 0.43 for EQ-5D, 0.35 for PHQ-9, and 0.47 for PDQ.¹⁷ Asher and colleagues⁴⁰ used the Neuropoint QOD to create a model predicting patient satisfaction after 1- or 2-level anterior cervical discectomy and fusion. Their model had an AUC of 0.66, and found that geographical region, socioeconomic status, baseline disability and symptom duration all contributed to postoperative outcome. Devin et al³³ also utilized the QOD for cervical spine surgery patients and found that predictors of returning to work within 90 days included age, employment, occupation, workers' compensation, baseline Neck Disability Index score, presentation, and levels fused. They used bootstrapping to validate their cohort and achieved an AUC of 0.81.³³ And Merali et al³⁹ used the AOSpine prospective registry to predict postoperative SF-6D and mJOA quality-of-life outcomes in patients undergoing surgery for cervical spondylotic myelopathy. Their models used machine learning tools to predict 6-, 12-, and 24-month outcomes, and their best performing model had an average AUC of 0.7.

A final outcome of interest is opioid use following degenerative spine surgery. Associations between spine surgery and opioid use are well established.^{48,49} Karhade et al³⁴⁻³⁶ endeavored to build predictive models of sustained opioid use after cervical and lumbar spine surgery, defined as >90 days of uninterrupted prescription filling. Their models had AUCs ranging from 0.7 to 0.8 and were deployed as web-based calculators to potentially enable a surgeon, at the bedside, to identify an individual's specific risk.

Limitations and Future Directions

There is an increasing body of literature looking at predicting outcomes in degenerative spine surgery. Some focus on administrative outcomes such as readmission, emergency department visits, and reoperation, whereas others focus on patient reported outcomes and complications. Heterogeneity also exists in how the data is collected, how the analyses are performed and models validated, and the mechanisms by which the data is reported. To be integrated into clinical practice, prediction models need to have the data collected in a systematic way, preferably prospectively, with detailed clinical information. Models based on the Current Procedural Terminology and diagnosis codes of administrative databases are therefore inherently limited. Models need to assess for discrimination and calibration and should preferably have AUC >0.7. Details of how the analysis is performed should be explicitly reported. Validation should be performed with a patient population that is different from which the model was generated, ideally at another institution. If validation is performed on patients from the same institution, this limits the model's generalizability outside of the primary hospital setting.

Future directions include the generation a grading system to help clinicians determine the relative strengths of the different published models. Additionally, studies are needed to determine the usefulness of such prediction models. Better understanding is needed whether the use of a prediction model leads

to greater patient satisfaction, outcome, or value. Lastly, it is important to remember that regardless of how accurate the prediction model is, it cannot replace clinical judgment. There are innumerable clinical and social variables that are taken into account when helping patients decide on a treatment course. The goal is to create prediction calculators that can help the physician provide more accurate and individualized descriptions of the risk/benefit profile for a given patient.

Conclusion

The continued emphasis on value-based care in American health care and the variability in degenerative spine surgery outcomes presents an important case for clinical prediction modeling. The current body of clinical prediction for degenerative spine surgery is heterogeneous with regard to data sets, outcome measures, and statistical learning methods. Importantly, external validation of proposed models must be emphasized and executed. While the promise of clinical prediction in degenerative spine surgery for patients, hospitals, and health systems is significant, further efforts are required before current models are appropriate for clinical deployment





Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Daniel M. Sciubba is a consultant for Baxter, DePuy-Synthes, Globus Medical, K2M, Medtronic, NuVasive, Stryker, and receives unrelated grant support from Baxter Medical, North American Spine Society, and Stryker. The other authors have no disclosures to make.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This supplement was supported by a grant from AO Spine North America.

ORCID iD

Daniel Lubelski, MD  <https://orcid.org/0000-0002-9403-9509>
 Andrew Hersh, BA  <https://orcid.org/0000-0003-1755-3974>
 Zachary Pennington, BS  <https://orcid.org/0000-0001-8012-860X>
 Daniel M. Sciubba, MD  <https://orcid.org/0000-0001-7604-434X>

References

- Lubelski D, Alentado VJ, Williams SK, et al. Variability in surgical treatment of spondylolisthesis among spine surgeons. *World Neurosurg.* 2018;111:e564-e572. doi:10.1016/j.wneu.2017.12.108
- Alvin MD, Lubelski D, Alam R, et al. Spine surgeon treatment variability: the impact on costs. *Global Spine J.* 2018;8:498-506. doi:10.1177/2192568217739610
- Mroz TE, Lubelski D, Williams SK, et al. Differences in the surgical treatment of recurrent lumbar disc herniation among spine surgeons in the united states. *Spine J.* 2014;14:2334-2343. doi:10.1016/j.spinee.2014.01.037
- Azad TD, Vail D, O'Connell C, Han SS, Veeravagu A, Ratliff JK. Geographic variation in the surgical management of lumbar spondylolisthesis: characterizing practice patterns and outcomes. *Spine J.* 2018;18:2232-2238. doi:10.1016/j.spinee.2018.05.008
- Bailey CS, Rasoulinejad P, Taylor D, et al. Surgery versus conservative care for persistent sciatica lasting 4 to 12 months. *N Engl J Med.* 2020;382:1093-1102. doi:10.1056/NEJMoa1912658
- Ghogawala Z, Dziura J, Butler WE, et al. Laminectomy plus fusion versus laminectomy alone for lumbar spondylolisthesis. *N Engl J Med.* 2016;374:1424-1434. doi:10.1056/NEJMoa1508788
- Försth P, Ólafsson G, Carlsson T, et al. A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. *N Engl J Med.* 2016;374:1413-1423. doi:10.1056/NEJMoa1513721
- Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus non-surgical therapy for lumbar spinal stenosis. *N Engl J Med.* 2008;358:794-810. doi:10.1056/NEJMoa0707136
- Azad TD, Veeravagu A, Mittal V, et al. Neurosurgical randomized controlled trials-distance travelled. *Neurosurgery.* 2018;82:604-612. doi:10.1093/neuros/nyx319
- Mansouri A, Cooper B, Shin SM, Kondziolka D. Randomized controlled trials and neurosurgery: The ideal fit or should alternative methodologies be considered? *J Neurosurg.* 2016;124:558-568. doi:10.3171/2014.12.JNS142465
- 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:E13. doi:10.3171/2015.8.FOCUS15338
- 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;14:291-299. doi:10.1016/j.spinee.2013.10.043
- Kim JS, Merrill RK, Arvind V, et al. Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. *Spine (Phila Pa 1976).* 2018;43:853-860. doi:10.1097/BRS.0000000000002442
- Ratliff JK, Balise R, Veeravagu A, et al. Predicting occurrence of spine surgery complications using "big data" modeling of an administrative claims database. *J Bone Joint Surg Am.* 2016;98:824-834. doi:10.2106/JBJS.15.00301
- Han SS, Azad TD, Suarez PA, Ratliff JK. A machine learning approach for predictive models of adverse events following spine surgery. *Spine J.* 2019;19:1772-1781. doi:10.1016/j.spinee.2019.06.018
- Siccoli A, de Wispelaere MP, Schröder ML, Staartjes VE. Machine learning-based preoperative predictive analytics for lumbar spinal stenosis. *Neurosurg Focus.* 2019;46:E5. doi:10.3171/2019.2.FOCUS18723
- Lubelski D, Alentado V, Nowacki AS, et al. Preoperative nomograms predict patient-specific cervical spine surgery clinical and quality of life outcomes. *Neurosurgery.* 2017;83:104-113. doi:10.1093/neuros/nyx343
- Goyal A, Ngufor C, Kerezoudis P, McCutcheon B, Storlie C, Bydon M. Can machine learning algorithms accurately predict discharge to nonhome facility and early unplanned readmissions following spinal fusion? Analysis of a national surgical registry.

- J Neurosurg Spine*. 2019;31:568-578. doi:10.3171/2019.3.SPINE181367
19. Hopkins BS, Yamaguchi JT, Garcia R, et al. Using machine learning to predict 30-day readmissions after posterior lumbar fusion: an NSQIP study involving 23,264 patients. *J Neurosurg Spine*. 2019;32:399-406. doi:10.3171/2019.9.SPINE19860
 20. Lubelski D, Ehresman J, Feghali J, et al. Prediction calculator for nonroutine discharge and length of stay after spine surgery. *Spine J*. 2020;20:1154-1158. doi:10.1016/j.spinee.2020.02.022
 21. Harada GK, Basques BA, Samartzis D, Goldberg EJ, Colman M, An HS. Development and validation of a novel scoring tool for predicting facility discharge after elective posterior lumbar fusion. *Spine J*. Published online March 2, 2020. doi:10.1016/j.spinee.2020.02.014
 22. Karhade AV, Ogink P, Thio Q, et al. Development of machine learning algorithms for prediction of discharge disposition after elective inpatient surgery for lumbar degenerative disc disorders. *Neurosurg Focus*. 2018;45:E6. doi:10.3171/2018.8.FOCUS18340
 23. Ogink PT, Karhade AV, Thio QCBS, et al. Development of a machine learning algorithm predicting discharge placement after surgery for spondylolisthesis. *Eur Spine J*. 2019;28:1775-1782. doi:10.1007/s00586-019-05936-z
 24. Ogink PT, Karhade AV, Thio QCBS, et al. Predicting discharge placement after elective surgery for lumbar spinal stenosis using machine learning methods. *Eur Spine J*. 2019;28:1433-1440. doi:10.1007/s00586-019-05928-z
 25. Guan J, Knightly JJ, Bisson EF. Development of a predictive score for discharge disposition after lumbar fusion using the quality outcomes database. *Neurosurgery*. 2018;83:452-458. doi:10.1093/neuros/nyx436
 26. 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;27:357-369. doi:10.3171/2016.11.SPINE16526
 27. Spratt KF, Keller TS, Szpalski M, Vandeputte K, Gunzburg R. A predictive model for outcome after conservative decompression surgery for lumbar spinal stenosis. *Eur Spine J*. 2004;13:14-21. doi:10.1007/s00586-003-0583-2
 28. Khor S, Lavallee D, Cizik AM, et al. Development and validation of a prediction model for pain and functional outcomes after lumbar spine surgery. *JAMA Surg*. 2018;153:634-642. doi:10.1001/jamasurg.2018.0072
 29. Pennings JS, Devin CJ, Khan I, Bydon M, Asher AL, Archer KR. Prediction of Oswestry Disability Index (ODI) using PROMIS-29 in a national sample of lumbar spine surgery patients. *Qual Life Res*. 2019;28:2839-2850. doi:10.1007/s11136-019-02223-8
 30. Staub LP, Aghayev E, Skrivankova V, Lord SJ, Haschtmann D, Mannion AF. Development and temporal validation of a prognostic model for 1-year clinical outcome after decompression surgery for lumbar disc herniation. *Eur Spine J*. 2020;29:1742-1751. doi:10.1007/s00586-020-06351-5
 31. De Silva T, Vedula SS, Perdomo-Pantoja A, et al. SpineCloud: image analytics for predictive modeling of spine surgery outcomes. *J Med Imaging (Bellingham)*. 2020;7:031502. doi:10.1117/1.JMI.7.3.031502
 32. Asher AL, Devin CJ, Archer KR, et al. An analysis from the quality outcomes database, part 2. predictive model for return to work after elective surgery for lumbar degenerative disease. *J Neurosurg Spine*. 2017;27:370-381. doi:10.3171/2016.8.SPINE16527
 33. Devin CJ, Bydon M, Mohammed AA, et al. A predictive model and nomogram for predicting return to work at 3 months after cervical spine surgery: an analysis from the quality outcomes database. *Neurosurg Focus*. 2018;45:E9. doi:10.3171/2018.8.FOCUS18326
 34. Karhade AV, Ogink PT, Thio QCBS, et al. Machine learning for prediction of sustained opioid prescription after anterior cervical discectomy and fusion. *Spine J*. 2019;19:976-983. doi:10.1016/j.spinee.2019.01.009
 35. Karhade AV, Ogink PT, Thio QCBS, et al. Development of machine learning algorithms for prediction of prolonged opioid prescription after surgery for lumbar disc herniation. *Spine J*. 2019;19:1764-1771. doi:10.1016/j.spinee.2019.06.002
 36. Karhade AV, Cha TD, Fogel HA, et al. Predicting prolonged opioid prescriptions in opioid-naïve lumbar spine surgery patients. *Spine J*. 2020;20:888-895. doi:10.1016/j.spinee.2019.12.019
 37. Hegarty D, Shorten G. Multivariate prognostic modeling of persistent pain following lumbar discectomy. *Pain Physician*. 2012;15:421-434.
 38. Rundell SD, Pennings JS, Nian H, et al. Adding 3-month patient data improves prognostic models of 12-month disability, pain, and satisfaction after specific lumbar spine surgical procedures: development and validation of a prediction model. *Spine J*. 2020;20:600-613. doi:10.1016/j.spinee.2019.12.010
 39. Merali ZG, Witw CD, Badhiwala JH, Wilson JR, Fehlings MG. Using a machine learning approach to predict outcome after surgery for degenerative cervical myelopathy. *PLoS One*. 2019;14:e0215133. doi:10.1371/journal.pone.0215133
 40. Asher AL, Devin CJ, Kerezoudis P, et al. Predictors of patient satisfaction following 1- or 2-level anterior cervical discectomy and fusion: insights from the quality outcomes database. *J Neurosurg Spine*. 2019;31:835-843. doi:10.3171/2019.6.SPINE19426
 41. Janssen DMC, van Kuijk SMJ, d'Aumerie B, Willems P. A prediction model of surgical site infection after instrumented thoracolumbar spine surgery in adults. *Eur Spine J*. 2019;28:775-782. doi:10.1007/s00586-018-05877-z
 42. 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;27:81-91. doi:10.3171/2016.12.SPINE16969
 43. Sebastian A, Goyal A, Alvi MA, et al. Assessing the performance of national surgical quality improvement program surgical risk calculator in elective spine surgery: insights from patients undergoing single-level posterior lumbar fusion. *World Neurosurg*. 2019;126:e323-e329. doi:10.1016/j.wneu.2019.02.049
 44. Staartjes VE, Kernbach JM. Letter to the editor. Importance of calibration assessment in machine learning-based predictive analytics. *J Neurosurg Spine*. 2020;32:985-987. doi:10.3171/2019.12.SPINE191503
 45. Stopa BM, Robertson FC, Karhade AV, et al. Predicting nonroutine discharge after elective spine surgery: external validation of

- machine learning algorithms. *J Neurosurg Spine*. 2019;31:742-747. doi:10.3171/2019.5.SPINE1987
46. Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)*. 2000;25:2940-2952. doi:10.1097/00007632-200011150-00017
47. Quddusi A, Eversdijk HAJ, Klukowska AM, et al. External validation of a prediction model for pain and functional outcome after elective lumbar spinal fusion. *Eur Spine J*. 2020;29:374-383. doi:10.1007/s00586-019-06189-6
48. Vail D, Azad TD, O'Connell C, Han SS, Veeravagu A, Ratliff JK. Postoperative opioid use, complications, and costs in surgical management of lumbar spondylolisthesis. *Spine (Phila Pa 1976)*. 2018;43:1080-1088. doi:10.1097/BRS.00000000000002509
49. O'Connell C, Azad TD, Mittal V, et al. Preoperative depression, lumbar fusion, and opioid use: an assessment of postoperative prescription, quality, and economic outcomes. *Neurosurg Focus*. 2018; 44:E5. doi:10.3171/2017.10.FOCUS17563
50. Kasperek MF, Boettner F, Rienmueller A, et al. Predicting medical complications in spine surgery: evaluation of a novel online risk calculator. *Eur Spine J*. 2018;27:2449-2456. doi:10.1007/s00586-018-5707-9