The Spine Journal 19 (2019) 1772−1781

Clinical Study

A machine learning approach for predictive models of adverse events following spine surgery

Summer S. Han, PhD, Tej D. Azad, MS, Paola A. Suarez, MS, John K. Ratliff, MD*

Department of Neurosurgery, Stanford University School of Medicine, Stanford, CA 94305, USA
Received 14 March 2019; revised 19 May 2019; accepted 17 June 2019

Abstract

BACKGROUND: Rates of adverse events following spine surgery vary widely by patient-, diagnosis-, and procedure-related factors. It is critical to understand the expected rates of complications and to be able to implement targeted efforts at limiting these events.
PURPOSE: To develop and evaluate a set of predictive models for common adverse events after spine surgery.
STUDY DESIGN: A retrospective cohort study.
PATIENT SAMPLES: We extracted 345,510 patients from the Truven MarketScan (MKS) and MarketScan Medicaid Databases and 760,724 patients from the Centers for Medicare and Medicaid Services (CMS) Medicare database who underwent spine surgeries between 2009 and 2013.
OUTCOME MEASURES: Overall adverse event (AE) occurrence and types of AE occurrence during the 30-day postoperative follow-up.
METHODS: We applied a least absolute shrinkage and selection operator regularization method and a logistic regression approach for predicting the risks of an overall AE and the top six most commonly observed AEs. Predictors included patient demographics, location of the spine procedure, comorbidities, type of surgery performed, and preoperative diagnosis.
RESULTS: The median ages of MKS and CMS patients were 49 years and 69, respectively. The most frequent individual AE was a cardiac dysfunction in CMS (10.6%) patients and a pulmonary complication (4.7%) in MKS. The area under the curve (AUC) of a prediction model for an overall AE was 0.7. Among the six individual prediction models, the model for predicting the risk of a pulmonary complication showed the greatest accuracy (AUC 0.76), and the range of AUC for these six models was 0.7 and 0.76. Medicaid status was one of the most important factors in predicting the occurrences of AEs; Medicaid recipients had increased odds of AEs by 20%−60% compared with non-Medicaid patients (odds ratios 1.28−1.6; $p<10^{-10}$). Logistic regression showed higher AUCs than least absolute shrinkage and selection operator across these different models.
CONCLUSIONS: We present a set of predictive models for AEs following spine surgery that account for patient-, diagnosis-, and procedure-related factors which can contribute to patient-counseling, accurate risk adjustment, and accurate quality metrics.

Keywords: Spine surgery; Adverse event; Prediction model; Risk adjustment

FDA device/drug status: Not applicable.

Author Disclosures: ***SSH:*** Nothing to disclose. ***TDA:*** Nothing to disclose. ***PAS:*** Nothing to disclose. ***JKR:*** Grant: AHRQ (G); Consulting: Stryker (C); Medtronic (B); Nuvasive (B); Grants: The Glen and Angela Charles Foundation (E).

Level of evidence: III

* Corresponding author. Department of Neurosurgery, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94305, USA. Tel.: (650) 723-6469; fax: (650) 725-0390.

E-mail address: jratliff@stanford.edu (J.K. Ratliff).

Introduction

Adverse events (AEs) following spine surgery negatively impact patients, surgeons, and the health care system. The incidence of AEs in the 30 days following a spine surgery procedure range from 40% to 50% in single-center prospective studies [1−5] and 2%−23% in retrospective studies using administrative databases and national registries [6,7].

https://doi.org/10.1016/j.spinee.2019.06.018
1529-9430/

Post- and perioperative metrics have become a benchmark to evaluate surgeons and hospitals. As reimbursement becomes tied to quality metrics, it is critical to investigate factors that are associated with AEs and to develop risk stratification strategies that account for patient factors. Predictive models that account for patient-, diagnosis-, and procedure-related factors are necessary for patient-counseling, accurate risk adjustment, and accurate quality metrics.

The objective of this study was to build a set of models based on a wide array of patients to best reflect the overall population of patients undergoing spine surgery in the US. We used a variety of patients for model development, including an elderly patient population via records from the Centers for Medicare and Medicaid Services (CMS), privately insured patients from Truven MarketScan (MKS), and patients with lower socioeconomic status through patients from the Truven MKS Medicaid database. The potential predictors for these models include preoperative diagnosis, comorbidities, surgical procedures, demographics, and interaction among these factors. As a measure of lower socioeconomic status, we include Medicaid status in our analysis.

Methods

Study design and data sources

We conducted a retrospective administrative database study of more than one million patients who had undergone spine surgery in the United States from 2009 to 2013. We used the two databases: (1) the claims data (N=345,510 patients) from the Truven Health Analytics MarketScan Commercial Claims and Encounters and Medicare Supplemental and Coordination of Benefits databases (denoted MKS in the manuscript); and (2) CMS Medicare data (denoted CMS in the manuscript), including 760,724 Medicare beneficiaries. The databases used in this assessment included 157,895 Medicaid recipients from the MarketScan Medicaid and CMS databases who had undergone spine surgery procedures; the Medicaid cohort included both patients with Medicaid as a primary insurer and also dual-eligible Medicare and Medicaid beneficiaries. A previous approach using a smaller database of Marketscan privately insured patients following similar methodology, has been described [7].

Split-sample approach: training versuss validation set

We randomly divided the data into a training dataset (70%) and a validation dataset (30%) based on the suggested split proportions provided in the literature [8]. The training dataset was used to develop a set of prediction models for various types of AEs both a generalized liner regression model with a logit link function (ie, logistic regression model) and a least absolute shrinkage and selection operator (LASSO) regularization method (see the following subsections). The validation dataset was used to evaluate the performance of the prediction models developed using the training data.

Cohort definition

Our cohort of patients was defined by querying the overall MKS and CMS databases for patients with Common Procedural Terminology descriptors for spine surgeries (Supplemental Table S1). Patients were divided into four general preoperative diagnostic groups: degenerative disease, trauma, neoplasm, and infection (see Supplemental Table S2 for ICD-9-CM codes of each category). MKS and CMS are both longitudinal databases, allowing for tracking of patients over time and for identification of new diagnoses, not present before admission for a given surgery. Based on the retrospective assessment of diagnoses present before the admission for spine surgery, we evaluated comorbidities of patients using the databases (Supplemental Table S3). Both databases are compiled from billing records, and entries are made as part of the billing process by hospitals and physician practices; they are processed by CMS and by third party payers for health care. Common Procedural Terminology codes are based upon claims paid for physician services, and ICD-9-CM codes are entered by facilities, by hospital-based chart abstractors, or by individual physician practice coders.

Definition of the outcome variables: adverse events

We defined an AE as the occurrence of new ICD-9-CM codes either during the admission for a given spine surgery or during the patient's postoperative follow-up claims history. We restricted the analysis to the 30 days immediately after the date of the surgery. We limited the dataset to patients with at least 30-day follow-up, and hence patients with less follow-up were not included in the analysis. The approach of using longitudinal databases to capture complication occurrence has been explored previously [7]. Longitudinal databases such as Marketscan and Medicare have been shown to capture rates of AE occurrence after spine surgery procedures that are comparable to prospective data capture [9]. Definitions of the 15 different types of AEs are presented in Supplemental Table S4. An overall AE was defined as having at least one of the 15 different AEs. We were interested in developing prediction models for the top six most frequent AEs (with the prevalence rate at least larger than 2.5% in either CMS or MKS)—cardiac dysfunction, congestive heart failure, pulmonary complication, pneumonia, neurologic complication, and urinary tract infections. We note that this grouping of AEs into less granular categories of the most frequent AEs observed in our patient population risks sacrificing granularity for ease of use of the model.

Features

The following categories of features or variables were considered as potential predictors for the risks of AEs, including (1) demographic factors, (2) preoperative diagnosis, (3) procedure-based cohort indicator for anterior cervical, posterior cervical, anterior thoracolumbar, and posterior thoracolumbar procedure groups; (4) comorbidities, and (5) surgery procedure. It is possible that the effects of some features may vary by Medicaid or Medicare status, by type of surgical approach, or by other factors in the data, and hence we allowed pairwise interactions between all the features. The possibility of three-way interaction (ie, a two-way interaction that varies across levels of a third variable) was considered *a priori* and assessed using a likelihood ratio test.

Predictive modeling approaches

In building predictive models, we used a LASSO regression approach based on a penalized regression to obtain shrinkage estimators for the regression coefficients [10], which has been widely applied in predicting complications rates after various surgical procedures using administrative claims data [7,11]. Although the statistical model used under LASSO is a generalized linear model with a logit link function, by its nature, LASSO uses a regularization method and shrinkage estimators to impose a constraint on the model parameters that causes regression coefficients for some variables to shrink toward zero. Therefore, the selection of the features and the estimation of the parameters under LASSO are different from those based on a stepwise selection method using logistic regression. In conducting the LASSO analysis, we used a 10-fold cross-validation to find a tuning parameter for each predictive model. The second approach to build prediction models was logistic regression. In order to select features for logistic regression models, we conducted a backward stepwise selection procedure based on Akaike information criterion. To capture the nonlinearity of continuous variables, we used a natural cubic spline method that is constructed of piecewise third-order polynomials [12]. Variable importance was evaluated based on the absolute value of the z-statistic for each model parameter used.

Performance evaluation

To evaluate the performance of the predictive models, discrimination was assessed using the receiver operator characteristic area under the curve (AUC) on both training and validation data. Calibration was assessed by plotting the observed incidence of each AE against the model-predicted probability of incidence. In a well-calibrated model, we expect the predictions to be close to a 45° diagonal line. We also calculated the Brier score that is measured as the average squared difference between the predicted probability and the observed outcome of each model (that ranges between 0 and 1, where the brier score value of 0 shows a perfect fit). Sensitivity and specificity were also calculated using the median predicted probability as a threshold.

Sensitivity analysis

For sensitivity analysis, we created other “overall” AE variables to examine if a different definition of an overall AE impacts the performance of predictive models. First, we defined a medical related overall AE variable that excludes any of the following surgical complications—(1) wound hematoma, (2) wound infection, (3) other wound complication, and (4) infection. Second, we defined a surgical related overall AE so that the variable is coded as 1 (vs. 0) if a subject has at least one of the following complications listed in (1)−(4) or a neurologic complication.

In order to explore how the sample size of data affects the performance of a prediction model for an overall AE, we conducted a simulation study; we randomly selected a subset of data from a varying size from N=10,000 to N=1,000,000 patients and repeated the model fitting procedure for predicting an overall AE using a logistic regression to calculate AUC. To increase the robustness of the results, we repeated the random sampling 10 times for each sample size and we reported the average AUC for each sample size.

Benchmark model

We compared the performance of our new model for an overall AE that incorporates more diverse study subjects to the performance of the existing model (“benchmark model”) [7] that was developed using the similar database (MKS database) but without elderly patients in CMS. We applied the benchmark model to the entire data (based both on MKS and CMS cohort) and calculated the performance metrics.

Results

Cohort characteristics

Our patient cohort characteristics are shown in Table 1**,** procedure descriptors for each of the patient cohorts are tabulated in Tables 2 and 3. The average age of the entire cohort was 62 years (standard deviation [SD]: 14.74 years). The mean age of the CMS-Medicare cohort was 69 (SD10.7), older than the mean age for the MKS cohort of 49 (SD 12.5). The overall AE rate (ie, the proportion of subjects who have at least one AE event) in the patient population was 24.7%, which was 27.6% in CMS versus 18.0% in MKS. The most common individual AE was a cardiac dysfunction in CMS (10.6%) patients and a pulmonary complication (4.7%) in MKS. Overall, the patients in CMS also had higher comorbidity rates compared with MKS.

Prediction of an overall AE

The receiver operator characteristic curve of the prediction model for an overall AE based on training data is shown in Figure 1, with a corresponding AUC value of 0.7. A calibration plot for the prediction model is displayed in

Table 1
Patient characteristics for MKS and CMS

	MKS (2009−2013)		CMS (2009−2013)	
Cohort attributes	N (343,509)	% (std)	N (760,724)	% (std)
Gender				
Male	160,988	47.0	347,022	45.6
Age				
Average age at time of surgery (yrs)	————	49 (12)	————	69 (11)
Medicaid				
Yes	31,101	9.1	126,794	16.7
No	312,408	90.9	633,723	83.3
Unknown/Other	0	0.0	207	0.0
Spine procedure type				
Cervical-unambiguous	123,782	36.0	184,969	24.3
Thoracolumbar-unambiguous	226,127	65.8	586,651	77.1
Anterior cervical	104,834	30.5	141,203	18.6
Posterior cervical	23,849	6.9	50,020	6.6
Anterior thoracolumbar	32,892	9.6	42,436	5.6
Posterior thorocolumbar	213,360	62.1	570,429	75.0
Had fusion	213,522	62.0	391,700	51.5
Had instrumentation	253,231	74.0	431,569	56.7
Used additonal level	177,377	53.0	467,208	61.4
Used BMP	24,350	7.1	90,507	11.9
Diagnosis of				
Degenerative disease	334,678	97.4	754,379	99.2
Neoplasm	9,170	2.7	14,955	2.0
Trauma	18,221	5.3	19,325	3.0
Infection	3,558	1.0	4,369	<1
Pre-existing comorbidities:				
Any comorbidities	221,728	65.0	665,639	87.5
Pulmonary disorder	36,103	10.5	136,370	17.9
Neurological disorder/deficit	22,003	6.4	62,659	8.2
Hypercholesterolemia	52,304	15.2	283,422	37.3
Smoking	62,301	18.1	184,581	24.3
Hypertension	115,230	33.5	496,143	65.2
Cardiac disorder other than hypertension	324	<1	4,525	<1
Diabetes mellitus	39,042	11.4	181,906	23.9
Cancer	17,213	5.0	91,184	12.0
Gastroesophageal disorder	2,389	0.7	13,231	1.7
ETOH/drug use	6,462	1.9	488	<1
Psychiatric disorder	52,271	15.2	150,303	19.8
Complications (within 30 days postsurgery)				
Overall (any) complication	60,958	18.0	209,646	27.6
Cardiac dysrhythmia	14,689	4.3	80,822	10.6
Pulmonary	16,138	4.7	40,046	5.3
Urinary tract infection (UTI)	11,410	3.3	46,786	6.2
Neurological complications	7,317	2.1	29,462	4.0
Congestive heart failure (CHF)	3,538	1.0	26,989	3.6
Pneumonia	6,629	1.9	21,861	2.9
Deep vein thrombosis (DVT)	6,055	1.8	18,344	2.4
Wound hematoma	5,523	1.6	17,700	1.6
Other wound complications	4,383	1.3	9,352	1.0
Myocardial infarction (MI)	2,429	0.7	10,724	1.4
Pulmonary embolism (PE)	2,251	0.7	7,651	1.0
Renal failure	2,021	0.6	7,040	0.9
Delirium	1,539	<1	8,478	1.1

Figure 1, which shows that the predicted probability and observed outcome are in good agreement (intercept 0; slope 1). Other performance metrics that include the brier score, sensitivity, and specificity are shown in Supplemental Table S6. The top 20 most important variables for predicting an overall AE based on the final model selected by logistic regression (that included 274 variables) on training data are shown in Figure 2; the corresponding odds ratios (OR) and p values are shown in Supplemental Table S7. According to these results, Medicaid status is significantly associated with the risk of an overall AE with an OR of 1.26 (95% Confidence Interval 1.24−1.28, $p<1 \times 10^{-10}$), indicating

Table 2
MKS comorbidities and complication occurrence in cervical and thoracolumbar spine surgery

Cohort attributes	Cervical		Cohort attributes	Thoracolumbar	
Total number	N (123,782)	% (SD)	Total number	N (226,127)	% (SD)
Gender			*Gender*		
Male	58,699	47.0	Male	105,585	47.0
Age			*Age*		
Average age at time of surgery (yrs)	-	50 (10)	Average age at time of surgery (yrs,(std))	-	48 (14)
Spine procedure type			*Spine procedure type*		
Cervical procedures			*Thoracolumbar procedures*		
Anterior cervical decompression and fusion (ACDF) single level	18,633	15.1	Posterior thoracic decompression (PTD)	5,118	2.3
ACDF single + instrumentation	18,225	14.7	PTD + instrumentation	502	<1
ACDF single + bone morphogenic protein (BMP)	537	<1	PTD + bone morphogenic protein (BMP)	26	<1
ACDF, multiple level	49,168	39.7	Posterior thoracic decompression and fusion (PTDF)	4,488	2.0
ACDF multiple level + instrumentation	48,653	39.3	PTDF + instrumentation	4,268	1.9
ACDF multiple + BMP	1,537	1.2	PTDF + BMP	381	<1
Anterior cervical corpectomy	16,441	13.3	Posterior lumbar decompression (PLD)	75,802	33.5
ACC + instrumentation	15,806	12.8	PLD + instrumentation	2,604	1.2
ACC + BMP	460	<1	PLD + BMP	418	<1
Posterior cervical decompression (PCD)	11,400	9.2	Posterior lumbar decompression and fusion (PLDF)	140,195	62.0
PCD + instrumentation	1,556	1.3	PLDF + instrumentation	135,755	60.0
PCD + BMP	105	<1	PLDF + BMP	20,161	8.9
Posterior cervical decompression with fusion (PCDF)	10,038	8.1	Anterior thoracolumbar decompression and fusion (ATCDF)	7,591	3.4
PCDF + instrumentation	9,721	7.9	ATCDF + instrumentation	7,456	3.3
PCDF + BMP	690	0.6	ATCDF + BMP	903	<1
Total instrumentation	111,336	89.9	Total instrumentation	147,686	65.3
Total BMP	3,974	3.2	Total BMP	20,829	9.2
Primary diagnosis of			*Primary diagnosis of*		
Degenerative disease	118,375	98.0	Degenerative disease	219,343	97.0
Neoplasm	1,237	1.0	Neoplasm	4,702	2.1
Trauma	6,189	5.0	Trauma	5,777	2.6
Infection	396	<1	Infection	1,503	<1
Other	740	<1	Other	10,152	4.5
Pre-existing comorbidities:			*Pre-existing comorbidities:*		
Pulmonary disorder	13,655	11.0	Pulmonary disorder	23,172	10.2
Neurological disorder/deficit	7,609	6.1	Neurological disorder/deficit	15,279	6.8
Hypercholesterolemia	18,952	15.3	Hypercholesterolemia	34,168	15.1
Smoking	25,251	20.4	Smoking	38,101	16.8
Hypertension	41,727	33.7	Hypertension	75,588	33.4
Cardiac disorder other than hypertension	121	<1	Cardiac disorder other than hypertension	213	<1
Diabetes mellitus	14,312	11.6	Diabetes mellitus	25,443	11.3
Cancer	5,487	4.4	Cancer	12,353	5.5
Gastroesophageal disorder	942	0.8	Gastroesophageal disorder	1,518	<1
ETOH/drug use	2,616	2.1	ETOH/drug use	4,138	1.8
Psychiatric disorder	19,567	15.8	Psychiatric disorder	33,713	14.9
Complications (within 30 days postsurgery)			*Complications (within 30 days postsurgery)*		
Overall (any) complication	18,064	14.6	Overall (any) complication	45,103	19.9
Cardiac dysrhythmia	5,034	4.1	Cardiac dysrhythmia	10,178	4.5
Pulmonary	5,764	4.7	Pulmonary	11,323	5.0
Urinary tract infection (UTI)	3,236	2.6	Urinary tract infection (UTI)	8,584	3.8
Neurological complications	2,579	2.1	Neurological complications	5,087	2.2
Pneumonia	2,497	2.0	Pneumonia	4,547	2.0
Deep vein thrombosis (DVT)	1,777	1.4	Deep vein thrombosis (DVT)	4,548	2.0
Wound hematoma	1,411	1.1	Wound hematoma	4,272	1.9
Other wound complications	847	0.7	Other wound complications	3,708	1.6
Myocardial infarction (MI)	683	0.6	Myocardial infarction (MI)	1,821	0.8
Pulmonary embolism (PE)	607	<1	Pulmonary embolism (PE)	1,737	0.8
Renal failure	519	<1	Renal failure	1,565	0.7
Congestive heart failure (CHF)	1,182	1.0	Congestive heart failure (CHF)	2,493	1.1
Delirium	473	<1	Delirium	1,143	0.5

Table 3
CMS comorbidities and complication occurrence in cervical and thoracolumbar spine surgery

	Cervical			Thoracolumbar	
Cohort attributes	N (184,969)	% (SD)	Cohort attributes	N (586,651)	% (SD)
Gender			*Gender*		
Male	88,515	48.0	Male	263,775	45.0
Age			*Age*		
Average age at time of surgery (yrs)	-	66 (11)	Average age at time of surgery (yrs,(std))	-	70 (10)
Spine procedure type			*Spine procedure type*		
Cervical procedures			*Thoracolumbar procedures*		
Anterior cervical decompression and fusion (ACDF) single level	20,259	11.0	Posterior thoracic decompression (PTD)	7,672	1.3
ACDF single + instrumentation	19,832	10.7	PTD + instrumentation	771	<1
ACDF single + bone morphogenic protein (BMP)	1,222	0.7	PTD + bone morphogenic protein (BMP)	137	<1
ACDF, multiple level	45,848	24.8	Posterior thoracic decompression and fusion (PTDF)	4,621	0.8
ACDF multiple level + instrumentation	45,035	24.3	PTDF + instrumentation	4,227	0.7
ACDF multiple + BMP	2,838	1.5	PTDF + BMP	957	<1
Anterior cervical corpectomy	20,657	11.2	Posterior lumbar decompression (PLD)	307,520	52.4
ACC + instrumentation	20,051	10.8	PLD + instrumentation	14,844	2.5
ACC + BMP	1,227	0.7	PLD + BMP	4,268	0.7
Posterior cervical decompression (PCD)	25,373	13.7	Posterior lumbar decompression and fusion (PLDF)	285,024	48.6
PCD + instrumentation	3,541	1.9	PLDF + instrumentation	259,622	44.3
PCD + BMP	460	0.2	PLDF + BMP	75,970	12.9
Posterior cervical decompression with fusion (PCDF)	21,372	11.6	Anterior thoracolumbar decompression and fusion (ATCDF)	3,999	0.7
PCDF + instrumentation	20,371	11.0	ATCDF + instrumentation	3,904	0.7
PCDF + BMP	2,653	1.4	ATCDF + BMP	1,065	<1
Total instrumentation	161,385	87.2	Total instrumentation	279,811	47.7
Total BMP	11,544	6.2	Total BMP	80,492	14.0
Primary diagnosis of			*Primary diagnosis of*		
Degenerative disease	181,949	98.4	Degenerative disease	539,718	92.0
Neoplasm	554	<1	Neoplasm	2,816	<1
Trauma	4,439	2.4	Trauma	5,163	<1
Infection	287	<1	Infection	1,290	<1
Other	6,103	3.3	Other	36,372	6.2
Pre-existing comorbidities:			*Pre-existing comorbidities:*		
Pulmonary disorder	39,521	21.4	Pulmonary disorder	99,059	16.9
Neurological disorder/deficit	17,606	9.5	Neurological disorder/deficit	46,454	7.9
Hypercholesterolemia	63,361	34.3	Hypercholesterolemia	223,985	38.2
Smoking	54,638	29.5	Smoking	132,679	22.6
Hypertension	115,304	62.3	Hypertension	387,890	66.1
Cardiac disorder other than hypertension	927	<1	Cardiac disorder other than hypertension	3,677	<1
Diabetes mellitus	45,207	24.4	Diabetes mellitus	139,112	23.7
Cancer	20,103	10.9	Cancer	72,621	12.4
Gastroesophageal disorder	3,427	1.9	Gastroesophageal disorder	10,015	1.7
ETOH/drug use	188	<1	ETOH/drug use	314	<1
Psychiatric disorder	43,485	23.5	Psychiatric disorder	109,201	18.6
Complications (After surgery)			*Complications (After surgery)*		
Overall (any) complication	47,054	25.0	Overall (any) complication	169,019	29.0
Cardiac dysrhythmia	17,582	9.5	Cardiac dysrhythmia	64,949	11.1
Pulmonary	12,083	6.5	Pulmonary	29,320	5.0
Urinary tract infection (UTI)	9,067	4.9	Urinary tract infection (UTI)	38,552	6.6
Neurological complications	7,017	3.8	Neurological complications	23,086	3.9
Congestive heart failure (CHF)	6,286	3.4	Congestive heart failure (CHF)	21,345	3.6
Pneumonia	6,501	3.5	Pneumonia	15,940	2.7
Deep vein thrombosis (DVT)	3,762	2.0	Deep vein thrombosis (DVT)	14,951	2.5
Wound hematoma	2,917	1.6	Wound hematoma	9,255	1.6
Other wound complications	1,499	<1	Other wound complications	8,052	1.4
Myocardial infarction (MI)	2,349	1.3	Myocardial infarction (MI)	8,588	1.5
Renal failure	1,587	<1	Renal failure	5,588	<1
Pulmonary embolism (PE)	12,083	6.5	Pulmonary embolism (PE)	6,247	1.1
Delirium	1,829	<1	Delirium	6,857	1.2

Table 4
The discrimatory performance (AUC) of the prediction models for each adverse event

Adverse event	Training data	Validation data
Pulmonary complication	0.76	0.75
Congestive heart failure	0.75	0.75
Pneumonia	0.74	0.74
Urinary tract infections	0.71	0.71
Neurologic complication	0.7	0.69
Cardiac dysrthmia	0.72	0.72
Overall adverse event	0.7	0.7
Overall medical complication	0.71	0.7
Overall surgical complication	0.69	0.69

that patients who received Medicaid have 26% higher odds of developing an overall AE compared with non-Medicaid recipients. Also, the results show that preoperative infection is strongly associated with increased odds of developing a postoperative overall AE compared with those without preoperative infection. The performance of the prediction models based on LASSO was similar, but a bit lower compared with those using logistic regression, with an overall AUC of 0.69.

Prediction for individual AEs

We fitted separate prediction models for each of the following top most frequent AEs in our data—pulmonary complication, cardiac dysrhythmia, urinary tract infection, neurological complication, pneumonia, and congestive heart failure using logistic regression on training data. The highest AUC was observed for the prediction model for a pulmonary complication (AUC 0.76) (See Fig. 3). The performance of the model for a pulmonary complication showed that the brier score was 0.04, and sensitivity and specificity were 82% and 52%. respectively using the median predicted probability value of 0.033 as a threshold (see Supplemental Table S6). The next highest performance was shown in congestive heart failure symptoms (AUC 0.75) and pneumonia (AUC 0.74). The results for the performance of all individual AEs are shown in Table 4**.** Medicaid status was one of the most important factors in predicting individual AEs after spine surgery, including the following AE outcomes: congestive heart failure symptoms (OR 1.6; p<.0001), pneumonia (OR 1.42; p<.0001), and pulmonary complication (OR 1.28; p<.0001). The results based on LASSO were similar to those using logistic regression, with AUCs equal to or a bit lower than those using logistic regression. The comparisons between logistic regression and LASSO for each AE event are shown in (Supplemental Table S5).

Validation and sensitivity analysis

We used the remaining dataset (30%) for validation to evaluate the performance of the prediction models. Overall, the AUCs for validation analysis were comparable to those based on training data (see Table 4). We also conducted sensitivity analysis by redefining the overall AE variable (ie, medical related overall AE) to exclude surgical related AEs, and the prediction result for this outcome (AUC 0.71) was similar to the initial overall AE (AUC 0.7) (Table 4). Similarly, the prediction model for an overall surgical related AE showed AUC of 0.69.

The simulation study to examine the impact of the sample size on the performance of the prediction model (see

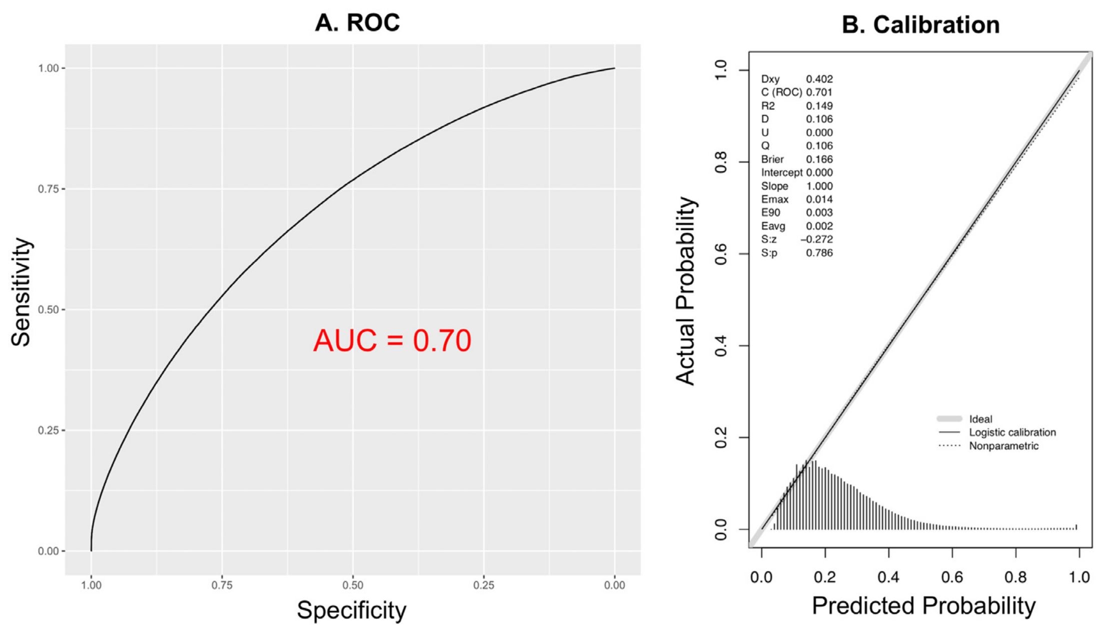

Fig. 1. The performance of the prediction model for the overall adverse event. The panel A shows the ROC curve and corresponding AUC value and the panel B shows the calibration plot of the prediction model for the overall adverse event. The training data set was used to obtain the results.

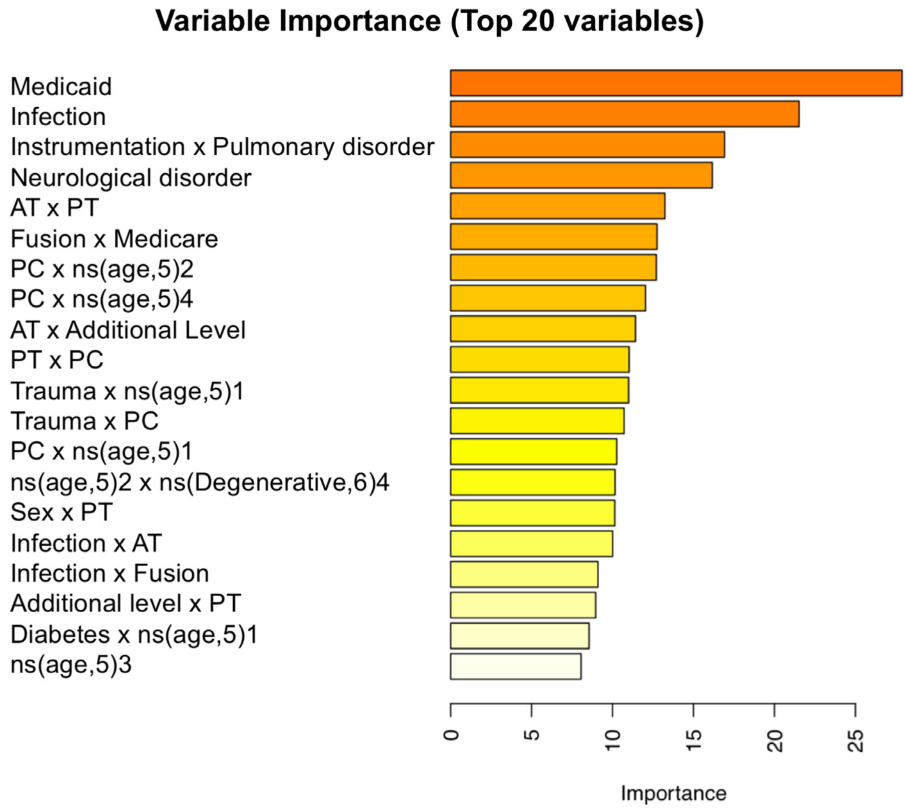

Fig. 2. Variable importance for the prediction model for the overall AE. Importance was calculated by taking an absolute value of t-statistic. The training data set was used to obtain the results.

Methods and Supplemental Figure S1) showed that by increasing the sample size from N=10,000 to 400,000, AUC of the prediction model increased from 0.67 to 0.70. AUC showed a plateau effect beyond a sample size of >400,000, with no improvement in accuracy as we increased sample size.

Comparison to the benchmark model

To examine how our new model for an overall AE that incorporates more diverse study subjects compare to the benchmark model, which is based solely upon a MKS database without elderly patients in CMS, we applied the earlier

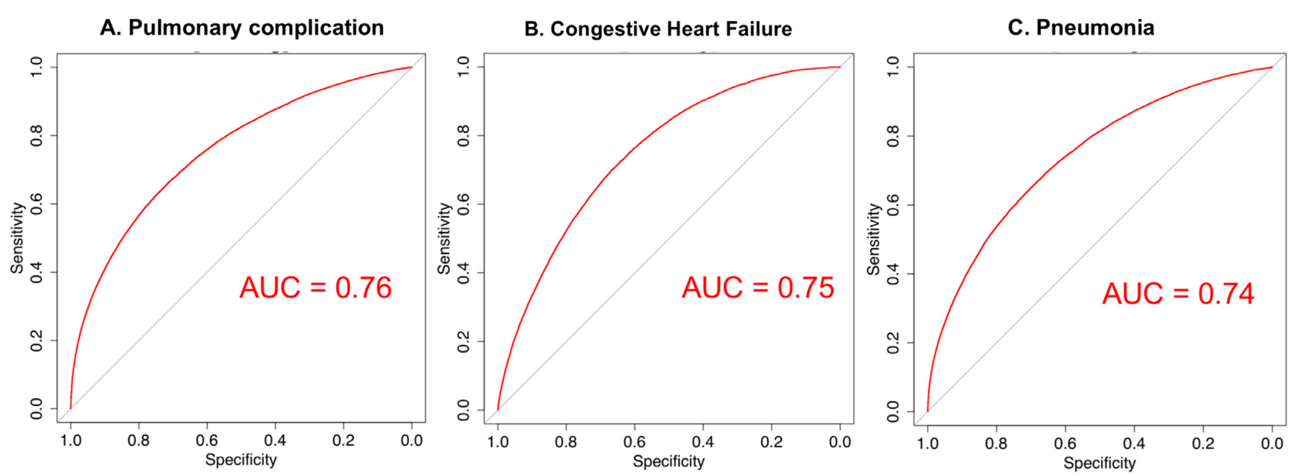

Fig. 3. The performance of the prediction models of the three adverse events selected based on the top three high-ranking AUC values. All models are fitted using the training data. The validation results are shown in Table 2.

model to the entire data (based both on MKS and CMS cohort). Overall AUC for the earlier model applied to the new dataset was 0.6 (compared with 0.7 using the new model) and the calibration plot showed that the risk of an overall AE is underestimated (intercept: −0.63; and slope: 0.301).

Discussion

This study presents a set of predictive models for an overall AE and the six most common individual AEs following spine surgery using data from over one million patients representing a variety of insurance types (private employer-based insurance, Medicare, and Medicaid). Compared with existing studies [7,13−16], this approach encompasses a wide array of patients to better reflect the population of patients undergoing spine surgery in the US. The predictive models for AEs built based on this data showed greater accuracy versus the previous models, with AUC ranging between 0.7 and 0.76, which account for patient-, diagnosis-, and procedure-related factors.

Patient socioeconomic status, as evidenced by Medicaid status, has a significant impact on AE occurrence. Our analysis showed that Medicaid status was the single most important factor in predicting the risks of various AEs after spine surgery (the range of OR of 1.24−1.6 with $p<.001$). These results indicate potential targets for quality improvement and further investigation. The finding that Medicaid status is strongly associated with AEs indicates that this is a population that may benefit from targeted interventions to reduce AEs. Medicaid patients had comorbidity rates significantly higher than non-Medicaid patients. The strong association between Medicaid and postoperative AEs in our study is consistent with previous findings in the spine surgery literature [17,18]. Some of these studies used a prospective registry and found Medicaid status to be associated with higher postoperative AEs (OR 1.7, $p=.001$) [17,18].

Although there are several existing studies that explored the predictive models for AEs following spine surgery [7,13−16], they have several limitations. For example, Bekelis et al. developed the predictive models for AEs based on the National Surgical Quality Improvement Program. However, a major limitation of the study based on National Surgical Quality Improvement Program is that the study population is young, with an average age of 55.7 that do not well capture elderly patients. In addition, the sample size of their study is moderate (N<14,000) compared with ours (N>1 million), and their models do not incorporate Medicaid status that has shown to be one of the most important factors in predicting AEs following spine surgery. In addition, the prediction models in our prior work focused on predicting overall AEs, and the models for predicting the risks of individual adverse events (eg, pulmonary complication) had limited discriminative accuracy [7]. When we applied our original model, developed from a relatively homogenous set of MKS-only patients, to our new dataset of MKS, CMS, and Medicaid patients, we found its performance was much worse in the more heterogenous population. Given that this heterogeneous population better reflects the patient population undergoing spine surgery procedures, the new models we developed will be more generalizable and better reflect AE occurrence observed in a real-world setting.

Our current study has several strengths. The risk prediction models we developed can be used to improve risk adjustment when assessing patient populations with varying comorbidities and demographic profiles. These models can inform patient counseling by enabling surgeons to tailor their assessment of risk to individual patient characteristics. By developing models for specific types of complications, we also provide a tool for surgeons and patients to focus on complications they have the highest suspicion or concern for.

Despite the merits of this study, there are some limitations. As with all administrative database studies, this study relies on accuracy of administrative coding of procedures, diagnoses, comorbidities, and complications. Although classically administrative data has been assumed to underestimate complication occurrence, use of longitudinal data decreases this inaccuracy [6]. However, database studies still are predicated on the assumption that postoperative de novo appearance of a diagnosis code in a longitudinal assessment indicates occurrence of an AE. Therefore, errors in coding or failure in preoperative comorbidity capture are sources of potential bias. Complication rates in our administrative data presented here are comparable to other prospective datasets.

Although we have a larger number of observations compared with our previous work and incorporated extensive feature variables such as comorbidity, preoperative diagnosis and surgery procedures, the AUCs of the models still have not exceeded 0.8. Even though the sample size increased to over a million, the number and the domains of the features for outcomes remained the same as before [7], hence limiting the performance of the prediction models. There may be explanatory features not captured in administrative data which contribute to the risk of AEs. This may indicate an intrinsic limitation of the ability of administrative data to predict AE occurrence. Applications of alternative machine-learning approaches (eg, tree-based methods or deep-learning methods) could be also explored to examine if these can help enhance predictive accuracy. To our best knowledge, there is not a well-defined optimal AUC value for predicting adverse outcomes after spine surgery. However, a number of studies reported AUC values ranging from 0.60 to 0.78 for predicting complications after spine surgery that include urinary tract infections, pulmonary embolism, and overall adverse events [7,14].

Given the limited accuracy of the predictive models, they could be primarily used for identifying expected complication rates in a population of patients and hence might be valuable for assessing O:E ratios for AE occurrence in a given practice. Being able to translate from a population-based approach to an individual patient could be more challenging.

We also found that the machine learning approach using LASSO did not show better performance compared with a classical generalized regression based approach, which was also observed in the previous study [7]. This could be because a penalized regression approach such as LASSO tends to perform better over the traditional approaches when the number of observations is relatively small compared with the number of features, which is not the case in our data.

Conclusions

We present predictive models of AE occurrence after spine surgery procedures, integrating multiple administrative claims databases and encompassing privately and publicly insured patients, which provide greater accuracy in predicting the risks of AEs following spine surgery. Our findings can inform patient counseling, risk adjustment, and quality assessment in spine surgery. Identifying variables of importance in predicting AEs may inform targeted interventions for quality improvement. Future directions include the implementation of these prediction models in software or applications, improvement of the granularity of individual AE prediction, and the evaluation of the performance of these models in independent prospective or retrospective studies.

Acknowledgment

The research is supported by AHRQ 1R01HS023800-01A1.

Supplementary materials

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.spinee.2019.06.018.

References

[1] Campbell PG, Yadla S, Malone J, Zussman B, Maltenfort MG, Sharan AD, et al. Early complications related to approach in cervical spine surgery: single-center prospective study. World Neurosurg 2010;74:363–8.

[2] Yadla S, Malone J, Campbell PG, Maltenfort MG, Sharan AD, Harrop JS, et al. Preoperative diagnosis and early complications in thoracolumbar spine surgery: a single center prospective study. J Spinal Disord Tech 2011;24:E16–20.

[3] Yadla S, Malone J, Campbell PG, Maltenfort MG, Harrop JS, Sharan AD, et al. Early complications in spine surgery and relation to preoperative diagnosis: a single-center prospective study. J Neurosurg Spine 2010;13:360–6.

[4] Bazaz R, Lee MJ, Yoo JU. Incidence of dysphagia after anterior cervical spine surgery: a prospective study. Spine (Phila Pa 1976) 2002;27:2453–8.

[5] Yadla S, Malone J, Campbell PG, Nasser R, Maltenfort MG, Harrop JS, et al. Incidence of early complications in cervical spine surgery and relation to preoperative diagnosis: a single-center prospective study. J Spinal Disord Tech 2011;24:50–4.

[6] Veeravagu A, Cole TS, Azad TD, Ratliff JK. Improved capture of adverse events after spinal surgery procedures with a longitudinal administrative database. J Neurosurg Spine 2015;23:374–82.

[7] Ratliff JK, Balise R, Veeravagu A, Cole TS, Cheng I, Olshen RA, et al. Predicting occurrence of spine surgery complications using "big data" modeling of an administrative claims database. JBJS 2016;98:824–34.

[8] Steyerberg E. Clinical prediction models: a practical approach to development, validation, and updating. Springer Science & Business Media; 2008.

[9] Veeravagu A, Li A, Swinney C, Tian Lu, Moraff A, Azad TD, et al. Predicting complication risk in spine surgery: a prospective analysis of a novel risk assessment tool. J Neurosurg Spine 2017;27:81–91.

[10] Tibshirani R. Regression shrinkage and selection via the lasso. J Royal Stat Soc Ser B (Methodological) 1996:267–88.

[11] Golubovsky JL, Ilyas H, Chen J, Tanenbaum JE, Mroz TE, Steinmetz MP. Risk factors and associated complications for postoperative urinary retention after lumbar surgery for lumbar spinal stenosis. Spine J 2018;18:1533–9.

[12] Hastie TJ. Generalized additive models. Statistical models in S. Routledge; 2017. p. 249–307.

[13] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chron Dis 1987;40:373–83.

[14] Bekelis K, Desai A, Bakhoum SF, Missios S. A predictive model of complications after spine surgery: the National Surgical Quality Improvement Program (NSQIP) 2005-2010. Spine J 2014;14:1247–55.

[15] Kimmell KT, Algattas H, Joynt P, Schmidt T, Jahromi BS, Silberstein HJ, et al. Risk modeling predicts complication rates for spinal surgery. Spine (Phila Pa 1976) 2015;40:1836–41.

[16] Passias PG, Diebo BG, Marascalchi BJ, Jalai CM, Horn SR, Zhou PL, et al. A novel index for quantifying the risk of early complications for patients undergoing cervical spine surgeries. J Neurosurg Spine 2017;27:501–7.

[17] Hacquebord J, Cizik AM, Malempati SH, Konodi MA, Bransford RJ, Bellabarba C, et al. Medicaid status is associated with higher complication rates after spine surgery. Spine (Phila Pa 1976) 2013;38:1393–400.

[18] Manoso MW, Cizik AM, Bransford RJ, Bellabarba C, Chapman J, Lee MJ. Medicaid status is associated with higher surgical site infection rates after spine surgery. Spine (Phila Pa 1976) 2014;39:1707–13.