



Development and validation of risk-adjustment models for elective, single-level posterior lumbar spinal fusions

David N. Bernstein¹, Aakash Keswani², Debbie Chi², James E. Dowdell², Samuel C. Overley², Saad B. Chaudhary², Addisu Mesfin¹

¹Department of Orthopaedics and Rehabilitation, University of Rochester Medical Center, Rochester, NY, USA; ²Department of Orthopaedic Surgery, Mount Sinai Hospital, New York, NY, USA

Contributions: (I) Conception and design: DN Bernstein, A Keswani; (II) Administrative support: JE Dowdell, SC Overley, SB Chaudhary, A Mesfin; (III) Provision of study materials or patients: A Keswani; (IV) Collection and assembly of data: DN Bernstein, A Keswani; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Addisu Mesfin, MD. Associate Professor of Orthopaedic Surgery, University of Rochester School of Medicine and Dentistry, 601 Elmwood Ave, Box 665, Rochester, NY 14642, USA. Email: amesfin@gmail.com.

Background: There is a paucity of literature examining the development and subsequent validation of risk-adjustment models that inform the trade-off between adequate risk-adjustment and data collection burden. We aimed to evaluate patient risk stratification by surgeons with the development and validation of risk-adjustment models for elective, single-level, posterior lumbar spinal fusions (PLSFs).

Methods: Patients undergoing PLSF from 2011–2014 were identified in the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP). The derivation cohort included patients from 2011–2013, while the validation cohort included patients from 2014. Outcomes of interest were severe adverse events (SAEs) and unplanned readmission. Bivariate analysis of risk factors followed by a stepwise logistic regression model was used. Limited risk-adjustment models were created and analyzed by sequentially adding variables until the full model was reached.

Results: A total of 7,192 and 4,182 patients were included in our derivation and validation cohorts, respectively. Full model performance was similar for the derivation and validation cohorts in both 30-day SAEs (C-statistic =0.66 vs. 0.69) and 30-day unplanned readmission (C-statistic =0.62 vs. 0.65). All models demonstrated good calibration and fit ($P \geq 0.58$). Intraoperative variables, laboratory values, and comorbid conditions explained >75% of the variation in 30-day SAEs; ASA class, laboratory values, and comorbid conditions accounted for >80% of model risk prediction for 30-day unplanned readmission. Four variables for the 30-day SAE models (age, gender, ASA ≥ 3 , operative time) and 3 variables for the 30-day unplanned readmission models (age, ASA ≥ 3 , operative time) were sufficient to achieve a C-statistic within four percentage points of the full model.

Conclusions: Risk-adjustment models for PLSF demonstrated acceptable calibration and discrimination using variables commonly found in health records and demonstrated only a limited set of variables were required to achieve an appropriate level of risk prediction.

Keywords: Posterior lumbar spinal fusion (PLSF); national database; big data; risk stratification; adverse events; readmission

Submitted Oct 05, 2018. Accepted for publication Dec 11, 2018.

doi: 10.21037/jss.2018.12.11

View this article at: <http://dx.doi.org/10.21037/jss.2018.12.11>

Introduction

Following the passage of the Affordable Care Act (ACA) in 2010, there has been increased focus on delivering quality care at a lower cost, driving the shift from traditional fee-for-service reimbursement structure to a pay-for-performance model, especially for orthopaedic procedures (1-8). To succeed in these programs and maximize value for the patient, orthopaedic providers require data-driven tools to efficiently allocate resources before, during, and after the target procedure. Effectively managing patient care and optimizing long-term outcomes must remain the main focus for orthopaedic surgeons. Risk stratification models can help ensure good patient care and allow for appropriate reimbursement for this clinically complex patient population.

Risk stratification accomplishes the above objectives as it involves an important trade-off between data collection burden and the ability to identify and manage high-risk patients. There are numerous risk-adjustment models used to calculate appropriate capitated payment for services (9-11). However, several studies have demonstrated the limitations of claims-based risk-adjustment models that are focused on preventing short-term complications (9-12). Only a handful of models such as the America Joint Replacement Registry's 90-day infection/1-year revision risk calculator for total hip/knee arthroplasty (THA/TKA) patients utilize clinical data (13,14). The need for additional orthopaedic-centered risk-adjustment models is evident.

Over the past two decades, the volume of elective, single-level, posterior lumbar spinal fusions (PLSFs) performed has increased by 137%, targeting the procedure for bundled care models (15). There are many studies evaluating the link between patient characteristics and provider factors (i.e., surgeon volume, operative technique) on short-term outcomes following PLSF (16-18). However, there is a paucity of literature that seeks to identify a limited-set of "most-predictive" risk factors that can be practically implemented.

The goal of this study was to develop two sets of risk-adjustment models specific to 30-day severe adverse events (SAEs) and unplanned readmission following elective PLSF using a large, nationally representative clinical database. A second aim was to provide a basis for spine surgeons and hospitals to make informed decisions when constructing clinical data-driven risk-adjustment models for PLSF to balance the trade-off between improved risk stratification

accuracy versus clinical data collection burden.

Methods

Study population

CPT codes (22612, 22630, 22633) were used to identify individuals 17 years or older who underwent single-level PLSF with or without use of an interbody device within 2011–2014 American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) data (19). Those with additional CPT codes not corresponding to single-level PLSF were removed, as were patients with pre-operative wound infection, disseminated cancer, or those who underwent emergent spine surgery. Procedures from 2011 to 2013 constituted our derivation cohort while procedures from 2014 were used as a validation cohort for our risk-adjustment models.

Outcome measures

The 30-day SAE and 30-day unplanned readmission were the two outcome variables of interest; 30-day SAE included death, myocardial infarction, cerebrovascular accident, renal failure, pulmonary embolism, venous thromboembolism, sepsis, septic shock, unplanned intubation, paraplegia, deep wound infection, organ/space infection, and return to operating room.

Patient characteristics

Table 1 displays patient-level covariates considered for our PLSF model analyses. These included: age, American Society of Anesthesiologists (ASA) physical status classification, comorbidities, laboratory values, vital signs-based comorbid conditions and intraoperative variables. Vital signs-based comorbid conditions included systemic inflammatory response syndrome (SIRS) and septic shock. Intraoperative variables were comprised of operating time and total hospital length of stay. Laboratory values were denoted as normal or abnormal based on accepted norms (20). Age was also reported as a quadratic with the assumption of improved fit. More information regarding each variable can be found in the ACS NSQIP Participant Use Data File (PUF) (21). A category for patients with missing laboratory data was created, and the outcomes for this group were compared and pooled into the reference group (22).

Table 1 Patient characteristics of derivation [2011–2013] and validation [2014] cohorts

Characteristics	Lumbar fusion [2011–2013] (N=7,192)		Lumbar fusion [2014] (N=4,182)	
	Number of patients	%	Number of patients	%
Demographics				
Age (yr)*	59.2±13.7	–	58.9±13.4	–
Sex				
Male	3,113	43.3	1,924	46.0
Female	4,079	56.7	2,258	54.0
ASA classification				
1 or 2	3,941	54.8	2,189	52.3
3 or 4	3,251	45.2	1,993	47.7
Comorbid conditions				
BMI (kg/m²)				
Greater than 40	565	7.9	340	8.1
Less than 40	6,627	92.1	3,842	91.9
Takes hypertension medications	3,909	54.4	2,302	55.0
Bleeding disorder	99	1.4	65	1.6
History of smoking	1,606	22.3	889	21.3
COPD	285	4.0	207	5.0
Dyspnea				
With exertion	470	6.5	200	4.8
At rest	28	0.4	5	0.1
Transfusion pre-op. (≥1 unit PRBC pre-op.)	11	0.2	2	0.05
Diabetes mellitus	1,152	16.0	737	17.6
Corticosteroid use	254	3.5	152	3.6
Congestive heart failure	18	0.3	12	0.3
Disseminated cancer	–	–	–	–
Dialysis-dependent	6	0.08	6	0.1
Renal failure	5	0.07	–	–
Functional status				
Partially dependent	160	2.2	69	1.7
Dependent	6	0.08	1	0.02
Laboratory results within 90 days pre-op. (%)				
Low WBC count (<4,500/mcL)	354	4.9	202	4.8
High WBC count (>10,000/mcL)	716	10.0	443	10.6
Low hematocrit (<30%)	73	1.0	41	1.0
Low platelets (<150,000/mcL)	917	12.8	598	14.3
High INR (>1.1)	235	3.3	115	2.7
Low sodium (<135 mEq/L)	336	4.7	227	5.4
High sodium (>145 mEq/L)	53	0.7	35	0.8

Table 1 (continued)

Table 1 (continued)

Characteristics	Lumbar fusion [2011–2013] (N=7,192)		Lumbar fusion [2014] (N=4,182)	
	Number of patients	%	Number of patients	%
High creatinine (>1.3 mg/dL)	349	4.9	188	4.5
High blood urea nitrogen (>30 mg/dL)	185	2.6	105	2.5
High bilirubin (>1.9 mg/dL)	20	0.3	5	0.1
Low albumin (<3.4 g/dL)	95	1.3	68	1.6
Vital signs-based comorbid conditions				
SIRS	–	–	–	–
Sepsis	–	–	–	–
Septic shock	–	–	–	–
Intraoperative variables				
Operating time*	195.0±92.2	–	191.7±96.1	–
Outcomes, 30-day				
Mortality rate	10	0.1	4	0.1
Severe adverse event	383	5.3	196	4.7
Minor adverse event	218	3.0	110	2.6
Unplanned readmission	345	4.8	178	4.3
Outcomes, specific adverse events, 30-day				
Urinary tract infection	106	1.5	52	1.2
Pneumonia	47	0.7	26	0.6
Sepsis	50	0.7	26	0.6
Septic shock	12	0.2	6	0.1
Re-intubation	24	0.3	9	0.2
Myocardial infarction	24	0.3	17	0.4
Surgical site infection				
Superficial	69	1.0	34	0.8
Deep wound	52	0.7	28	0.7
Organ space	15	0.2	12	0.3
Deep vein thrombosis	52	0.7	32	0.8
Cardiac arrest requiring CPR	8	0.1	3	0.07
On ventilator >48 hours	13	0.2	4	0.1
Renal insufficiency				
Progressive	8	0.1	5	0.1
Acute	10	0.1	2	0.05
Cerebral vascular accident	11	0.2	5	0.1
Wound dehiscence	21	0.3	15	0.4
Return to OR	238	3.3	123	2.9

*, age and operating time are given as the mean and the standard deviation. PRBC, packed red blood cells; CPR, cardiopulmonary resuscitation; OR, operating room.

Bivariate analysis

Bivariate analysis of risk factors was conducted for each of the outcome variables. *Table 2* displays odds ratios of the statistically significant patient-level covariates for the derivation cohort and the validation cohort.

Specification of full risk-adjustment models

Variables with a P value ≤ 0.10 in bivariate analysis were included in separate stepwise regression models for 30-day SAE and 30-day unplanned readmission. All variables that entered and exited the model with a P value of 0.10 were retained, and those that were statistically significant are shown in *Table 3*.

Evaluation and validation of full risk-adjustment models

The models were evaluated and compared for predictive performance (discrimination) and goodness-of-fit (calibration) (*Table 4*). C-statistics described the predictive value of each model while results from the Hosmer-Lemeshow test were used to assess calibration of the models. Derivation and validation models were evaluated and compared against each other to assess overall model performance and consistency.

Predictor contribution to full risk-adjustment models

We determined the relative contribution of each variable to each model by sequentially removing variables one at a time (*Table 4*). The exact contribution was calculated by analyzing the change in the model log-likelihood value when a given variable was removed. Variable categories included age, sex, ASA classification, comorbid conditions, laboratory values and intraoperative variables.

Limited risk-adjustment models and prediction value

All models were analyzed using fewer covariates and built up sequentially to the full model to understand the predictive value of models without the full complement of clinical data (*Table 5*). Covariates were added back starting with ASA classification and progressing through the addition of comorbid conditions, laboratory values, age and intraoperative variables until the full model was

formed. Prediction performance was based on C-statistics and calculations of the continuous net reclassification improvement [NRI (>0)] (23).

Results

Patient characteristics, intra-operative variables, and 30-day outcomes

The derivation and validation cohorts consisted of 7,192 and 4,182 patients respectively and were similar in terms of demographics, characteristics, clinical comorbidities, pre-operative laboratory data, and intraoperative variables (*Table 1*). There were no statistically significant differences in post-operative 30-day outcomes.

Bivariate analysis

Both validation and derivation cohorts identified age, ASA 3/4, BMI >40 , hypertension, bleeding-causing disorders, diabetes mellitus, corticosteroid use, dependent functional status, low hematocrit, high INR, high creatinine, high BUN, and low albumin as factors associated with SAEs or unplanned 30-day readmissions ($P < 0.05$) (*Table 2*).

Multivariate risk-adjustment models

Stepwise logistic regression in the derivation cohort yielded models with 12 and 11 significant independent predictors of 30-day SAEs and unplanned readmission (*Table 3*) while 12 and 10 predictors were generated for the validation cohort. The derivation and validation models for 30-day SAEs both identified age, BMI >40 , and high INR as significant predictors while those for 30-day unplanned readmission identified ASA 3/4 and BMI >40 (*Table 3*).

Evaluation of risk-adjustment model performance

Model performance was similar for the derivation and validation cohorts (*Table 4*). The risk models were more predictive of 30-day SAEs than for 30-day unplanned readmission. The C-statistics for models predicting 30-day SAEs were 66.1% (derivation cohort) and 68.5% (validation cohort), while those for unplanned readmission were 61.6% (derivation cohort) and 65.3% (validation cohort). All models demonstrated good calibration and fit ($P \geq 0.58$ for all).

Table 2 Bivariate analysis of risk factors for severe adverse events or unplanned readmission

Outcome/risk factor	Derivation cohort [2011–2013] (N=7,192)				Validation cohort [2014] (N=4,182)			
	Severe adverse event		Unplanned readmission		Severe adverse event		Unplanned readmission	
	Odds ratio (95% CI)	P value	Odds ratio (95% CI)	P value	Odds ratio (95% CI)	P value	Odds Ratio (95% CI)	P value
Demographics								
Age	0.99 (0.98–1.00)	0.009	0.99 (0.98–1.00)	0.002	0.98 (0.97–0.99)	0.0006	0.97 (0.96–0.99)	<0.0001
ASA classification								
3 or 4	1.80 (1.46–2.22)	<0.0001	1.53 (1.23–1.91)	0.0001	1.70 (1.27–2.28)	0.0003	2.07 (1.51–2.83)	<0.0001
Comorbid conditions								
BMI (>40)	1.88 (1.38–2.56)	0.0002	1.77 (1.28–2.47)	0.001	1.70 (1.10–2.63)	0.02	2.01 (1.31–3.10)	0.003
Takes hypertension medications	1.21 (0.98–1.49)	0.08	1.26 (1.01–1.57)	0.04	1.53 (1.14–2.07)	0.005	1.37 (1.01–1.87)	0.05
Bleeding disorder	2.02 (1.04–3.92)	0.04	1.52 (0.70–3.30)	0.34	2.51 (1.13–5.57)	0.03	1.49 (0.55–4.13)	0.36
COPD	1.93 (1.28–2.90)	0.004	1.89 (1.22–2.91)	0.007	1.03 (0.54–1.99)	0.87	1.28 (0.68–2.40)	0.48
Dyspnea with exertion	1.37 (0.95–1.99)	0.11	1.72 (1.20–2.46)	0.005	1.57 (0.89–2.76)	0.12	2.36 (1.42–3.93)	0.002
Diabetes mellitus	1.66 (1.30–2.12)	0.0001	1.49 (1.14–1.94)	0.004	1.99 (1.45–2.74)	<0.0001	1.62 (1.14–2.29)	0.009
Corticosteroid use	2.10 (1.38–3.20)	0.001	1.45 (0.87–2.39)	0.18	1.62 (0.86–3.05)	0.17	2.81 (1.63–4.83)	0.0006
Dialysis-dependent	3.56 (0.42–30.57)	0.28	1.00 (1.00–1.00)	0.99	10.26 (1.87–56.37)	0.03	22.86 (4.58–114.09)	0.001
Functional status								
Partially dependent	2.02 (1.19–3.42)	0.02	1.78 (1.00–3.18)	0.06	1.97 (0.84–4.60)	0.14	1.78 (0.71–4.48)	0.22
Dependent	8.93 (1.63–48.91)	0.04	9.98 (1.82–54.65)	0.03	1.01 (1.00–1.02)	0.05	1.00 (1.00–1.00)	0.99
Laboratory results within 90 days pre-op. (%)								
High WBC count (>10,000/mcL)	1.56 (1.16–2.10)	0.005	1.38 (1.00–1.91)	0.05	0.90 (0.56–1.46)	0.81	1.14 (0.71–1.81)	0.62
Low hematocrit (<30%)	2.87 (1.46–5.64)	0.005	1.15 (0.42–3.18)	0.78	3.56 (1.48–8.58)	0.01	1.16 (0.28–4.82)	0.69
High INR (>1.1)	1.99 (1.28–3.10)	0.004	1.58 (0.95–2.61)	0.09	1.33 (0.61–2.90)	0.50	1.48 (0.68–3.22)	0.34
Low sodium (<135 mEq/L)	1.39 (0.91–2.14)	0.130	1.13 (0.69–1.84)	0.6	1.83 (1.10–3.03)	0.02	1.65 (0.95–2.84)	0.09
High creatinine (>1.3 mg/dL)	2.53 (1.79–3.57)	<0.0001	1.72 (1.14–2.59)	0.01	2.58 (1.59–4.21)	0.0005	2.89 (1.77–4.72)	0.0001
High blood urea nitrogen (>30 mg/dL)	2.90 (1.88–4.48)	<0.0001	2.20 (1.34–3.63)	0.004	2.20 (1.13–4.29)	0.03	3.04 (1.63–5.66)	0.002
Low albumin (<3.4 g/dL)	4.01 (2.35–6.84)	<0.0001	2.65 (1.40–5.02)	0.006	5.08 (2.72–9.46)	<0.0001	4.56 (2.35–8.86)	0.0001
Intraoperative variables								
Operating time	1.00 (1.00–1.00)	<0.0001	1.00 (1.00–1.00)	0.002	1.00 (1.00–1.00)	0.001	1.00 (1.00–1.00)	0.99

Table 3 Risk-adjustment model results based off variables that met criteria for stepwise selection (PR0.10 for entry and exit)

Outcome/risk factor	Derivation cohort [2011–2013] (N=7,192), odds ratio (95% CI)		Validation cohort [2014] (N=4,182), odds ratio (95% CI)	
	Severe adverse event	Unplanned readmission	Severe adverse event	Unplanned readmission
Demographics				
Age		1.01 (1.00–1.02)*		
Age Sqr	1.00 (1.00–1.00)		1.00 (1.00–1.00)*	1.00 (1.00–1.00)*
ASA classification				
3 or 4	1.33 (1.12–1.57)*	1.19 (1.00–1.42)*	1.22 (0.98–1.53)	1.33 (1.04–1.69)*
Comorbid conditions				
BMI (>40)	1.45 (1.14–1.85)*	1.33 (1.02–1.72)*	1.95 (1.45–2.64)*	2.13 (1.55–2.92)*
Bleeding disorder			1.83 (1.00–3.33)*	
History of smoking			1.49 (1.16–1.91)*	1.33 (1.01–1.76)*
COPD		1.56 (1.15–2.12)*		
Dyspnea with exertion		1.29 (0.99–1.69)		1.50 (1.03–2.18)*
Diabetes mellitus	1.21 (1.00–1.47)*	1.19 (0.97–1.45)*	1.29 (1.01–1.63)	
Corticosteroid use				2.03 (1.37–3.00)*
Congestive heart failure			2.93 (1.07–8.07)*	
Disseminated cancer	2.42 (1.27–4.59)*		3.93 (1.86–8.29)*	3.26 (1.45–7.29)*
Dialysis-dependent			4.26 (1.39–13.11)	6.22 (2.13–18.21)*
Functional status, partially dependent	1.62 (1.14–2.31)*		1.64 (0.98–2.73)	
Laboratory results within 90 days pre-op. (%)				
Low WBC count (<4,500/mcL)		1.37 (0.99–1.82)*		
High WBC count (>10,000/mcL)	1.63 (1.32–2.02)*	1.43 (1.14–1.79)*		
Low hematocrit (<30%)	1.51 (0.93–2.46)		2.04 (1.13–3.68)*	
Low platelets (<150,000/mcL)	1.53 (1.14–2.06)*			
High INR (>1.1)	1.43 (1.03–1.98)*			
Low sodium (<135 mEq/L)		1.35 (0.99–1.83)	1.70 (1.21–2.38)*	
High creatinine (>1.3 mg/dL)	1.37 (1.00–1.87)*	1.41 (1.06–1.89)*	1.63 (1.14–2.34)*	1.63 (1.11–2.38)*
High bilirubin (>1.9 mg/dL)	4.09 (1.61–10.37)*	5.05 (2.11–12.08)*	7.14 (1.32–38.70)*	
Low albumin (<3.4 g/dL)	1.62 (1.10–2.39)*	2.05 (1.39–3.01)*	2.80 (1.79–4.39)*	2.03 (1.22–3.39)*
Intraoperative variables				
Operating time	1.00 (1.00–1.00)*	1.00 (1.00–1.00)*	1.00 (1.00–1.00)*	

*, indicates significant P value <0.05.

Table 4 Full risk-adjustment model performance and the relative contribution of predictors

Variable	Severe adverse event		Unplanned readmission	
	Derivation cohort, N=7,192	Validation cohort, N=4,182	Derivation cohort, N=7,192	Validation cohort, N=4,182
C-statistic (full model)	66.1%	68.5%	61.6%	65.3%
Pr χ^2 for Hosmer-Lemeshow goodness-of-fit test [†]	0.78	0.62	0.77	0.58
Relative contribution of predictors [‡]				
Age	0.04	0.09	0.08	0.20
Sex	0.01	0.01	–	–
ASA classification (3/4 vs. 1/2)	0.16	0.13	0.12	0.22
Intraoperative variables	0.28	0.11	0.10	0.01
Laboratory values	0.34	0.51	0.50	0.33
Comorbid conditions	0.17	0.15	0.20	0.24

[†]Pr χ^2 = probability of chi-squared statistic; df = degrees of freedom. A P value of >0.05 for the Hosmer-Lemeshow goodness-of-fit test demonstrates good model calibration; a P value of <0.05 for the Hosmer-Lemeshow goodness-of-fit test suggests a poor model fit. [‡]We assessed the relative contribution of variables grouped into different categories by examining changes in the model C-statistic when each variable category was retained and then removed from the full model.

Relative contribution of predictors to risk-adjustment models

Table 4 displays the risk-adjustment contribution to the C-statistic of each predictor group. Intraoperative variables, laboratory values, and comorbid conditions explained >75% of the variation in 30-day SAEs for both the validation and derivation cohorts. For 30-day unplanned readmission, ASA class, laboratory values, and comorbid conditions accounted for >80% of model risk prediction in the derivation cohort while age provided a 20% contribution in the validation cohort.

Table 5 demonstrates the change in performance of the risk models with sequential addition of each of the variable group (comorbid conditions, labs, intraoperative, age, gender, ASA 3/4) starting with comorbid conditions. This process was also repeated but in reverse order. For the 30-day SAE derivation and validation models, four variables (age, gender, ASA 3/4, operative time) were sufficient to achieve a C-statistic within 4 percentage points of the full model (16 variables). These four variables were also sufficient to achieve a C-statistic within 2 percentage points of models in using only labs, comorbidities, and operative time (derivation model C-statistic: 0.63; validation model C-statistic: 0.65) (Table 5). Among the aforementioned

four variables, ASA 3/4 and operative time improved risk prediction (NRI >0) via reclassification of events and nonevents while gender only improved reclassification of nonevents (Table 5). Among the 30-day unplanned readmission models, three variables (age, ASA 3/4, operative time) were sufficient for achieving a C-statistic within 4 percentage points of the full model (13 variables). These three variables also achieved a C-statistic within 0–3 points of models using only labs, comorbidities, and operative time (derivation model C-statistic: 0.58 validation model: 0.62) (Table 5). In both derivation and validation models, operative time improved risk prediction (NRI >0) via reclassification of nonevents.

Discussion

As the US healthcare landscape continues to shift towards a value-based payment system, the need for adequate risk-adjustment in making meaningful outcomes comparisons has increased. The great variability within the patient population must be taken into account to ensure hospitals are not penalized for treating more complicated patients. Refining our understanding of the most predictive risk factors of adverse events can guide efforts to modify these factors pre-operatively. We created a set of PLSF risk-

Table 5 Any adverse outcome or unplanned readmission model performance with sequential addition of variable groups and measures of prediction increment

Variable	No. of data elements	C-statistic (%)	Event NRI (>0) (%)	Nonevent NRI (>0) (%)	Combined NRI (>0)	P value, HL test
Severe adverse event-models [§] (derivation cohort)						
Age	1	0.54	–	–	–	0.87
+ Gender	2	0.55	15	–11	0.04	0.29
+ ASA class 3/4	3	0.58	23	6	0.29	0.79
+ Intraoperative variables	4	0.63	1	30	0.31	0.12
+ Laboratory values	12	0.66	–26	54	0.28	0.97
+ Comorbid conditions (full model)	16	0.66	–33	53	0.20	0.78
Severe adverse event-models [§] (validation cohort)						
Age	1	0.57	–	–	–	0.06
+ Gender	2	0.58	12	–8	0.04	0.20
+ ASA class 3/4	3	0.61	32	1	0.33	0.24
+ Intraoperative variables	4	0.65	2	29	0.31	0.20
+ Laboratory values	12	0.68	–29	59	0.30	0.06
+ Comorbid conditions (full model)	16	0.69	–24	53	0.29	0.62
Unplanned readmission-models [§] (derivation cohort)						
Age	1	0.54	–	–	–	0.86
+ Gender	–	–	–	–	–	–
+ ASA class 3/4	2	0.57	17	5	0.22	0.85
+ Intraoperative variables	3	0.58	–12	23	0.11	0.06
+ Laboratory values	9	0.61	–29	52	0.23	0.98
+ Comorbid conditions (full model)	13	0.62	–29	43	0.14	0.77
Unplanned readmission-models [§] (validation cohort)						
Age	1	0.59	–	–	–	0.10
+ Gender	–	–	–	–	–	–
+ ASA class 3/4	2	0.62	33	0	0.33	0.85
+ Intraoperative variables	3	0.62	–10	20	0.10	0.49
+ Laboratory values	9	0.64	–38	58	0.20	0.30
+ Comorbid conditions (full model)	13	0.65	–47	68	0.21	0.58

[§], as groups of variables were sequentially added. A P value of >0.05 for the Hosmer-Lemeshow goodness-of-fit test demonstrates good model calibration; a P value of <0.05 for the Hosmer-Lemeshow goodness-of-fit test suggests a poor model fit.

adjustment models to help address these issues for a high-volume procedure using a large, nationally representative database. The models for 30-day SAEs and for 30-day unplanned readmission aim to identify the most predictive clinical variables for risk adjustment in PLSF patients and assess the predictive ability of specific combinations of variable groups (13).

Performance of full and limited set risk models

Our model results indicate that it is possible to provide acceptable risk-adjustment for SAEs and unplanned readmission within thirty days following PLSF using only the most predictive variables. The full derivation and validation models provide moderate patient-level discrimination (C-statistic >0.65) for SAEs and unplanned readmission across the full spectrum of patient risk. Previous work using the ACS-NSQIP database to develop and validate risk-adjustment models for HFR, THA, and TKA noted similar discrimination for SAEs (C-statistic >0.60) compared to the model we present (14), and other specialties have likewise noted acceptable predictive ability for morbidity (C-statistic >0.70) in risk-adjustment models for five common general surgery procedures.

In contrast, we found laboratory values comprised seven of the 10 variables with the highest standardized coefficients for risk of SAE/unplanned readmission in PLSF patients. Multiple database and retrospective studies have noted the importance of nutritional status, liver disease, and pre-operative infection on short-term complication risk after lumbar and cervical fusion (24-26). The remaining three variables of the 10 most predictive in our models were clinical comorbidities (such as diabetes, COPD, BMI >40), which is consistent with the literature (13,24).

Our analysis of the relative contribution of each variable group revealed that age, ASA class 3/4, and gender collectively contributed 21–23% explanatory value to our four PLSF models, although this is less than the 38–55% explanatory value contributed in TKA, THA, and HF models (14). However, another study found a similar discrimination as compared to the full models, albeit with different variables (13). In our limited-set models, the comorbid condition, laboratory value, and intraoperative variable groups (Table 5) achieved discrimination within 1 percentage point of the full models.

Implications for risk models in spine surgery

The implications of our findings are best understood when taking into consideration both data collection burden and unexplained variation in 30-day SAEs/readmissions for spine surgery. Hospitals are generally unable to leverage existing healthcare information technology infrastructure to reduce duplication of clinician data collection efforts and respondent burden (27). Thus, most risk-adjustment models for readmission and mortality have been based on claims data because they are easily available and provide a longitudinal view of outcomes and resource utilization (27). Limitations of claims data include incorrect recording of diagnoses, under-coding of clinical complications, and can lead to bias and inadequate risk-adjustment. However, they must be weighed against the labor and financial costs of clinical data collection. Since claims provide a total cost of care perspective, the ideal data set would integrate claims and clinical data to provide the most accurate longitudinal view of outcomes and cost.

Our analysis suggests that laboratory values and comorbid conditions individually provide the most explanatory value to the PLSF model's predictive ability. Age, gender, and operative time are readily available in the HER, facilitating input into a clinical registry/risk model. In contrast, collection of comorbid conditions is challenging because most registries require a trained clinician to extract these data from the chart. As was shown in our analysis of limited set PLSF models, the models with laboratory values, clinical comorbidities, and operative time achieved a 2–3 percentage point higher discrimination as compared to the model with age, gender, ASA class, and operative time.

Our full and limited set PLSF models for SAEs and unplanned readmission were adequate (C-statistic >0.65) but less robust than other published ACS-NSQIP models like those predicting 30-day mortality (C-statistic >0.90) (13,14). First, we recognize that SAE includes 15 different adverse events and that unplanned readmission in PLSF patient can be due to numerous reasons. One explanation for only moderate patient-level discrimination is that individual adverse event and readmission types are affected differently by different variables. We hypothesize that other important data elements from literature such as provider-related factors (surgical approach, surgeon volume, hospital quality performance, perioperative protocols), patient-reported outcomes (PROS, pre-operative pain, function, quality

of life, mental health status), and psychosocial factors (socioeconomic status, insurance status, home support) may all play important roles in addressing the unexplained variation in a patient's risk for 30-day complications, as shown in a recent model for six different conditions/procedures (including THA/TKA) that demonstrated a 45-point increase in explanatory value of the overall model via the addition of sociodemographic status (SDS) (19).

There are several limitations inherent to large databases, one of which is missing data. An analysis of 80,000 spine surgery patients in the ACS-NSQIP found that 5% were missing demographic data, 72% were missing comorbidities, and 80% were missing at least one laboratory value (23). The authors developed three different approaches for handling the missing values leading to differences in variables that entered risk models and beta coefficients for those variables. In our analysis, we excluded patients missing critical data elements (such as age, gender, ASA class, operative time). Since ~70% of our PLSF cohort were missing at least 1 laboratory value, we followed a pooled approach for replacing the missing values similar to that of other similar published analyses. A second limitation is that the ACS-NSQIP focuses on general surgery cases and is built for optimizing care of those procedures. A spine specific registry that captures more relevant perioperative clinical data and extends the post-procedure period to 90 or 180 days may prove more beneficial for PLSF patients (13,14). Another limitation of our work is that we only took patient-specific (e.g., BMI) or operating room-specific (e.g., operating time) risk factors into account. Additional risk factors, such as insurance status and ethnicity/race, may also impact our findings; while these are not modifiable by the surgeon preoperatively, such knowledge could assist surgeons in delivering care. Future work can seek to analyze the impact of these factors on 30-day SAEs and unplanned readmission in elective, single-level PLSF cases. Lastly, we are inherently limited by the accuracy of the data entered into ACS NSQIP. However, previous spine research has suggested that ACS NSQIP is better suited for adverse-event studies compared to large claims databases [e.g., nationwide inpatient sample (NIS)] (28).

In summary, the goal of this study was to improve the design of risk-adjustment models in spine surgery, including those used prospectively in healthcare delivery and retrospectively in alternative payment models. We believe our analysis demonstrates the important trade-offs physicians, hospitals, and payers/employers must take into account when deciding which data to include in risk models

for high-volume, relatively homogenous procedures such as PLSF. Future work can evaluate whether alternative methods of developing risk-adjustment models—such as those created by machine learning—perform better than more traditional statistical approaches—such as those presented in the current study.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: ACS NSQIP is a publicly available database with no patient identifiers; thus, IRB approval is not needed for its use.

References

1. Bhattacharyya T, Freiberg AA, Mehta P, et al. Measuring the report card: the validity of pay-for-performance metrics in orthopedic surgery. *Health Aff (Millwood)* 2009;28:526-32.
2. Bhattacharyya T, Mehta P, Freiberg AA. Hospital characteristics associated with success in a pay-for-performance program in orthopaedic surgery. *J Bone Joint Surg Am* 2008;90:1240-3.
3. Bosko T, Dubow M, Koenig T. Understanding Value-Based Incentive Models and Using Performance as a Strategic Advantage. *J Healthc Manag* 2016;61:11-4.
4. Burwell SM. Setting value-based payment goals--HHS efforts to improve U.S. health care. *N Engl J Med* 2015;372:897-9.
5. Lansky D, Nwachukwu BU, Bozic KJ. Using financial incentives to improve value in orthopaedics. *Clin Orthop Relat Res* 2012;470:1027-37.
6. Wei DH, Hawker GA, Jevsevar DS, et al. Improving Value in Musculoskeletal Care Delivery: AOA Critical Issues. *J Bone Joint Surg Am* 2015;97:769-74.
7. Bozic KJ, Ward L, Vail TP, et al. Bundled payments in total joint arthroplasty: targeting opportunities for quality improvement and cost reduction. *Clin Orthop Relat Res* 2014;472:188-93.
8. Iorio R, Clair AJ, Inneh IA, et al. Early Results of Medicare's Bundled Payment Initiative for a 90-Day

- Total Joint Arthroplasty Episode of Care. *J Arthroplasty* 2016;31:343-50.
9. Buchner F, Goepffarth D, Wasem J. The new risk adjustment formula in Germany: implementation and first experiences. *Health Policy* 2013;109:253-62.
 10. Chang RE, Lai CL. Use of diagnosis-based risk adjustment models to predict individual health care expenditure under the National Health Insurance system in Taiwan. *J Formos Med Assoc* 2005;104:883-90.
 11. Donato R, Richardson J. Diagnosis-based risk adjustment and Australian health system policy. *Aust Health Rev* 2006;30:83-99.
 12. Lawson EH, Louie R, Zingmond DS, et al. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. *Ann Surg* 2012;256:973-81.
 13. Dimick JB, Osborne NH, Hall BL, et al. Risk adjustment for comparing hospital quality with surgery: how many variables are needed? *J Am Coll Surg* 2010;210:503-8.
 14. Schilling PL, Bozic KJ. Development and Validation of Perioperative Risk-Adjustment Models for Hip Fracture Repair, Total Hip Arthroplasty, and Total Knee Arthroplasty. *J Bone Joint Surg Am* 2016;98:e2.
 15. Rajae SS, Bae HW, Kanim LE, et al. Spinal fusion in the United States: analysis of trends from 1998 to 2008. *Spine (Phila Pa 1976)* 2012;37:67-76.
 16. Lee NJ, Kothari P, Phan K, et al. Incidence and Risk Factors for 30-Day Unplanned Readmissions After Elective Posterior Lumbar Fusion. *Spine (Phila Pa 1976)* 2018;43:41-8.
 17. Pugely AJ, Martin CT, Gao Y, et al. Causes and risk factors for 30-day unplanned readmissions after lumbar spine surgery. *Spine (Phila Pa 1976)* 2014;39:761-8.
 18. Su AW, Habermann EB, Thomsen KM, et al. Risk Factors for 30-Day Unplanned Readmission and Major Perioperative Complications After Spine Fusion Surgery in Adults: A Review of the National Surgical Quality Improvement Program Database. *Spine (Phila Pa 1976)* 2016;41:1523-34.
 19. Nagasako EM, Reidhead M, Waterman B, et al. Adding socioeconomic data to hospital readmissions calculations may produce more useful results. *Health Aff (Millwood)* 2014;33:786-91.
 20. Ayers DC, Fehring TK, Odum SM, et al. Using joint registry data from FORCE-TJR to improve the accuracy of risk-adjustment prediction models for thirty-day readmission after total hip replacement and total knee replacement. *J Bone Joint Surg Am* 2015;97:668-71.
 21. Hammill BG, Hernandez AF, Peterson ED, et al. Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. *Am Heart J* 2009;157:995-1000.
 22. Tabak YP, Johannes RS, Silber JH. Using automated clinical data for risk adjustment: development and validation of six disease-specific mortality predictive models for pay-for-performance. *Med Care* 2007;45:789-805.
 23. Basques BA, McLynn RP, Fice MP, et al. Results of Database Studies in Spine Surgery Can Be Influenced by Missing Data. *Clin Orthop Relat Res* 2017;475:2893-904.
 24. Fu MC, Buerba RA, Grauer JN. Preoperative Nutritional Status as an Adjunct Predictor of Major Postoperative Complications Following Anterior Cervical Discectomy and Fusion. *Clin Spine Surg* 2016;29:167-72.
 25. Kong CG, Kim YY, Ahn CY, et al. Diagnostic usefulness of white blood cell and absolute neutrophil count for postoperative infection after anterior cervical discectomy and fusion using allograft and demineralized bone matrix. *Asian Spine J* 2013;7:173-7.
 26. Liao JC, Chen WJ, Chen LH, et al. Complications associated with instrumented lumbar surgery in patients with liver cirrhosis: a matched cohort analysis. *Spine J* 2013;13:908-13.
 27. Gliklich RE, Dreyer NA, Leavy MB. editors. *Registries for Evaluating Patient Outcomes: A User's Guide*. AHRQ Methods for Effective Health Care. Rockville (MD): Agency for Healthcare Research and Quality (US), 2014.
 28. Somani S, Di Capua J, Kim JS, et al. Comparing National Inpatient Sample and National Surgical Quality Improvement Program: An Independent Risk Factor Analysis for Risk Stratification in Anterior Cervical Discectomy and Fusion. *Spine (Phila Pa 1976)* 2017;42:565-72.

Cite this article as: Bernstein DN, Keswani A, Chi D, Dowdell JE, Overley SC, Chaudhary SB, Mesfin A. Development and validation of risk-adjustment models for elective, single-level posterior lumbar spinal fusions. *J Spine Surg* 2019;5(1):46-57. doi: 10.21037/jss.2018.12.11