Accepted Manuscript online: 2022-06-25 Article published online: 2022-08-25

Machine Learning Models for Predicting Postoperative Outcomes following Skull Base Meningioma Surgery

Adrian E. Jimenez[1] Jose L. Porras[1] Tej D. Azad[1] Pavan P. Shah[1] Christopher M. Jackson[1] Gary Gallia[1] Chetan Bettegowda[1] Jon Weingart[1] Debraj Mukherjee[1]

[1] Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States

J Neurol Surg B Skull Base 2022;83:635–645.

Address for correspondence Debraj Mukherjee, MD, MPH, Department of Neurosurgery, Johns Hopkins University School of Medicine, 1800 Orleans St, Baltimore, MD 21287, United States (e-mail: dmukher1@jhmi.edu).

Abstract

Objective While predictive analytic techniques have been used to analyze meningioma postoperative outcomes, to our knowledge, there have been no studies that have investigated the utility of machine learning (ML) models in prognosticating outcomes among skull base meningioma patients. The present study aimed to develop models for predicting postoperative outcomes among skull base meningioma patients, specifically prolonged hospital length of stay (LOS), nonroutine discharge disposition, and high hospital charges. We also validated the predictive performance of our models on out-of-sample testing data.

Methods Patients who underwent skull base meningioma surgery between 2016 and 2019 at an academic institution were included in our study. Prolonged hospital LOS and high hospital charges were defined as >4 days and >$47,887, respectively. Elastic net logistic regression algorithms were trained to predict postoperative outcomes using 70% of available data, and their predictive performance was evaluated on the remaining 30%.

Results A total of 265 patients were included in our final analysis. Our cohort was majority female (77.7%) and Caucasian (63.4%). Elastic net logistic regression algorithms predicting prolonged LOS, nonroutine discharge, and high hospital charges achieved areas under the receiver operating characteristic curve of 0.798, 0.752, and 0.592, respectively. Further, all models were adequately calibrated as determined by the Spiegelhalter *Z*-test ($p > 0.05$).

Conclusion Our study developed models predicting prolonged hospital LOS, nonroutine discharge disposition, and high hospital charges among skull base meningioma patients. Our models highlight the utility of ML as a tool to aid skull base surgeons in providing high-value health care and optimizing clinical workflows.

Keywords

- meningioma
- neurovascular structures
- central nervous system
- overall survival
- outcomes

Introduction

Meningiomas are the most common type of primary intracranial tumor affecting the central nervous system (CNS).[1,2] They can be classified into two subtypes based on anatomical location: skull base and non-skull base.[3] While maximal resection of the tumor and its dural attachment is the standard surgical treatment for all meningiomas types, the utility of extent-of-resection as a meaningful prognostic

received December 31, 2021
accepted after revision June 20, 2022
accepted manuscript online June 25, 2022
article published online August 25, 2022

DOI https://doi.org/10.1055/a-1885-1447.
ISSN 2193-6331.

variable for skull base meningioma patients specifically has recently been brought into question.[4] Further, the location of many skull base meningiomas adjacent to critical neurovascular structures often makes aggressive resection difficult or unfeasible, and puts patients at risk of serious postoperative complications.[5] While overall survival has not been shown to be significantly different between skull base and non-skull base meningiomas, patients with skull base tumors are approximately twice as likely to undergo retreatment of their meningiomas (either by surgery or radiotherapy) and have significantly shorter retreatment-free survival.[3] Therefore, being able to better predict the postoperative course of skull base meningioma patients specifically may aid in reducing patient morbidity and optimizing the delivery of high-value health care after surgery.

While predictive analytic techniques have been used to study meningioma postoperative outcomes generally, to our knowledge, there have been no studies that have investigated the utility of machine learning (ML) models in prognosticating outcomes among skull base meningioma patients specifically. The goal of the present study was to validate a workflow aimed at (1) developing predictive models for postoperative outcomes among skull base meningioma patients, specifically including prolonged hospital length of stay (LOS), nonroutine discharge disposition, and high hospital charges; and (2) validating the predictive performance of these models of out-of-sample testing data. We hope that this proof-of-concept study demonstrates the validity of using ML to develop effective prognostic tools for skull base meningioma patients.

Methods

Patient Selection and Recorded Variables

The present study was conducted using data from 265 patients who received surgical resection of their skull base meningiomas at a single academic institution between January 1st, 2016 and December 31st, 2019. Using Al-Mefty's anatomical classification system, the following meningioma subtypes were defined as "skull base": tuberculum sellae, planum sphenoidale, olfactory groove, sphenoid wing/spheno-orbital, clinoidal, cavernous sinus, clival and petroclival, tentorial, cerebellopontine angle, foramen magnum meningiomas, meningiomas of the middle fossa floor, and temporal bone meningiomas.[6] Our Institutional Review Board (IRB), acting as a Health Insurance Portability and Accountability Act (HIPAA) Privacy Board, reviewed and approved the waiver of informed consent for this retrospective study (IRB00181593). Manual chart review of electronic medical records was used to obtain demographic and clinical information. Tumor size and location were determined using post-contrast magnetic resonance images, with tumor volume measured using tumor dimensions in axial (x), coronal (y), and sagittal (z) planes via the following formula: $\frac{x \cdot y \cdot z}{2}$. An American Society of Anesthesiology physical status classification system (ASA) score was documented for each patient, and patient frailty was quantified using the 5-factor modified frailty index (mFI-5).[7,8] A symptomatic presentation was defined as a meningioma diagnosis on the basis of a workup prompted by any of the following symptoms: seizures, headaches, nausea/vomiting, diplopia, decreased hearing, vertigo, dysarthria, dysphagia, confusion, bladder incontinence, motor deficit, sensory deficit, language deficit, visual deficit, cognitive deficit, or gait deficit. Surgeon years of experience was defined as the number of years since a surgeon completed their residency training to the date of surgery, in line with prior research.[9,10]

Regarding postoperative outcomes, prolonged hospital LOS and high hospital charges were both analyzed as dichotomous variables using a cutoff of the upper quartile of each outcome (>4 days for LOS and >$47,887 for hospital charges), as described previously.[11–16] For the present study, routine discharge disposition was defined as discharge to home (either with self-care or health care service assistance) and nonroutine discharge was defined as discharge to a rehabilitation facility, a skilled nursing facility, or a hospice facility.[17]

Statistical Analysis

Data were collected using Microsoft Excel (version 2016, Microsoft Corp.) and statistical analyses were conducted using R statistical software (version 4.0.2, r-project.org). Bivariate analyses were conducted using Fisher's exact test and the Mann-Whitney U test for continuous and categorical variables, respectively. ML algorithms were trained using the Caret package.[18] For this study, elastic net logistic regression ML algorithms were used. Briefly, the elastic net is a statistical technique used to prevent model overfitting by applying a penalty function to regression β-coefficients.[19] Addition of this penalty function to either linear or logistic regression model coefficients allows for better predictive performance compared with ordinary least-squares regression coefficients by effectively removing model covariates that do not contribute to optimizing predictive performance, thereby creating more parsimonious models that perform better when evaluated on out-of-sample testing data.[19]

Patient data was separated into training and independent holdout testing subsets based on an 70/30 ratio, respectively. Fivefold cross validation repeated 10 times was used to tune model hyperparameters on the 70% training dataset, and hyperparameter optimization was conducted using a random search.[20] Following training (i.e., hyperparameter optimization), the predictive abilities of the finalized models were evaluated on 30% holdout testing dataset. Models were compared based on their discrimination and calibration, quantified by the area under the receiver operating characteristic curve (AUROC) and by the Brier score, respectively.[21,22] An AUROC (also known as the c-statistic) of 0.70 is generally taken to indicate that a predictive model demonstrates clinically-useful discrimination.[21] For AUROC and Brier score metrics, 95% confidence intervals were obtained using 2,000 bootstrapped replicates while a 95% confidence interval for accuracy was calculated as described by Clopper and Pearson.[23] Spiegelhalter's Z-test was also used to assess for adequate calibration of the final models for prolonged LOS, nonroutine discharge, and high hospital charges, with p <0.05 indicating inadequate calibration.[22] Variable importance plots were also created for each model to depict the relative importance of each variable toward calculating the predicted outcome.

Results

Patient Demographics and Outcomes

►**Table 1** demonstrates the demographic and clinical characteristics of our patient cohort. Our cohort was comprised of patients with the following types of skull base

Table 1 Patient demographic and clinical characteristics ($n = 265$)

Characteristic	*n* (%)
Mean age ± SD	58.89 ± 12.91
Sex	
Female	206 (77.7)
Male	59 (22.3)
Race	
White or Caucasian	168 (63.4)
Black or African-American, Asian, or Other	97 (36.6)
Insurance	
Private	182 (68.7)
Medicare or Medicaid	83 (31.3)
Marital status	
Married	185 (69.8)
Not married	80 (30.2)
Admission type	
Elective surgery	244 (92.1)
Non-elective	21 (7.9)
WHO Grade	
I	249 (94.0)
II/III	16 (6.0)
Mean tumor volume (cm^3) ± SD	17.17 ± 20.99
Tumor location	
Anterior fossa	47 (17.7)
Middle fossa	111 (41.9)
Posterior fossa	107 (40.4)
Mean ASA score ± SD	2.64 ± 0.53
Mean mFI-5 score ± SD	0.79 ± 0.84
Hypertension requiring medication	138 (52.1)
Diabetes	46 (17.4)
Chronic obstructive pulmonary disease	13 (4.9)
Congestive heart failure	9 (3.4)
Functional status	4 (1.5)
Symptomatic presentation	
Yes	241 (90.9)
No	24 (9.1)
Mean surgeon years of experience ± SD	16.64 ± 10.71
Surgical approach	
Endoscopic endonasal resection	9 (3.4)
Craniotomy	256 (96.6)

(*Continued*)

Table 1 (*Continued*)

Characteristic	*n* (%)
Mean hospital LOS ± SD	4.86 ± 7.18
Prolonged hospital LOS (>4 d)	
Yes	66 (24.9)
No	199 (75.1)
Discharge disposition	
Non-routine	30 (11.3)
Routine	235 (88.7)
Mean hospital charges in U.S. dollars ($) ± SD	$44,740.88 ± 29,547.03
High hospital charges (>$47,887.39)	
Yes	66 (24.9)
No	199 (75.1)

Abbreviations: LOS, length of stay; SD, standard deviation.

meningiomas: 72 sphenoid wing meningiomas (no spheno-orbital meningiomas), 54 cerebellopontine angle meningiomas, 43 tentorial meningiomas, 27 tuberculum sella meningiomas, 20 cavernous sinus meningiomas, 15 olfactory groove meningiomas, nine meningiomas of the middle fossa floor, eight meningiomas of the temporal bone, six foramen magnum meningiomas, five planum sphenoidal meningiomas, three clinoidal meningiomas, and three clival/petroclival meningiomas. Overall, our cohort had a mean age (± SD) of 58.89 ± 12.91 years, was majority female (77.7%), and was mostly Caucasian (63.4%). Most patients had private insurance (68.7%), were married (69.8%), and underwent an elective resection of their skull base meningioma (92.1%). Most patients had WHO grade I tumors (94.0%), and the mean tumor size (± SD) of our cohort was 17.17 ± 20.99 cm^3. A total of 47 (17.7%), 111 (41.9%), and 107 (40.4%) patients had tumors located in the anterior, middle, and posterior fossa of the skull base, respectively. The mean ASA and mFI-5 scores (± SD) for our patients were 2.64 ± 0.53 and 0.79 ± 0.84, respectively. The majority of our patient cohort had a symptomatic presentation (90.9%). The mean surgeon years of experience (± SD) among our cohort was 16.64 ± 10.71 years, with a small minority of surgeons utilizing an endoscopic endonasal approach for tumor resection (3.4%). The mean surgery duration (± SD) in our cohort was 4.84 ± 1.92 hours, while the mean hospital LOS (± SD) was 4.86 ± 7.18 days. Most patients had a routine discharge disposition postoperatively (88.7%), and the mean hospital charges (± SD) incurred among our patients were $44,740.88 ± $29,547.03.

Bivariate Analysis

►**Table 2** displays the results of our bivariate analysis assessing for significant relationships between patient demographic/clinical variables and our three postoperative outcomes of interest: prolonged LOS, nonroutine discharge, and high hospital charges. Regarding LOS, older patient age ($p = 0.017$), Medicare or Medicaid insurance status ($p = 0.0021$), non-elective admission ($p < 0.0001$), greater tumor volume ($p = 0.0017$), higher ASA score ($p < 0.0001$), greater mFI-5 score ($p = 0.0032$), less surgeon years of experience ($p = 0.0045$), and longer surgery duration ($p = 0.017$) were all significantly associated with prolonged hospital LOS. Regarding discharge disposition, older patient age ($p < 0.0001$), Medicare or Medicaid insurance status ($p = 0.035$), greater tumor volume ($p < 0.0001$), higher ASA score ($p < 0.001$), and higher mFI-5 score ($p = 0.020$) were all significantly associated with nonroutine discharge. Finally, non-Caucasian race ($p = 0.012$), Medicare or Medicaid insurance status ($p = 0.032$), non-elective admission ($p = 0.0010$), greater tumor volume ($p < 0.0001$), higher ASA score ($p = 0.031$), less surgeon years of experience ($p < 0.001$), and longer surgery duration ($p < 0.0001$) were all significantly associated with high hospital charges.

Predictive Modeling Results

►**Table 3** displays the regularized coefficients for our fully-trained elastic net logistic regression models in addition to their respective odds ratios. Further, our elastic net models predicting prolonged LOS, nonroutine discharge, and high hospital charges had α values of 0.38, 0.017, and 0.95, respectively. While all patient demographic and clinical variables were utilized to predict nonroutine discharge disposition, elastic net regularization only selected the following variables for predicting prolonged hospital LOS: insurance status, admission type, tumor volume, ASA score, and surgeon years of experience. Further, elastic net regularization removed the following variables from logistic regression analysis when seeking to optimize the predictive accuracy of high hospital charges: patient age, patient sex, patient race, marital status, WHO grade, tumor location, ASA score, mFI-5 score, symptomatic presentation, and surgical approach.

Importantly, our analysis demonstrated that Medicare or Medicaid insurance status (odds ratio [OR] = 1.11), non-elective admission (OR = 1.30), greater tumor volume (OR

Table 2 Bivariate analysis of patient demographic/clinical characteristics and postoperative outcomes ($n = 265$)

Characteristic	Hospital LOS			Discharge disposition			Hospital charges		
	>4 d (*n* = 66)	≤4 d (*n* = 199)	*p*-Value	Non-routine (*n* = 30)	Routine (*n* = 235)	*p*-Value	>$47,887.39 (*n* = 66)	≤$47,887.39 (*n* = 199)	*p*-Value
Mean age ± SD	62.36 ± 13.17	57.73 ± 12.65	***0.017***[a]	67.87 ± 11.58	57.74 ± 12.64	<***0.0001***[a]	61.32 ± 14.49	58.08 ± 12.28	0.074
Sex									
Male	13 (19.7)	46 (23.1)	0.61	4 (13.3)	55 (23.4)	0.25	15 (22.7)	44 (22.1)	1.00
Female	53 (80.3)	153 (76.9)		26 (86.7)	180 (76.6)		51 (77.3)	155 (77.9)	
Race									
Black or African-American, Asian, or Other	31 (47.0)	66 (33.2)	0.055	7 (23.3)	90 (38.3)	0.16	33 (50.0)	64 (32.2)	***0.012***[a]
White or Caucasian	35 (53.0)	133 (66.8)		23 (76.7)	145 (61.7)		33 (50.0)	135 (67.8)	
Insurance									
Medicare or Medicaid	31 (47.0)	52 (26.1)	***0.0021***[a]	15 (50.0)	68 (28.9)	***0.035***[a]	28 (42.4)	55 (27.6)	***0.032***[a]
Private	35 (53.0)	147 (73.9)		15 (50.0)	167 (71.1)		38 (57.6)	144 (72.4)	
Marital status									
Married	42 (63.6)	143 (71.9)	0.22	16 (53.3)	169 (71.9)	0.055	42 (63.6)	143 (71.9)	0.22
Not married	24 (36.4)	56 (28.1)		14 (46.7)	66 (28.1)		24 (36.4)	56 (28.1)	
Admission type									
Non-elective	16 (24.2)	5 (2.5)	<***0.0001***[a]	2 (6.7)	19 (8.1)	1.00	12 (18.2)	9 (4.5)	***0.0010***[a]
Elective	50 (75.8)	194 (97.5)		28 (93.3)	216 (91.1)		54 (81.8)	190 (95.5)	
WHO Grade									
II/III	5 (7.6)	11 (5.5)	0.56	4 (13.3)	12 (5.1)	0.092	5 (7.6)	11 (5.5)	0.56
I	61 (92.4)	188 (94.5)		26 (86.7)	223 (94.9)		61 (92.4)	188 (94.5)	
Mean tumor volume (cm^3) ± SD	24.34 ± 24.30	14.79 ± 19.25	***0.0017***[a]	31.25 ± 23.94	15.37 ± 19.93	<***0.0001***[a]	26.52 ± 25.88	14.07 ± 18.13	<***0.0001***[a]
Tumor location									
Anterior fossa	14 (21.2)	33 (16.6)	Reference	4 (13.3)	43 (18.3)	Reference	15 (22.7)	32 (16.1)	Reference
Middle fossa	25 (37.9)	86 (43.2)	0.42	14 (46.7)	97 (41.3)	0.59	28 (42.4)	83 (41.7)	0.44
Posterior fossa	27 (40.9)	80 (40.2)	0.56	12 (40.0)	95 (40.4)	0.78	23 (34.8)	84 (42.2)	0.23
Mean ASA score ± SD	2.88 ± 0.41	2.56 ± 0.55	<***0.0001***[a]	2.97 ± 0.32	2.60 ± 0.54	<***0.001***[a]	2.77 ± 0.49	2.60 ± 0.54	***0.031***[a]
Mean mFI-5 score ± SD	1.09 ± 0.97	0.69 ± 0.77	***0.0032***[a]	1.20 ± 1.06	0.74 ± 0.79	***0.020***[a]	0.98 ± 0.98	0.73 ± 0.78	0.099

(*Continued*)

Table 2 (*Continued*)

Characteristic	Hospital LOS			Discharge disposition			Hospital charges		
	>4 d (*n* = 66)	≤4 d (*n* = 199)	*p*-Value	Non-routine (*n* = 30)	Routine (*n* = 235)	*p*-Value	>$47,887.39 (*n* = 66)	≤$47,887.39 (*n* = 199)	*p*-Value
Symptomatic presentation									
Yes	63 (95.5)	178 (89.4)	0.21	29 (96.7)	212 (90.2)	0.49	63 (95.5)	178 (89.4)	
No	3 (4.5)	21 (10.6)		1 (3.3)	23 (9.8)		3 (4.5)	21 (10.6)	
Mean surgeon years of experience ± SD	13.26 ± 10.57	17.76 ± 10.54	***0.0045***[a]	14.47 ± 9.22	16.92 ± 10.87	0.47	12.41 ± 10.15	18.05 ± 10.54	***<0.001***[a]
Surgical approach									
Endoscopic endonasal resection	3 (4.5)	6 (3.0)	0.69	0 (0.0)	9 (3.8)	0.60	4 (6.1)	5 (2.5)	0.23
Craniotomy	63 (95.5)	193 (97.0)		30 (100.0)	226 (96.2)		62 (93.9)	194 (97.5)	

Note: Bold is used to highlight statistical significance.

= 1.04 per 1 cm^3 increase), and higher ASA score (OR = 1.22 per 1 point increase) were all associated with higher odds of prolonged hospital LOS. Further, greater surgeon years of experience was associated with a lower odds of prolonged hospital LOS (OR = 0.95 per additional year of experience). Regarding discharge disposition, older patient age (OR = 1.55 per 1 year increase), Medicare or Medicaid insurance status (OR = 1.11), WHO grade II/III vs I (OR = 1.13), greater tumor volume (OR = 1.51 per 1 cm^3 increase), middle relative to anterior fossa tumor location (OR = 1.11), higher ASA score (OR = 1.36 per 1 point increase), higher mFI-5 score (OR = 1.07 per 1 point increase), and symptomatic presentation (OR = 1.05) were all associated with increased odds of non-routine discharge. Male sex (OR = 0.74), non-Caucasian race (OR = 0.83), married marital status (OR = 0.87), non-elective admission status (OR = 0.79), posterior relative to anterior fossa tumor location (OR = 0.93), greater surgeon years of experience (OR = 0.86 per additional year of experience), and endoscopic endonasal approach relative to craniotomy (OR = 0.85) were all associated with decreased odds of nonroutine discharge. Finally, Medicare or Medicaid insurance status (OR = 1.03), non-elective admission (OR = 1.13), and greater tumor volume (OR = 1.46 per 1 cm^3 increase) were all associated with increased odds of high hospital charges. Greater surgeon years of experience (OR = 0.85 per 1 year increase) was the sole variable associated with decreased odds of incurring high hospital charges.

►**Table 4** displays the predictive performance metrics for our elastic net logistic regression models on our holdout validation datasets. Models predicting prolonged LOS, nonroutine discharge, and high hospital charges achieved AUROCs of 0.798, 0.752, and 0.592, respectively. ROC plots for all three models are displayed in ►**Fig. 1**. The elastic net logistic regression model predicting prolonged LOS achieved an accuracy of 82.1% on its holdout dataset, while the model predicting nonroutine discharge achieved an accuracy of 89.9% and the model predicting high hospital charges achieved an accuracy of 73.4%. Further, the Brier scores of models predicting prolonged LOS, nonroutine discharge, and high hospital charges were of 0.15, 0.084, and 0.19, respectively. All three models demonstrated adequate calibration via the Spiegelhalter *Z*-test ($p > 0.05$). Variable importance plots for all three elastic net models are depicted in ►**Fig. 2A–C**.

Discussion

Prior Research

Previous research within the neurosurgical literature has utilized ML to predict postoperative outcomes such as LOS, discharge disposition, and hospital charges in brain tumor patients. A 2017 study by Muhlestein et al analyzed the impact of medical comorbidities on discharge disposition and LOS following craniotomy for brain tumor patients listed within the National Inpatient Sample.[13] The authors created ML ensemble models predicting discharge disposition and LOS >7 days with AUROC of 0.796 and 0.824, respectively. The authors also found that preoperative paralysis,

Table 3 Elastic net logistic regression coefficients and odds ratios for models predicting LOS, discharge, hospital charges, and readmission

	Hospital LOS		Nonroutine disposition		Hospital charges	
Characteristic	β-coefficient	Odds ratio	β-coefficient	Odds ratio	β-coefficient	Odds ratio
Age	–	–	0.44	1.55	–	–
Sex						
Male	–	–	−0.30	0.74	–	–
Female	–	–			–	–
Race						
Black or African-American, Asian, or Other	–	–	−0.19	0.83	–	–
White or Caucasian	–	–			–	–
Insurance						
Medicare or Medicaid	0.10	1.11	0.11	1.11	0.026	1.03
Private						
Marital status						
Married	–	–	−0.15	0.87	–	–
Not married	–	–			–	–
Admission type						
Non-elective	0.26	1.30	−0.24	0.79	0.12	1.13
Elective						
WHO Grade						
II/III	–	–	0.12	1.13	–	–
I	–	–			–	–
Tumor volume (cm^3)	0.039	1.04	0.41	1.51	0.38	1.46
Tumor location						
Anterior fossa	Reference	Reference	Reference	Reference	Reference	Reference
Middle fossa	–	–	0.11	1.11	–	–
Posterior fossa	–	–	−0.076	0.93	–	–
ASA score	0.20	1.22	0.30	1.36	–	–
mFI-5 score	–	–	0.067	1.07	–	–
Symptomatic presentation						
Yes	–	–	0.052	1.05	–	–
No	–	–			–	–
Surgeon years of experience	−0.055	0.95	−0.15	0.86	−0.17	0.85
Surgical approach						
Endoscopic endonasal resection	–	–	−0.16	0.85	–	–
Craniotomy	–	–			–	–

fluid/electrolyte abnormalities, and other non-paralysis neurological defects most strongly influenced the ensemble model predicting prolonged LOS.[13] In a separate study, Muhlestein et al also developed ML models that directly predicted total charges for transsphenoidal surgery for pituitary tumors, and the investigators identified extended LOS, nonelective admission type, non-Southern hospital region, minority race, postoperative complication, and private investor hospital ownership as drivers of total charges and therefore as potential targets for cost-lowering interventions.[24] Within the orthopaedics literature, a 2018 study by Navarro et al used a naïve Bayesian ML algorithm to predict LOS and inpatient costs following total knee arthroplasty. The authors noted that their ML algorithm demonstrated high validity for predicting LOS and hospital charges, with AUROCs of 0.782 and 0.738, respectively.[25] While our

Table 4 Elastic net logistic regression models predictive performance metrics and 95% confidence intervals on holdout validation sets

Metric	Hospital LOS ($n = 78$)	Nonroutine discharge ($n = 79$)	Hospital charges ($n = 79$)
AUROC	0.798 (0.662–0.900)	0.752 (0.581–0.906)	0.592 (0.445–0.731)
Accuracy (%)	82.1% (71.7%–89.8%)	89.9% (81.0%–95.5%)	73.4% (62.3%–82.7%)
Brier score	0.15 (0.11–0.19)	0.084 (0.037–0.13)	0.19 (0.14–0.24)
Spiegelhalter *Z*-test *p*-value	0.16	0.83	0.64

Abbreviation: LOS, length of stay.

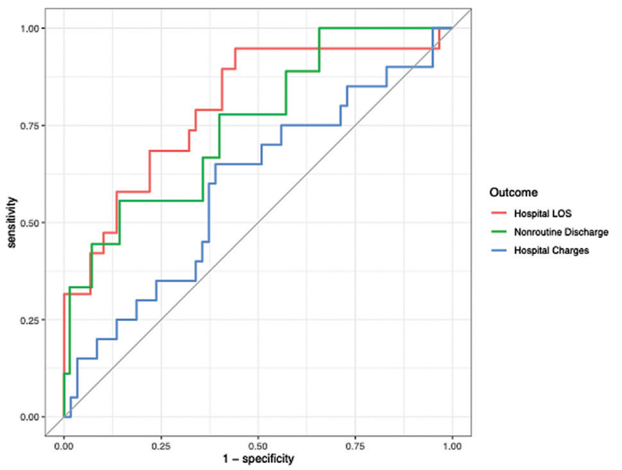

Fig. 1 ROC curves for LOS, discharge, and hospital charges on holdout validation sets. LOS, length of stay; ROC, receiver operating characteristic curve.

prior work has utilized logistic regression models and inferential statistics to identify predictors of high-value care outcomes among skull base meningioma patients as well as to obtain predictive performance metrics for these models, the present study represents the first effort (to our knowledge) of applying ML methods to model these outcomes.[26]

Present Study

The present work sought to apply an ML-workflow to prognosticating postoperative outcomes among skull base meningioma patients. Additionally, we also evaluated the discrimination and calibration of our predictive algorithms to assess whether preoperative patient demographic and clinical characteristics could effectively predict high-value care metrics. Overall, our two elastic net logistic regression models predicting prolonged hospital LOS and nonroutine discharge disposition demonstrated adequate discrimination (AUROC >0.70) and calibration (Spiegelhalter's *Z*-test *p*-value >0.05). Importantly, while our model predicting high hospital charges demonstrated adequate calibration, its AUROC of 0.592 suggests inadequate discrimination that would likely not be clinically useful. One potential reason for the our limited ability to predict high hospital charges preoperatively is that the charges incurred by a patient during their hospital stay are mainly influenced by intra- and postoperative variables such as surgery duration,

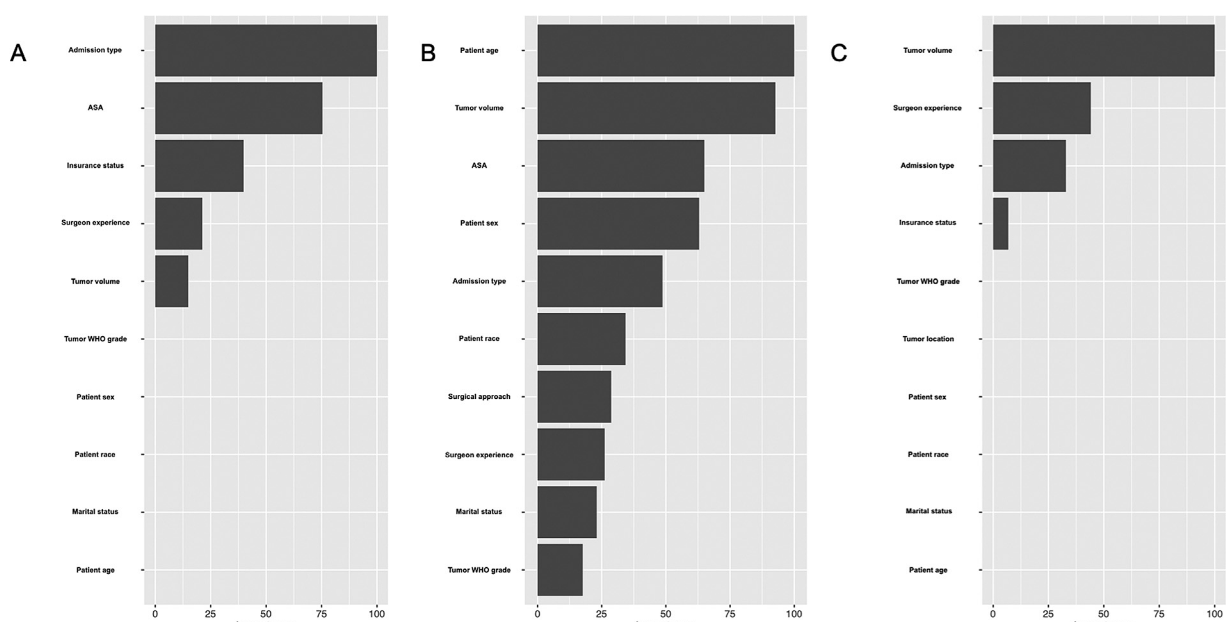

Fig. 2 Variable importance plots for elastic net logistic regression models predicting **(A)** LOS, **(B)** discharge, and **(C)** hospital charges. LOS, length of stay.

postoperative complication, and total hospital LOS, as detailed in prior studies.[24,27,28] Our results highlight that while ML algorithms may be useful for predicting postoperative risk of prolonged LOS and nonroutine discharge disposition, the preoperative variables utilized in the present study are likely not sufficient for effectively predicting a patient's risk of incurring high hospital charges.

Further, our use of the elastic net logistic regression algorithm allowed us to calculate odds ratios for each model covariate to determine whether a given variable was associated with a higher or lower odds of a given postoperative outcome. Our results demonstrating that Medicare or Medicaid insurance, non-elective admission, greater tumor volume, higher ASA score, and less surgeon years of experience were all associated with increased odds of prolonged hospital LOS corroborates prior findings within the neurosurgical and spine surgery literature.[11,29–31] Further, the fact that older patient age, Medicare or Medicaid insurance, greater tumor volume, higher ASA score, and higher mFI-5 score are all associated with higher odds of nonroutine discharge also validates prior research findings.[32–35] Finally, the fact that Medicare or Medicaid insurance, non-elective admission, greater tumor volume, and less surgeon years of experience were all associated with higher odds in high hospital charges according to our predictive model is also in line with previous research.[16,30,36–40]

Interestingly, several associations that were significant in bivariate analysis were excluded from the final predictive models during the training process (such as the significant association between older patient age and higher odds of prolonged LOS), and some associations that were not significant in bivariate analysis were included as inputs in the final elastic net models (such as the association between patient sex and discharge disposition). This also highlights the importance of differentiating between inferential statistics and predictive analytics. While inferential methods such as generalized linear models make use of probabilistic assumptions and hypothesis testing to provide a mathematical guarantee regarding the underlying structure and behavior of associations observed in datasets, predictive analytic methods such as ML algorithms focus mainly on achieving superior predictive performance (AUROC, calibration) on out-of-sample datasets with a lesser emphasis on representing the data-generating mechanism between model input and output.[41] Linear and logistic regression methods represent examples of inferential approaches where the probabilistic, stochastic structure of the models allows for calculation of meaningful confidence intervals and p-values that can be interpreted to yield insight regarding specific relationships between model inputs and outputs. Deep neural networks, on the other hand, are examples of algorithms that can achieve excellent performance metrics when predicting complex, non-linear relationships, but also provide little insight regarding the underlying data-generating relationship and thus how such predictions are being made.[42,43]

The present study developed three predictive models using elastic net regularization, an ML algorithm that accomplishes data-driven variable selection and allowed us to train two models predicting prolonged LOS and nonroutine discharge disposition which demonstrated good calibration and discrimination on out-of-sample datasets. Elastic net logistic regression was also useful because we are able to calculate β-coefficients and odds ratios to gain some degree of understanding regarding how our model inputs were producing our model outputs. However, given that the elastic net regularization method does not have an underlying probabilistic structure like ordinary-least-square linear or logistic regression that would allow for the calculation of confidence intervals and p-values, it is important to keep in mind that we are limited regarding our inferences about the statistical relationships among model inputs and output.[44,45] Overall, our study demonstrates that ML methodology can be applied to prognosticate postoperative outcomes among skull base meningioma patients with reasonable predictive performance. However, investigators must be mindful about whether their priority is achieving superior predictive performance or attaining a better understanding of the underlying statistical relationships in their data. Further, given the mixed results of ML predictive performance relative to traditional statistical techniques within the medical literature, one cannot assume a priori that ML always leads to superior predictive performance compared with linear or logistic regression and therefore should empirically assess the performance of their algorithms to see if they attain acceptable levels of discrimination and calibration for their outcomes of interest.[46–49]

Limitations

The present study is retrospective and is limited in its analysis of patient data from a single academic, medical institution during a restricted time period (2016–2019). The retrospective design of our study prevents us from commenting on any causal relationships that may exist between the variables that we analyzed. External validation of our findings in an independent cohort of skull base meningioma patients would be ideal to ensure the generalizability of our findings; this validation provides an avenue for future research. As all patients in this study were surgically managed, the model is not valid for patients who are treated only with non-surgical approaches such as radiotherapy. Additionally, the present study only estimated total hospital charges as opposed to costs. Hospital charges are initial hospital list prices for services while costs represent actual expenses incurred during a patient's hospitalization, and while charges and costs are not synonymous, we agree with previous investigators that charges may represent a useful proxy for costs.[24] Applying ML methods to estimate charges directly may serve as an avenue for future research efforts. Another important limitation is the small number of EEA cases in our patient cohort ($n = 9$). Additional research efforts incorporating greater number of EEA surgeries may be needed to definitively determine whether surgical approach is a useful prognostic variable to consider when using ML to predict high-value postoperative outcomes among skull base meningioma patients. Acknowledging these limitations, the present study has developed and internally validated predictive models that may be useful in optimizing postoperative outcomes and increasing the provision of high-value health care.

Conclusion

Our study developed three ML models predicting prolonged hospital LOS, nonroutine discharge disposition, and high hospital charges among skull base meningioma patients. Using preoperative demographic and clinical variables, our models predicting hospital LOS and nonroutine discharge disposition demonstrated adequate discrimination and calibration, and highlight the utility of ML as a tool to aid neurosurgeons in providing high-value health care and optimizing clinical workflows.

Reporting Guidelines

The authors found no applicable reporting guidelines that would apply to this article. By following the EQUATOR reporting guidelines decision tree, (http://www.equatornetwork.org/wp-content/uploads/2013/11/20160226-RG-decision-tree-for-Wizard-CC-BY-26- February-2016.pdf), we found that none of the most popular checklists are appropriate for our study design.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article. The authors acknowledge assistance for clinical data coordination and retrieval from the Core for Clinical Research Data Acquisition, supported in part by the Johns Hopkins Institute for Clinical and Translational Research (UL1TR001079).

Conflict of Interest

None declared.

References

1 Ostrom QT, Gittleman H, Fulop J, et al. CBTRUS statistical Report: primary brain and central nervous system tumors diagnosed in the United States in 2008–2012. Neuro-oncol 2015;17(suppl 4):iv1–iv62
2 Wang N, Osswald M. Meningiomas: overview and new directions in therapy. Semin Neurol 2018;38(01):112–120
3 Meling TR, Da Broi M, Scheie D, Helseth E. Meningiomas: skull base versus non-skull base. Neurosurg Rev 2019;42(01):163–173
4 Voß KM, Spille DC, Sauerland C, et al. The Simpson grading in meningioma surgery: does the tumor location influence the prognostic value? J Neurooncol 2017;133(03):641–651
5 Chen CM, Huang APH, Kuo LT, Tu YK. Contemporary surgical outcome for skull base meningiomas. Neurosurg Rev 2011;34(03):281–296, discussion 296
6 DeMonte F, McDermott M, Al-Mefty O. Al-Mefty's Meningiomas. 2nd ed. Thieme; 2011
7 Saklad M. Grading of patients for surgical procedures. Anesthesiology 1941;2(03):281–284
8 Subramaniam S, Aalberg JJ, Soriano RP, Divino CM. New 5-factor modified frailty index using American College of Surgeons NSQIP Data. J Am Coll Surg 2018;226(02):173–181.e8
9 Cahill PJ, Pahys JM, Asghar J, et al. The effect of surgeon experience on outcomes of surgery for adolescent idiopathic scoliosis. J Bone Joint Surg Am 2014;96(16):1333–1339
10 Lau D, Deviren V, Ames CP. The impact of surgeon experience on perioperative complications and operative measures following thoracolumbar 3-column osteotomy for adult spinal deformity: overcoming the learning curve. J Neurosurg Spine 2019;32(02):207–220
11 Dasenbrock HH, Liu KX, Devine CA, et al. Length of hospital stay after craniotomy for tumor: a National Surgical Quality Improvement Program analysis. Neurosurg Focus 2015;39(06):E12
12 Lakomkin N, Hadjipanayis CG. Resident participation is not associated with postoperative adverse events, reoperation, or prolonged length of stay following craniotomy for brain tumor resection. J Neurooncol 2017;135(03):613–619
13 Muhlestein WE, Akagi DS, Chotai S, Chambless LB. The impact of presurgical comorbidities on discharge disposition and length of hospitalization following craniotomy for brain tumor. Surg Neurol Int 2017;8:220
14 Kalakoti P, Missios S, Menger R, Kukreja S, Konar S, Nanda A. Association of risk factors with unfavorable outcomes after resection of adult benign intradural spine tumors and the effect of hospital volume on outcomes: an analysis of 18, 297 patients across 774 US hospitals using the National Inpatient Sample (2002-2011). Neurosurg Focus 2015;39(02):E4
15 Muhlestein WE, Akagi DS, Chotai S, Chambless LB. The impact of race on discharge disposition and length of hospitalization after craniotomy for brain tumor. World Neurosurg 2017;104:24–38
16 Khalafallah AM, Jimenez AE, Patel P, Huq S, Azmeh O, Mukherjee D. A novel online calculator predicting short-term postoperative outcomes in patients with metastatic brain tumors. J Neurooncol 2020;149(03):429–436
17 Brandel MG, Rennert RC, Wali AR, et al. Impact of preoperative endovascular embolization on immediate meningioma resection outcomes. Neurosurg Focus 2018;44(04):E6
18 Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28(05):1–26
19 Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol 2005;67(05):768
20 Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res 2012;13:281–305
21 Swets JA. Measuring the accuracy of diagnostic systems. Science 1988;240(4857):1285–1293
22 Rufibach K. Use of Brier score to assess binary predictions. J Clin Epidemiol 2010;63(08):938–939, author reply 939
23 Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934;26(04):404–413
24 Muhlestein WE, Akagi DS, McManus AR, Chambless LB. Machine learning ensemble models predict total charges and drivers of cost for transsphenoidal surgery for pituitary tumor. J Neurosurg 2018;131(02):507–516
25 Navarro SM, Wang EY, Haeberle HS, et al. Machine learning and primary total knee arthroplasty: patient forecasting for a patient-specific payment model. J Arthroplasty 2018;33(12):3617–3623
26 Jimenez AE, Khalafallah AM, Lam S, et al. Predicting high-value care outcomes after surgery for skull base meningiomas. World Neurosurg 2021;149:e427–e436
27 Little AS, Chapple K. Predictors of resource utilization in transsphenoidal surgery for Cushing disease. J Neurosurg 2013;119(02):504–511
28 Zacharia BE, Deibert C, Gupta G, et al. Incidence, cost, and mortality associated with hospital-acquired conditions after resection of cranial neoplasms. Neurosurgery 2014;74(06):638–647
29 McKee SP, Yang A, Gray M, et al. Intracranial meningioma surgery: value-based care determinants in New York State, 1995-2015. World Neurosurg 2018;118:e731–e744

30 Chapman EK, Doctor T, Gal JS, et al. The impact of non-elective admission on cost of care and length of stay in anterior cervical discectomy and fusion: a propensity-matched analysis. Spine 2021;46(22):1535–1541
31 Ahn J, Iqbal A, Manning BT, et al. Minimally invasive lumbar decompression-the surgical learning curve. Spine J 2016;16(08): 909–916
32 Huq S, Khalafallah AM, Patel P, et al. Predictive model and online calculator for discharge disposition in brain tumor patients. World Neurosurg 2021;146:e786–e798
33 Lakomkin N, Hadjipanayis CG. Non-routine discharge disposition is associated with post-discharge complications and 30-day readmissions following craniotomy for brain tumor resection. J Neurooncol 2018;136(03):595–604
34 Sastry RA, Pertsch NJ, Tang O, Shao B, Toms SA, Weil RJ. Frailty and outcomes after craniotomy for brain tumor. J Clin Neurosci 2020; 81:95–100
35 Muhlestein WE, Akagi DS, Kallos JA, et al. Using a guided machine learning ensemble model to predict discharge disposition following meningioma resection. J Neurol Surg B Skull Base 2018;79(02): 123–130
36 Kidwai SM, Yang A, Gray ML, et al. Hospital charge variability across New York State: sociodemographic factors in pituitary surgery. J Neurol Surg B Skull Base 2019;80(06):612–619
37 Hamill CS, Villwock JA, Sykes KJ, Chamoun RB, Beahm DD. Socioeconomic factors affecting discharge status of patients with uncomplicated transsphenoidal adenohypophysectomy. J Neurol Surg B Skull Base 2018;79(05):501–507
38 Abou-Al-Shaar H, Azab MA, Karsy M, Guan J, Couldwell WT, Jensen RL. Assessment of costs in open microsurgery and stereotactic radiosurgery for intracranial meningiomas. World Neurosurg 2018;119:e357–e365
39 Alvin MD, Lubelski D, Alam R, et al. Spine surgeon treatment variability: the impact on costs. Global Spine J 2018;8(05):498–506
40 Doumouras AG, Saleh F, Gmora S, Anvari M, Hong D. The value of surgical experience: excess costs associated with the Roux-en-Y gastric bypass learning curve. Surg Endosc 2019;33(06): 1944–1951
41 Breiman L. Statistical modeling: the two cultures. Stat Sci 2001;16 (03):199–215
42 Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019;25(01): 65–69
43 Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. Nature 2020;577 (7792):706–710
44 Lockhart R, Taylor J, Tibshirani RJ, Tibshirani R. A significance test for the lasso. Ann Stat 2014;42(02):413–468
45 Lee JD, Sun DL, Sun Y, Taylor JE. Exact post-selection inference, with application to the lasso. Ann Stat 2016;44(03):907–927
46 Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12–22
47 Marcus AP, Marcus HJ, Camp SJ, Nandi D, Kitchen N, Thorne L. Improved prediction of surgical resectability in patients with glioblastoma using an artificial neural network. Sci Rep 2020;10 (01):5143
48 Bae S, Massie AB, Caffo BS, Jackson KR, Segev DL. Machine learning to predict transplant outcomes: helpful or hype? A national cohort study. Transpl Int 2020;33(11):1472–1480
49 Oosterhoff JHF, Gravesteijn BY, Karhade AV, et al; Machine Learning Consortium. Feasibility of machine learning and logistic regression algorithms to predict outcome in orthopaedic trauma surgery. J Bone Joint Surg Am 2022;104(06): 544–551

