Machine learning ensemble models predict total charges and drivers of cost for transsphenoidal surgery for pituitary tumor

Whitney E. Muhlestein, MD,¹ Dallin S. Akagi, BS,² Amy R. McManus, MD,¹ and Lola B. Chambless, MD¹

¹Department of Neurological Surgery, Vanderbilt University, Nashville, Tennessee; and ²DataRobot, Inc., Boston, Massachusetts

OBJECTIVE Efficient allocation of resources in the healthcare system enables providers to care for more and needier patients. Identifying drivers of total charges for transsphenoidal surgery (TSS) for pituitary tumors, which are poorly understood, represents an opportunity for neurosurgeons to reduce waste and provide higher-quality care for their patients. In this study the authors used a large, national database to build machine learning (ML) ensembles that directly predict total charges in this patient population. They then interrogated the ensembles to identify variables that predict high charges.

METHODS The authors created a training data set of 15,487 patients who underwent TSS between 2002 and 2011 and were registered in the National Inpatient Sample. Thirty-two ML algorithms were trained to predict total charges from 71 collected variables, and the most predictive algorithms combined to form an ensemble model. The model was internally and externally validated to demonstrate generalizability. Permutation importance and partial dependence analyses were performed to identify the strongest drivers of total charges. Given the overwhelming influence of length of stay (LOS), a second ensemble excluding LOS as a predictor was built to identify additional drivers of total charges.

RESULTS An ensemble model comprising 3 gradient boosted tree classifiers best predicted total charges (root mean square logarithmic error = 0.446; 95% CI 0.439–0.453; holdout = 0.455). LOS was by far the strongest predictor of total charges, increasing total predicted charges by approximately \$5000 per day.

In the absence of LOS, the strongest predictors of total charges were admission type, hospital region, race, any postoperative complication, and hospital ownership type.

CONCLUSIONS ML ensembles predict total charges for TSS with good fidelity. The authors 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 potential targets for cost-lowering interventions.

<https://thejns.org/doi/abs/10.3171/2018.4.JNS18306>

KEYWORDS transsphenoidal surgery; machine learning; total charges; outcomes modeling; pituitary surgery

INCREASING emphasis is being placed on improving the quality of healthcare in the United States.¹⁶ Underscoring this effort is the development of new provider reimbursement strategies that incentivize high quality over high quantity care.⁷ Efforts to decrease cost of care are an important aspect of improving healthcare quality.^{2,11} Despite this, relatively few efforts have been made to understand drivers of cost for transsphenoidal surgery (TSS) for pituitary tumor. Additionally, there currently exist no clinical models that directly predict charges for this type

of surgery, hampering the ability of providers, insurance companies, and patients to allocate resources appropriately or to develop cost-saving measures.

Interest in using machine learning (ML) models to predict hospital charges has grown considerably in light of skyrocketing costs of healthcare. For example, ML models for colorectal and gastric cancer have been used to predict hospital charges and identify targets for cost containment.^{21,30} ML models have also been shown to increase physician awareness of total charge predictors and have

ABBREVIATIONS LOS = length of stay; ML = machine learning; NIS = National (Nationwide) Inpatient Sample; RMSLE = root mean square logarithmic error; TSS = transsphenoidal surgery.

SUBMITTED February 5, 2018. **ACCEPTED** April 5, 2018.

INCLUDE WHEN CITING Published online September 21, 2018; DOI: 10.3171/2018.4.JNS18306.

been used to create clinical pathways that aim to contain cost.^{19,22} However, development of these models in neurosurgery has been limited.

Here, we use ML techniques to build an ensemble model that uses the National (Nationwide) Inpatient Sample (NIS), a national, multiinstitution registry of hospitalizations in the United States, to predict total charges for TSS for pituitary tumor. We then use permutation importance and partial dependence analyses to identify the strongest predictors of total charges in this patient population with the goal of identifying and mitigating risk factors for high total charges to reduce waste in the healthcare system. We seek to demonstrate the utility of this type of model, which may also have broader applicability in the field of neurosurgery.

Methods

Database

We used the NIS, the largest all-payer inpatient database publicly available in the United States, to identify TSS performed between 2002 and 2011. The NIS contains approximately 8 million hospital stays from approximately 1000 hospitals, sampled to approximate a 20% stratified sample of US hospitals.¹⁴ The NIS is compiled and maintained by the Agency for Healthcare Research and Quality (AHRQ). This publicly available, de-identified database was considered exempt from IRB review.

Patient Selection

All 79,742,743 admissions registered in the NIS between 2002 and 2011 were screened for inclusion in the study. Eligible admissions were first identified by ICD 9 diagnosis codes for pituitary tumor (227.3, 237.0, 239.7, 234.8, 194.3, 198.89), and then by ICD 9 procedure codes for transphenoidal approach for total (07.65) or partial (07.63) excision of the pituitary gland. We further restricted our cohort to patients 18 years or older.

Variable Selection and Primary Outcome

Seventy-one variables, including patient, tumor, and operative characteristics; postoperative complications; and hospital characteristics were collected for each hospitalization (for list of all variables, see Supplementary Table 1; for description of variables defined by ICD 9 codes, see Supplementary Table 2). Patient comorbidities were identified using the Elixhauser Comorbidity Software administered by AHRQ. The primary outcome was total hospital charges, calculated in US dollars. All charges were adjusted for inflation (calculated with the Consumer Price Index of 179.9 in 2002 and 224.939 in 2011).⁴

Data Preprocessing

Missing numerical data were dealt with by imputing the median of the column and creating a new binary column to indicate the imputation took place. Numerical data were standardized in each column by subtracting the mean and dividing by the standard deviation. For tree-based algorithms categorical data were encoded with integers. The assignment of category values to integers was done randomly.

Guided ML Ensemble Construction and Validation

We used an innovative method of ML algorithm ranking, selection, and combination to build the most predictive model for our data.²⁴ Before training, 20% of the data set was randomly selected as the holdout and not used in training. The remaining data were divided into 5 mutually exclusive folds of data, 4 of which were used together as training, with the final fold used for validation. For each algorithm, training was performed 5 times, with each fold used once for validation. Cross-validation scores were calculated by taking the root mean square logarithmic error (RMSLE) of the 5 possible validation folds (the closer to 0 the RMSLE value, the more accurate the model, with RMSLE of 0 denoting zero error). Top-performing algorithms were combined with an average blender to form an ensemble model. The ensemble model was similarly trained and a cross-validation score calculated. To demonstrate generalizability to never-before-seen data, the RMSLE was calculated for predictions made on the holdout data set. The holdout was taken as a single sample of data, and so no confidence intervals were computed.

Given the overwhelming influence of length of stay (LOS) on the first ensemble, a second ensemble, which excluded LOS as a predictor, was built. Model construction was performed using ML software from DataRobot (DataRobot v. 3.0; DataRobot, Inc.).

We generated lift charts in order to better visualize how accurately each ensemble model predicts total charges. Predicted total charges for each hospitalization were ranked in increasing order and then grouped into 10, equal-sized “bins.” The mean predicted total charge for each bin was plotted against the mean actual total charge for each bin, demonstrating how accurately the ensemble predicts total charges for each decile of hospitalizations. For example, if an ensemble-generated prediction falls within the fourth decile, the lift chart can be consulted to see if and how much the ensemble overshoots or undershoots actual total charges for that range of predictions.

Permutation Importance

The relative importance of a variable to the ensemble model was assessed using permutation importance, as described by Breiman.³ Using the training data only, for each variable the ensemble was retrained on data with the values for the variable randomly permuted. The difference in performance in RMSLE between the ensemble built on the reference data and that of the data with the permuted variable was used to rank and compare the relative importance of the features to the ensemble.

Partial Dependence

To understand the independent impact of individual variables on the ensemble predictions, we constructed partial dependence plots as described by Friedman.¹² A subset of the training data were selected. For any variable, we made predictions from the ensemble after having replaced all the values for the variable with a constant test value and computing the mean of those predictions. We tested many values to observe how the ensemble reacts to changes in the variable of interest.

FIG. 1. Histogram demonstrating distribution of total charges (in dollars). The y-axis values represent the numbers of admissions.

Other Statistical Methods

We performed traditional statistical analysis on selected patient and hospital characteristics. Mean total charges were compared between different groups using the Mann-Whitney U-test or one-way ANOVA. Significance of changes in patient characteristics between 2002 and 2011 were calculated using chi-square test for categorical variables and the Mann-Whitney U-test for continuous variables. Statistical analysis was performed using the open source statistical tools in SciPy (v. 0.17).

Results

Patient Characteristics

We reviewed 15,999 admissions for TSS for pituitary tumor for the analysis; 512 admissions had no total charges recorded, leaving 15,487 admissions for inclusion in the study. Overall, the average total charge per hospitalization was \$60,491 (with an SD of \$63,493, suggesting nonnormal distribution of charges). The lowest recorded charge was \$396 and the highest recorded charge was \$1,596,198 (Fig. 1).

Total charges increased roughly linearly from 2002 to 2011, even after adjusting for inflation, and the average total charges in 2011 were significantly higher than average total charges in 2002 (\$76,228 vs \$46,443, $p < 0.001$) (Fig. 2). Simultaneously, the average length of hospitalization dropped significantly from 5.5 days (median 4.0) in 2002 to 4.2 days (median 3.0) in 2011 ($p < 0.001$).

Patients with higher total charges tended to be older, non-white, and to have nonelective admissions, preoperative hormonal abnormalities, and visual field deficits. Nearly all postoperative complications and preoperative comorbidities, as well as longer hospitalizations, were associated with significantly higher total charges. Patients who were seen in large, urban, nonteaching hospitals under private investor control, or non-Southern hospitals had higher total charges than their counterparts seen elsewhere (Table 1, Supplementary Table 3).

FIG. 2. Chart demonstrating changes in length of hospitalization and total charges over the years included in the training database. Length of hospitalization is measured in days, and total charges are measured to the nearest dollar. Solid line denotes length of hospitalization; dashed line denotes total charges.

RMSLE and Lift Chart

An ensemble model (from here on referred to as ensemble 1) comprising 3 gradient boosted tree regressors best predicted total charges after TSS (RMSLE 0.446, 95% CI 0.438–0.453), and generalized well to never-before-seen data (holdout RMSLE 0.446). A second ensemble (from here on referred to as ensemble 2), also made up of 3 gradient boosted tree regressors, best predicted the outcome of interest in the absence of LOS as a variable, though with less fidelity than the first ensemble (RMSLE 0.522, 95% CI 0.514–0.530, holdout 0.526). Lift charts were constructed to visualize the accuracy of each ensemble (Fig. 3).

Permutation Importance and Partial Dependence

Permutation importance and partial dependence analyses demonstrate which variables are most important to, and how they independently influence, each of the ensemble models. For ensemble 1, LOS, followed by non-Southern hospital region, nonelective admission, nonwhite race, and high case volume (defined as > 50 TSS procedures per calendar year) were the strongest predictors of higher total charges (Fig. 4). LOS increases predicted total charges from \$36,172 for a single day to \$142,081 for 18 days; elective admission was predicted to be cheaper (\$55,428) than urgent (\$58,969) or emergent (\$59,611) admissions; surgery in the Northeast (\$65,748), West (\$63,036), and Midwest (\$60,633) had higher predicted total charges than surgery in the South (\$51,015); white patients had lower predicted total charges (\$60,054) than black (\$62,384), Hispanic (\$63,675), and Native American (\$60,468) race, and total charges were roughly comparable to Asian/Pacific Islander (\$58,988) and other race patients (\$60,031). We note that roughly 20% of patient race data were missing in the NIS data set. For this reason, we kept “missing race” as its own variable in the ensemble. Patients with missing race had total predicted charges of \$55,427. Patients seen at hospitals with a high volume of TSS procedures had higher predicted charges (\$65,995) than patients seen at low-volume hospitals (\$57,578) (Fig. 5).

For ensemble 2 (excluding LOS), the strongest predictors of higher total charges were nonelective admission,

TABLE 1. Selected patient characteristics

Variable	Total No. (%)	Mean Total Charges (SD)*	p Value†
Total hospitalizations	15487	\$60491 (\$63493)	
Mean age, yrs	51.7 (15.9)		<0.001
≤40	4025 (26.0)	\$56921 (\$60468)	
40–51	3438 (22.2)	\$59348 (\$61418)	
52–63	4129 (26.7)	\$62741 (\$68492)	
64+	3895 (25.1)	\$62803 (\$62657)	
Sex			0.08
Female	7857 (50.7)	\$58886 (\$56784)	
Male	7504 (48.5)	\$62141 (\$70165)	
Missing	126 (0.8)	\$62303 (\$30731)	
Race			<0.001
White	7664 (49.5)	\$58679 (\$60010)	
Black	1999 (12.9)	\$68862 (\$72065)	
Hispanic	1514 (9.8)	\$73953 (\$71603)	
Asian/Pacific Islander	456 (2.9)	\$74226 (\$97415)	
Native American	53 (0.3)	\$73009 (\$56735)	
Other	406 (2.6)	\$67572 (\$80096)	
Missing	3395 (21.9)	\$50761 (\$50744)	
Admission type			<0.001
Elective	11539 (74.5)	\$50988 (\$49078)	
Emergent	1034 (6.7)	\$93058 (\$96111)	
Urgent	967 (6.2)	\$75372 (\$71416)	
Missing	1947 (12.6)	\$92123 (\$90298)	
Tumor type			<0.001
Benign	14501 (93.6)	\$59616 (\$60854)	
Malignant	103 (0.7)	\$84355 (\$16789)	
Secondary malignant	72 (0.5)	\$97050 (\$107184)	
Missing	811 (5.2)	\$69848 (\$77010)	
Cushing's disease			0.27
Yes	1168 (7.5)	\$61090 (\$62310)	
No	14319 (92.5)	\$60443 (\$63590)	
Acromegaly			<0.001
Yes	1063 (6.9)	\$51927 (\$39795)	
No	14424 (93.1)	\$61122 (\$64853)	
Noniatrogenic panhypopituitarism			<0.001
Yes	1039 (6.7)	\$88517 (\$11690)	
No	14448 (93.3)	\$58475 (\$57260)	
Other hormone-secreting tumor			0.11
Yes	244 (1.6)	\$57247 (\$77366)	
No	15243 (98.4)	\$60543 (\$63247)	
Visual field deficit			<0.001
Yes	1481 (9.6)	\$72908 (\$66477)	
No	14006 (90.4)	\$59178 (\$63028)	
Resection type			0.002
Partial	11650 (75.2)	\$61365 (\$64896)	
Total	3837 (24.8)	\$57837 (\$58956)	

CONTINUED IN NEXT COLUMN »

» CONTINUED FROM PREVIOUS COLUMN

TABLE 1. Selected patient characteristics

Variable	Total No. (%)	Mean Total Charges (SD)*	p Value†
CSF leak			<0.001
Yes	1884 (12.2)	\$91174 (\$10684)	
No	13603 (87.8)	\$56241 (\$53486)	
CNS infection			<0.001
Yes	49 (0.3)	\$229435 (\$256951)	
No	15438 (99.7)	\$59954 (\$61220)	
Diabetes insipidus			<0.001
Yes	1724 (11.1)	\$86261 (\$10521)	
No	13763 (88.9)	\$57263 (\$55289)	
Any postop complication			<0.001
Yes	3867 (25.0)	\$91212 (\$104211)	
No	11620 (75.0)	\$50267 (\$36618)	
Length of hospitalization			<0.001
1	1157 (7.5)	\$36973 (\$22616)	
2	3671 (23.7)	\$42881 (\$25170)	
3	3575 (23.1)	\$45900 (\$25851)	
4	2511 (16.2)	\$52346 (\$31142)	
5	1282 (8.3)	\$61545 (\$37081)	
6–10	2308 (14.9)	\$76770 (\$4611)	
≥11	977 (6.3)	\$185548 (\$172749)	
Hospital region			<0.001
South	5928 (38.3)	\$51941 (\$50261)	
West	3866 (24.9)	\$72619 (\$73679)	
Northeast	3019 (19.5)	\$66489 (\$79763)	
Midwest	2674 (17.3)	\$55136 (\$47715)	
Hospital volume			0.001
High volume	4520 (29.2)	\$61845 (\$67312)	
Low volume	10967 (70.8)	\$59933 (\$61844)	

* All values rounded to the nearest US dollar.

† Mann-Whitney U-test (comparison of 2 groups) and one-way ANOVA test (comparison of more than 2 groups) for significant difference between groups. Analyses do not include missing data.

non-Southern hospital region, postoperative complication, non-white race, and private investor hospital control (Fig. 4). Again, elective admission was predicted to be cheaper (\$54,422) than urgent (\$63,683) or emergent (\$67,221) admissions; surgery in the Northeast (\$64,984), West (\$63,411), and Midwest (\$63,799) had higher predicted total charges than surgery in the South (\$54,824); any complication increased predicted charges from \$56,227 to \$67,953; white patients had lower predicted total charges (\$60,526) than black (\$64,913), Hispanic (\$68,116), Native American (\$64,307), Asian/Pacific Islander (\$62,136), and other (\$61,462) race patients and were higher than those for missing (\$54,653) race patients; and patients treated at hospitals under the control of private investors had higher total charges (\$82,696) than those seen at private not-for-profit hospitals (\$60,457), government/nonfederal hospitals (\$57,072), or hospitals under unknown control (\$60,384) (Fig. 5).

FIG. 3. A and B: Lift charts demonstrating graphically the accuracy of predicted total charges relative to actual total charges for each ensemble (with and without LOS as a variable). Predicted total charges are divided into 10 equal bins, or deciles. Mean predicted total charges and mean actual total charges are calculated and plotted for each decile bin. Solid line denotes actual total charges; dashed line denotes predicted total charges.

Discussion

The more efficiently resources are allocated in the healthcare system, the more patients are able to access these resources, including physicians, equipment, supplies, and time. Total charges are one representation of resource utilization during a hospitalization, and understanding risk factors for high total charges provides practitioners an opportunity to address potentially reversible drivers of high charges.

In this study, we used ML to build ensemble models that directly predict total charges following TSS for pituitary tumor. To our knowledge, this represents the first attempt to directly model total charges in this patient population. It is important to note that our model predicts total hospital charges and not total costs. Charges are the initial list prices a hospital sets for individual items and services it provides while costs represent the actual expenses incurred during a hospitalization. Medicare requires hospitals to submit total hospital charges for regulatory purposes, but it is rarely the case that insurance companies or consumers are asked to pay full charges. Charges and costs are intrinsically linked, however, and decreasing charges will likely result in decreasing costs. Developing an ensemble that predicts total costs directly is an important future direction for this work.

Univariate analysis of the 71 included variables demonstrated that many of the variables, including patient comorbidities and postoperative complications, are associated with significantly higher total charges. However, the incidence of any of these variables in a given hospitalization is relatively rare, making them poor targets for interventions meant to decrease costs for TSS more broadly. Permutation importance analysis shows us which variables influence the ensemble model predictions on a global level, potentially identifying more actionable targets.

We found that LOS is by far the strongest predictor of total charges, increasing predicted total charges by roughly \$5000 per day. The influence of this variable is so strong that, by comparison, the relative impact of all other variables is nearly negligible (Fig. 4). These findings are consistent with previous studies investigating costs for pituitary surgery.^{17,25} Decreasing LOS is therefore an obvious target in any cost-reduction strategy for this patient population, and has already been shown to be feasible and safe for patients who are medically ready for discharge.¹⁰ The use of evidence-based guidelines, as well as specialized pituitary centers, may increase a patient's ability to obtain early postoperative discharge and decrease the main driver of increased hospital charges.^{27,29} We note, however, that high-volume TSS centers were associated with higher total

FIG. 4. A and B: Permutation importance analyses demonstrating the relative importance of the 5 most influential variables on the predictions of both ensembles. The most important variable is assigned the value “1.0” and all other variables are assigned numerical values based on their importance relative to the most important variable.

charges, an association also seen in the treatment of cerebrovascular malformations.⁸ This is despite the fact that high-volume TSS centers had decreased average LOS (3.6 days vs 5.1 days, $p < 0.001$), healthier patients (average preoperative comorbidities 1.29 vs 1.45, $p < 0.001$), and similar complication rates (30.7% vs 30.6%, $p = 0.89$) compared to low-volume centers (Supplementary Table 4). It is possible that higher total charges at high-volume centers may be due to the ability of more experienced centers to demand higher procedure charges, though further work is needed to understand differential pricing at different volume centers.

Interestingly, the mean LOS after TSS decreased significantly between 2002 and 2011 (from 5.5 days to 4.2 days, $p < 0.001$), but the mean total charges rose significantly over the same time period (from \$46,443 to \$76,228, $p < 0.001$), even after controlling for inflation (Fig. 2). Bivariate analysis demonstrated that, when compared to patients in 2002, patients in 2011 tended to have more preoperative comorbidities, more preoperative hormonal abnormalities, and higher rates of complications. The geographic distribution of operations also changed, with a greater proportion of operations performed in the Northeast in 2011 compared to 2002 (Supplementary Table 5).

It is possible that increased total charges seen in 2011 compared to 2002 can be explained by higher numbers of operations taking place in the more expensive Northeast region. It is also possible that increased total charges in 2011 are a consequence of operating on sicker patients. As use of electronic health records became more widespread over this interval, it is also possible that patients

treated later were more likely to have comorbidities or complications coded. Partial dependence analysis demonstrates that, even in the presence of LOS, there remains a positive linear relationship between number of preoperative comorbidities and total charges (\$58,345 for patients with no comorbidities versus \$62,486 for patients with 8 comorbidities; Supplementary Fig. 1). We hypothesized that sicker patients may also be more prone to postoperative complication, which itself is independently associated with increased total charges.²⁸ We performed bivariate analysis, which demonstrated that risk of postoperative complication increases linearly with number of preoperative comorbidities (Table 2). The fact that, on average, sicker patients were operated on in 2011 than 2002 might partially explain both the increased number of postoperative complications and higher total charges in 2011.

That LOS declined between 2002 and 2011 without a corresponding decrease in total charges also suggests that decreasing LOS alone may not be sufficient to maintain a sustainable reduction of charges and that other drivers of charges need to be identified and intervened on. With this in mind, and hoping to unearth independent predictors whose effect on total charges had been initially obscured by the overwhelming influence of LOS, we built a second ensemble model excluding LOS as a variable. Analysis of ensemble 2 showed that nonelective admission, non-Southern hospital geography, the presence of any postoperative complication, non-white race, and surgery at private investor hospital most strongly predict higher total charges.

Given their importance in ensemble 1, it is likely that admission type, hospital region, and patient race exert their influence on ensemble predictions at least partially independent of their influence on LOS. Hospital ownership type was the 6th most important variable in ensemble 1, so we were also unsurprised to see it appear in the top 5 most important variables in ensemble 2 (Supplementary Fig. 2). The presence of a complication rose significantly in importance between ensemble 1 and ensemble 2 (from a relative importance of 0.02 to a relative importance of 0.28), suggesting that its predictive value is likely as a pseudomarker for LOS. Conversely, the relative importance of hospital volume actually remained similar between the ensembles (0.07 for ensemble 1 vs 0.09 for ensemble 2), despite dropping below other variables in terms of importance to ensemble predictions, suggesting that the influence of hospital volume on total charges is relatively independent of LOS. This is consistent with the fact that high-volume hospitals have shorter LOS but higher total charges.

Regardless of the extent to which these variables directly or indirectly influence charges, they can still be targeted for cost reduction. We discuss several of these variables below.

Geography

Healthcare spending in the United States in general, and for TSS specifically, varies widely by geography.^{5,6,13,20,23} It is not always the case, however, that higher costs in different geographic regions correspond to better health outcomes.²⁶ Understanding and reconciling regional differences in practice and referral patterns is thus an important step in reducing wasteful charges for TSS. Some have also

FIG. 5. Partial dependence plots demonstrating the independent impact of individual variables on the ensemble models. Left-side x-axis represents patient incidence for each patient group and corresponds to *bars*. Right-side x-axis represents predicted total charges and corresponds to *round heads*. **A–E:** Graphs depicting variables in ensemble 1 (with LOS). **F–J:** Graphs illustrating variables in ensemble 2 (without LOS). NFP = not for profit.

TABLE 2. Percentage of patients with postoperative complications by number of preoperative comorbidities

No. of Comorbidities	Complication (%)			p Value*
	Total (n = 15487)	Yes	No	
None	4795 (31.0)	967 (20.2)	3828 (79.8)	<0.001
1	4384 (28.3)	1050 (24.0)	3334 (76.0)	
2	3381 (21.8)	883 (26.1)	2498 (73.9)	
3	1765 (11.4)	529 (30.0)	1236 (70.0)	
4–6	1118 (7.2)	411 (36.8)	707 (63.2)	
≥7	44 (0.3)	27 (61.4)	17 (38.6)	

* Chi-square test used to calculate significance between groups.

advocated for greater transparency of hospital charges, which can help patients make more informed decisions about where they want to receive care, which may in turn introduce a level of competition that incentivizes charge reduction.²⁰

Postoperative Complication

Previous research has demonstrated an association between the presence of postoperative complications and increased LOS and increased hospital costs.¹⁸ Using the same preoperative variables available in our total charges ensemble model, we attempted to model postoperative complication. Though the accuracy of the model was not excellent (area under the curve of the receiver-operator characteristic curve 0.66, 95% CI 0.64–0.68; Supplementary Fig. 3), conclusions drawn from the model may be informative: We found that younger patient age, the presence of preoperative electrolyte abnormalities, and nonelective operations were the most important predictors of postoperative complication (Supplementary Fig. 4). We were surprised to find that younger patients were more likely to have postoperative complications, but we did note that the risk of complications decreases only 5% from the youngest to the oldest patients (from 28% to 23%), suggesting that a confluence of factors with small influences come together to define a patient's risk of complication.

Minority Race

It is well established that racial disparities exist in healthcare,^{1,15} and our study demonstrates that minority race predicts higher total charges for TSS. It is important to note that our ensemble models predict total charges rather than total costs, so this disparity does not reflect a difference in access to insurance between minority and white patients. It is possible that decreased access to healthcare results in minority patients presenting to providers with more advanced disease, driving up charges. We attempted to capture this using elective versus urgent or emergent admission type, though this may be a poor surrogate. It may also be the case that minority race acts as a marker for extended LOS, as has been shown in previous studies.²⁴ However, minority race exerts an independent impact on total charges in our ensemble 1, which includes LOS as a variable, suggesting that the relationship between minority race and LOS does not entirely explain higher charges for

minority patients. Further work is thus needed to understand how race influences total hospital charges, and how to mitigate its effects.

Hospital Ownership

We found that for-profit, private investor hospital ownership is associated with higher total charges for TSS, even in the presence of specific patient characteristics, including preoperative patient comorbidities and postoperative complications. Bivariate analysis also demonstrated no difference in average number of patient comorbidities or complications between private investor hospitals and other hospital types (Supplementary Table 6). This is in line with previous research that found that the greatest difference in charges between hospital ownership types lies not in underlying differences in patient populations but in charges for ancillary services, such as laboratory services and diagnostic radiology. Higher utilization of these more discretionary services at private investor hospitals may reflect the for-profit nature of these institutions.⁹

We also investigated the influence of hospital teaching status on predicted total charges. We found very little difference in total predicted charges between patients seen at urban teaching (\$61,327) and urban nonteaching (\$60,662) hospitals. Furthermore, the relative importance of hospital teaching status on the ensemble predictions was only 0.04 (compared to 1 for nonelective admission), suggesting that the presence of trainees in the hospital setting has only a minor independent influence on total charges (Supplementary Fig. 5).

We do not dismiss the importance of recognizing opportunities to decrease total charges on the level of the individual patient. For this reason, we generated partial dependence plots for all patient-specific variables included in ensemble 2 and calculated the added expense to patients in the presence of each variable. We ranked the variables by how much cost was added to identify the most expensive patient characteristics and found that the presence of peripheral vascular complications (including deep venous thrombosis, pulmonary embolism, and placement of an inferior vena cava filter) (added expense of \$51,612) was the most expensive patient characteristic, followed by hydrocephalus (\$43,403), respiratory complications (\$35,315), intracranial hemorrhage (\$28,129), and CNS infection (\$25,694). These, too, represent potential points of intervention to decrease total charges on the level of the individual patient.

Strengths and Limitations

Our study has several strengths: We used the NIS, a large, multiinstitution database that is well known and well validated, allowing us to draw broadly generalizable conclusions from our findings. Our ML technique is another strength. By ranking ML algorithms before final algorithm selection, we were able to identify which algorithms most accurately predict our outcome of interest, allowing us to confidently build the most predictive models from our data. This increased our confidence in the validity of the clinical insights gained from the model itself.

Our study also has some important limitations: First, we trained and validated our predictive ensembles on the

same retrospective database. Though we used a holdout validation technique to demonstrate generalizability of the ensembles to data never used in training, demonstration of generalizability to a separate database, or to prospectively collected data, would serve as a stronger validation. We also relied on data coded and collected outside of our own institution, which is prone to human error. Also, given the lack of uniformity in NIS data collection after 2011, we were limited to utilizing only cases between 2002 and 2011. Repeating the study with more current data is an important future direction for this work.

Conclusions

Identifying drivers of total charges provides an opportunity to reduce waste in the healthcare system, ensuring that resources are allocated as broadly and equitably as possible. In this study, we built an ML ensemble that directly predicts total charges for TSS with good fidelity and that can be used by physicians, insurers, and patients to better understand total charges following surgery. Similar modeling may be developed to understand charges for other aspects of neurosurgical practice.

LOS is the strongest predictor of total charges following TSS. Nonelective admission, non-Southern hospital location, postoperative complication, minority patient race, and private investor hospital control also predict higher charges, potentially by influencing LOS. Interventions aimed at minimizing the effects of these variables can improve efficiency in the resource-limited healthcare system, leading to higher quality care and improved outcomes for more patients.

Acknowledgments

W.E.M. received financial support from the Vanderbilt Medical Scholars Program and an NIH Clinical and Translational Science Awards Program grant (UL1 RR 024975).

References

- Agency for Healthcare Research and Quality: **2015 National Healthcare Disparities Report and Disparities Report and 5th Anniversary Update on the National Quality Strategy**. Rockville, MD: U.S. Department of Health and Human Services, 2016 (<http://www.ahrq.gov/research/findings/nhqrdr/nhqrdr15/index.html>) [Accessed May 29, 2018]
- Bodenheimer T, Fernandez A: High and rising health care costs. Part 4: can costs be controlled while preserving quality? **Ann Intern Med** **143**:26–31, 2005
- Breiman L: Random forests. **Mach Learn** **45**:5–32, 2001
- Bureau of Labor Statistics: **Consumer Price Index (CPI) Databases**. Washington, DC: U.S. Department of Labor, 2017 (<https://www.bls.gov/cpi/data.htm>) [Accessed May 29, 2018]
- Burke MA, Fournier GM, Prasad K: **Physician Social Networks and Geographical Variation in Medical Care**. Washington, DC: Brookings Institute, 2003 (https://www.brookings.edu/wp-content/uploads/2016/06/07healthcare_burke.pdf) [Accessed May 29, 2018]
- Cebul RD, Rebitzer JB, Taylor LJ, Votruba ME: Organizational fragmentation and care quality in the U.S healthcare system. **J Econ Perspect** **22**:93–113, 2008
- Centers for Medicare and Medicaid Services: **CMS' Value-Based Programs**. Baltimore: Centers of Medicare and Medicaid Services, 2017 (<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/Value-Based-Programs.html>) [Accessed May 29, 2018]
- Davies JM, Lawton MT: Improved outcomes for patients with cerebrovascular malformations at high-volume centers: the impact of surgeon and hospital volume in the United States, 2000–2009. **J Neurosurg** **127**:69–80, 2017
- Eskoz R, Peddecord KM: The relationship of hospital ownership and service composition to hospital charges. **Health Care Financ Rev** **6**:51–58, 1985
- Forbes JA, Wilkerson J, Chambless L, Shay SD, Elswick CM, Ablitt PW, et al: Safety and cost effectiveness of early discharge following microscopic trans-sphenoidal resection of pituitary lesions. **Surg Neurol Int** **2**:66, 2011
- Fraser I, Encinosa W, Glied S: Improving efficiency and value in health care: introduction. **Health Serv Res** **43**:1781–1786, 2008
- Friedman JH: Greedy function approximation: a gradient boosting machine. **Ann Stat** **29**:1189–1232, 2001
- Garber AM, Skinner J: Is American health care uniquely inefficient? **J Econ Perspect** **22**:27–50, 2008
- Healthcare Cost and Utilization Project Databases: **Nationwide Inpatient Sample**. Rockville, MD: Agency for Healthcare Research and Quality, 2018 (<http://www.hcup-us.ahrq.gov/nisoverview.jsp>) [Accessed May 29, 2018]
- Institute of Medicine: **Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care**. Washington, DC: National Academies Press, 2003
- Institute of Medicine (US) Committee on Quality of Health Care in America: **Crossing the Quality Chasm: A New Health System for the 21st Century**. Washington, DC: National Academies Press, 2001
- Karsy M, Brock AA, Guan J, Bisson EF, Couldwell WT: Assessment of cost drivers in transsphenoidal approaches for resection of pituitary tumors using the value-driven outcome database. **World Neurosurg** **105**:818–823, 2017
- Khan NA, Quan H, Bugar JM, Lemaire JB, Brant R, Ghali WA: Association of postoperative complications with hospital costs and length of stay in a tertiary care center. **J Gen Intern Med** **21**:177–180, 2006
- Kramolowsky EV, Wood NL, Rollins KL, Glasheen WP, Nelson CM: Impact of physician awareness on hospital charges for radical retropubic prostatectomy. **J Urol** **154**:139–142, 1995
- Lee CC, Kimmell KT, Lalonde A, Salzman P, Miller MC, Calvi LM, et al: Geographic variation in cost of care for pituitary tumor surgery. **Pituitary** **19**:515–521, 2016
- Lee SM, Kang JO, Suh YM: Comparison of hospital charge prediction models for colorectal cancer patients: neural network vs. decision tree models. **J Korean Med Sci** **19**:677–681, 2004
- Leibman BD, Dilliogluligil O, Abbas F, Tanli S, Kattan MW, Scardino PT: Impact of a clinical pathway for radical retropubic prostatectomy. **Urology** **52**:94–99, 1998
- MaCurdy T, Bhattacharya J, Perlroth D, Shafrin J, Au-Yeung A, Bashour H, et al: **Geographic Variation in Spending, Utilization, and Quality: Medicare and Medicaid Beneficiaries**. Washington, DC: National Academy of Sciences, 2013 (<http://www.nationalacademies.org/hmd/~media/Files/Report%20Files/2013/Geographic-Variation/Sub-Contractor/Acumen-Medicare-Medicaid.pdf>) [Accessed May 29, 2018]
- McLaughlin N, Martin NA, Upadhyaya P, Bari AA, Buxey F, Wang MB, et al: Assessing the cost of contemporary pituitary care. **Neurosurg Focus** **37**(5):E7, 2014
- Muhlestein WE, Akagi DS, Chotai S, Chambless LB: The impact of race on discharge disposition and length of hospi-

- talization after craniotomy for brain tumor. **World Neurosurg** 104:24–38, 2017
26. Newhouse JP, Garber AM: Geographic variation in health care spending in the United States: insights from an Institute of Medicine report. **JAMA** 310:1227–1228, 2013
 27. Sarkiss CA, Lee J, Papin JA, Geer EB, Banik R, Rucker JC, et al: Pilot study on early postoperative discharge in pituitary adenoma patients: effect of socioeconomic factors and benefit of specialized pituitary centers. **J Neurol Surg B Skull Base** 76:323–330, 2015
 28. Tetreault L, Tan G, Kopjar B, Côté P, Arnold P, Nugaeva N, et al: Clinical and surgical predictors of complications following surgery for the treatment of cervical spondylotic myelopathy: results from the multicenter, prospective AOSpine International Study of 479 patients. **Neurosurgery** 79:33–44, 2016
 29. Thomas JG, Gadgil N, Samson SL, Takashima M, Yoshor D: Prospective trial of a short hospital stay protocol after endoscopic endonasal pituitary adenoma surgery. **World Neurosurg** 81:576–583, 2014
 30. Wang J, Li M, Hu YT, Zhu Y: Comparison of hospital charge prediction models for gastric cancer patients: neural network vs. decision tree models. **BMC Health Serv Res** 9:161, 2009

Disclosures

D.S.A. is an employee of and data scientist at DataRobot, Inc.
W.E.M. is married to D.S.A.

Author Contributions

Conception and design: Muhlestein, Chambless. Acquisition of data: Muhlestein, Akagi. Analysis and interpretation of data: Muhlestein, Akagi. Drafting the article: Muhlestein, Akagi, McManus. Critically revising the article: Muhlestein, Chambless. Reviewed submitted version of manuscript: Muhlestein. Approved the final version of the manuscript on behalf of all authors: Muhlestein. Statistical analysis: Muhlestein. Study supervision: Chambless.

Supplemental Information

Online-Only Content

Supplemental material is available with the online version of the article.

Supplementary Tables and Figures. <https://thejns.org/doi/suppl/10.3171/2018.4.JNS18306>.

Correspondence

Whitney E. Muhlestein: Vanderbilt University, Vanderbilt University Medical Center, Nashville, TN. whitney.muhlestein@gmail.com.