J Shoulder Elbow Surg (2022) ■, 1–12

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

Identifying modifiable and nonmodifiable cost drivers of ambulatory rotator cuff repair: a machine learning analysis

Yining Lu, MD[a], Joshua R. Labott, MD[a], Harold I. Salmons IV, MD[a], Benjamin D. Gross, BS[b], Jonathan D. Barlow, MD, MS[a], Joaquin Sanchez-Sotelo, MD, PhD[a], Christopher L. Camp, MD[a,*]

[a]*Department of Orthopedic Surgery, Mayo Clinic, Rochester, MN, USA*
[b]*Icahn School of Medicine at Mount Sinai, New York, NY, USA*

Introduction: Implementing novel tools that identify contributors to the cost of orthopedic procedures can help hospitals maximize efficiency, minimize waste, improve surgical decision-making, and practice value-based care. The purpose of this study was to develop and internally validate a machine learning algorithm to identify key drivers of total charges after ambulatory arthroscopic rotator cuff repair and compare its performance with a state-of-the-art statistical learning model.
Methods: A retrospective review of the New York State Ambulatory Surgery and Services Database was performed to identify patients who underwent elective outpatient rotator cuff repair (RCR) from 2015 to 2016. Initial models were constructed using patient characteristics (age, gender, insurance status, patient income, Elixhauser Comorbidity Index) as well as intraoperative variables (concomitant procedures and services, operative time). These were subsequently entered into 5 separate machine learning algorithms and a generalized additive model using natural splines. Global variable importance and partial dependence curves were constructed to identify the greatest contributors to cost.
Results: A total of 33,976 patients undergoing ambulatory RCR were included. Median total charges after ambulatory RCR were \$16,017 (interquartile range: \$11,009-\$22,510). The ensemble model outperformed the generalized additive model and demonstrated the best performance on internal validation (root mean squared error: \$7112, 95% confidence interval: 7036-7188; logarithmic root mean squared error: 0.354, 95% confidence interval: 0.336-0.373, R^2: 0.53), and identified major drivers of total charges after RCR as increasing operating room time, patient income level, number of anchors used, use of local infiltration anesthesia/peripheral nerve blocks, non-White race/ethnicity, and concurrent distal clavicle excision. The model was integrated into a web-based open-access application capable of providing individual predictions and explanations on a case-by-case basis.
Conclusion: This study developed an ensemble supervised machine learning algorithm that outperformed a sophisticated statistical learning model in predicting total charges after ambulatory RCR. Important contributors to total charges included operating room time, duration of care, number of anchors used, type of anesthesia, concomitant distal clavicle excision, community characteristics, and patient demographic factors. Generation of a patient-specific payment schedule based on the Agency for Healthcare Research and Quality risk of mortality highlighted the financial risk assumed by physicians in flat episodic reimbursement schedules given variable patient comorbidities and the importance of an accurate prediction algorithm to appropriately reward high-value care at low costs.
Level of evidence: Level IV; Economic Analysis

Keywords: Machine learning; artificial intelligence; cost; value-based care; arthroscopic rotator cuff repair; orthopedic

Investigation performed at the Mayo Clinic, Rochester, MN, USA.
Institutional review board approval was not required for this economic analysis.

*Reprint requests: Christopher L. Camp, MD, Department of Orthopedic Surgery, Mayo Clinic, 200 First Street S.W., Rochester, MN 55905, USA.
E-mail address: camp.christopher@mayo.edu (C.L. Camp).

1058-2746/\$ - see front matter
https://doi.org/10.1016/j.jse.2022.04.008

Arthroscopic rotator cuff repair (ARCR) is the most common arthroscopic shoulder procedure performed in the United States.[8] Although the effective treatment of rotator cuff tears can range from nonoperative management (eg, physical therapy) to surgical intervention, there has been a substantial increase in the incidence of ARCR over the last decade.[2] Because of improvements in surgical, anesthesia, and pain management techniques, many ARCRs are now performed in an outpatient setting.[9,17] Outpatient surgery, especially ARCR, offers improved patient safety and satisfaction, less scheduling delays, and greater cost savings compared with traditional impatient surgery.[9,18]

Health care in much of the United States is slowly shifting from fee-for service to a value-based payment system.[15] Value-based care models connect reimbursement with the quality of care provided, rewarding physicians for efficiency and effectiveness[16] in order to provide better care for patients, improve population health management strategies, and reduce health care costs.[4,13] The success of ambulatory shoulder arthroscopy depends on minimizing procedural-related costs and complications in order to decrease cost. Thus, it is vital to identify and understand the drivers of cost for these procedures to optimize patient outcomes, lower costs, and ensure high-quality patient care.

Machine learning (ML) models have been increasingly used across orthopedic surgery to predict outcomes, complications, satisfaction, and discharge disposition, both successfully and accurately. ML models use algorithms that can learn from past examples and then adjust internal parameters to improve associations and data accuracy.[28,32] Such models can provide superior predictive ability compared with traditional statistics,[27] and particularly linear modeling approaches, especially when the contributions from cost drivers are nonlinear in nature. Another significant strength of ML predictive models is the level of flexibility to differentially assess the contribution of each input feature on an individual-by-individual basis, as opposed to traditional linear regressions in which each predictor is constrained to a single weight.[3] The efficiency and efficacy demonstrated by ML models in predicting cost and outcomes after other surgical procedures makes artificial intelligence–enabled cost prediction in outpatient RCR an attractive application.[19,26] Therefore, the purpose of this study was to develop and internally validate an ML algorithm to identify key drivers of total charges after ambulatory RCR and to compare its performance with a state-of-the-art statistical learning model. We hypothesize that ML algorithms will outperform traditional methods in predicting key drivers of total charges after ambulatory RCR.[28,32]

Methods

This is a retrospective cohort study of an anonymized national database. After data retrieval using an institutional review board–exempt anonymized national outcomes database, we performed interpretations of our data adherent to the Transparent Reporting of a multivariate prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines and the guidelines for developing and reporting machine learning models in biomedical research.[7,23]

Database

Data were collected using the State Ambulatory Surgery and Services Database (SASD) for the State of New York (NYSASD) (Healthcare Cost and Utilization Project [HCUP], Agency for Healthcare Research and Quality). The NYSASD was queried to identify all patients who underwent an outpatient elective ARCR repair from 2015 to 2016. SASD is one of many national databases that are part of the HCUP, a comprehensive and well-validated data source on health care sponsored by the Agency for Healthcare Research and Quality (Rockville, MD).2[6] The SASD includes encounter-level data for outpatient ambulatory surgery from hospital-owned and free-standing facilities from 35 states. We chose the New York state registry as it included integral variables for cost analysis such as operating room (OR) time and anesthesia type. The SASD also contains uniform clinical information on all patients, regardless of payer, including those covered by Medicare, Medicaid, private insurance, and the uninsured.

Data collection

The Current Procedural Terminology (CPT) code 29827 (arthroscopy, shoulder, surgical; with rotator cuff repair) was used to identify the subset of patients who underwent an ARCR. Patients who underwent the following concomitant procedures were excluded: pectoralis tendon repair, acromioclavicular joint reconstruction, tendon transfers, and distal biceps tendon repairs. In addition, ambulatory patients with unplanned extended postoperative stay were identified as those with total length of stay ≥ 1 day and were also excluded to limit our study population to same-day discharges.

Input variables

Feature selection was performed using more than 30 demographic, clinical, and intraoperative variables (Supplementary Table S1) abstracted from the database. The outcome of interest was total charges per care episode associated with each ambulatory RCR surgery. To adjust for month-to-month inflation, all charges were adjusted by multiplying by the appropriate consumer price index provided by the Bureau of Labor Statistics[37]; this adjustment was performed to prevent introducing temporal biases into the effects of any covariates. Patient comorbidities were standardized by calculating the Elixhauser Comorbidity Index (ECI) from International Classification of Diseases, Ninth Revision and International Classification of Diseases, Tenth Revision codes associated with each patient care episode[12] as well as secondarily the Agency for Healthcare Research and Quality (AHRQ) weight-adjusted risk of mortality.

Missing data and feature selection

To prevent the introduction of biases by missing data during model development, we treated missing data based on the established method of multiple imputation, following Rubin's rule for estimation of parameter variability and confidence intervals (CIs). Briefly, features with missing data were evaluated and imputed if missing in <30% of cases the "missForest" multiple imputation method was used to impute the missing cases; the threshold for missingness was established by Stekhoven et al,[34] who experimentally introduced missingness into training data and only found a significant deviation of the imputed values from true values at >30. A list of variables with missing data is provided in Table I. After treatment of missing data, input features were preliminarily evaluated for contribution to predictive performance through recursive feature elimination,[11] a technique that performs iterative modeling via backward elimination using different sets of inputs to characterize the influence of each feature (see Supplementary Table S2 for additional details). Among variables included in the final feature set, highly collinear features (defined as Spearman's correlation coefficients >0.5) were identified, and the lesser predictive of the collinear features were excluded. For validation of the feature selection process, we performed principal component analysis followed by unsupervised hierarchical clustering to demonstrate improved structural stratification.

Model training and selection

After input variable selection, modeling was performed using the same feature set on the following candidate algorithms: generalized linear model (GLM), random forest, extreme gradient boosted machine, elastic-net penalized linear regression, multilayer perceptron, and a gradient-boosted ensemble of the candidate algorithms. Models were trained with 10-fold cross-validation repeated 3 times and internally validated via 0.632 bootstrapping with 1000 random partitions of the dataset. This method has been found to optimize both model bias and variance and improve overall performance compared with internal validation through splitting the data into a single training and holdout set.[35] For comparison purposes, a trained general additive model with natural splines (representative of a state-of-the-art traditional statistics benchmark) as well as a GLM (most frequently encountered statistical model in the literature) was performed using the identical input feature set. For the GLM, multiple comparisons (n = 13) were adjusted using Bonferroni correction alpha set at $P < .004$.[1] Variable impact on total costs for the linear regression is presented as β coefficients and 95% CIs.

The final model was selected based on the optimization of the root mean squared error (RMSE) in units of dollars. This describes the aggregate errors (the absolute difference between predicted cost and actual cost) of the model averaged over the number of bootstraps; a normalized error was also provided in the form of logarithmic root mean squared error (RMSLE), which can aid in considering the relative error between the predicted and actual values of the total charges and has been shown to be less subjective to outlier effects.[26] The optimal model minimizes both the RMSE and the RMSLE. Finally, R^2, also known as the coefficient of determination, was estimated for the fitted algorithms and extracted. This value describes the amount of variation in the predicted charges a model can explain.[5,21] The model's predicted total charges (in dollars) as a function of observed total charges were summarized in lift plots to provide a visual calibration (Fig. 1). An elaboration of the candidate algorithms and modeling techniques are provided in Supplementary Appendix S1.

Model global and local transparency

After identifying the best fit model, interpretability enhancement was performed to offer transparency into model internal behavior. This was performed at both the global and local levels. Globally, variable importance plots and partial dependence plots were used to illustrate the overall ranking of feature performance contributions as well as the effect of individual features on model prediction, respectively. Partial dependence plots illustrate the changes in dollar amounts of total charges (*Y*-axis) as a function of perturbations in each input variable (*X*-axis), whereas other feature values are held constant. Bar plots are used for categorical/ordinal variables with discrete levels, and scatterplots with a best-fit curve are plotted for continuous variables. Local model-agnostic explanation models were generated via an independent explanation algorithm fitting the predicted probabilities to the input set of each specific patient, to provide insight into individual predictions. This was done to explain the broadest range of clinical scenarios captured within the best-performing model.

Risk-based assessment

Separate cost algorithms were then produced based on risk stratification using the AHRQ's Weighted-ECI for Risk of Mortality. Based on this definition, patients were categorized as mild (ECI $\leq$ 1), moderate (1 < ECI < 4), and severe risk of mortality (ECI > 4); predictive error margin for each model was compared and equated to the level of "risk." Theoretical risk-adjusted patient-specific payments were then calculated based on median charges for each risk category and the relative differences in model predictive error.

Digital application

The best-performing algorithm was integrated into a web-based application to demonstrate the potential integration of this application into the clinical space (following external validation). All data analysis was performed in R 4.1.1 using RStudio version 1.2.5001 (RStudio, Boston, MA, USA).

Results

Demographics

A total of 33,976 patients who underwent ambulatory RCR were included, with 57% males (n = 16,119). The median age within the cohort was 58 (interquartile range: 51-65). Most patients (65%) were White (Table I). The 2 lowest state income quartiles made up more than half of the patients (57.7%), with private insurance (48.1%) being the most used reimbursement mode. A total of 72.0% of the patients in this cohort reside in counties with a population of >1 million residents. Median total charges after ambulatory RCR in this cohort were $16,017 (interquartile range: $11,009-$22,510) (Fig. 2). Concomitant procedures identified in this cohort included the following: open biceps

Table I Baseline characteristic of study population, n = 33,976

Variables	N	Median cost ($)	IQR ($)	*P* value	Missing, N (%)
Sex					–
Male (reference)	19,397	16,119	(111,28-23,009)	.07	
Female	14,579	15,881	(10,745-22,333)		
Median income in state quartile				<.001	375 (1.10)
1 (reference)	6083	15,625	(10,693-21,806)		
2	8298	13,880	(10,170-19,792)		
3	9552	14,265	(10,035-20,828)		
4	10,043	19,652	(14,342-27,760)		
Anesthesia				<.001	5670 (16.7)
General (reference)	4390	15,308	(10,617-22,088)		
Regional	16,736	15,937	(11,441-28,167)		
General and regional	4307	18,644	(12,766-24,064)		
Other	8543	13,949	(10,081-18,636)		
Race				<.001	–
White (reference)	21,923	15,361	(10,736-22,085)		
Black	3287	17,768	(11,901-25,353)		
Hispanic	2726	17,771	(13,634-23,543)		
Asian or Pacific Islander	678	17,173	(9479-23,296)		
Native American	97	18,310	(12,661-22,047)		
Other	5265	15,884	(10,491-23,407)		
Hispanic ethnicity				<.001	9513 (28)
Not Hispanic (reference)	24,150	15,811	(10,988-22,297)		
Hispanic, White	5947	15,918	(10,535-24,451)		
Hispanic, Black	1685	16,409	(10,930-21,918)		
Hispanic, other	2194	17,982	(12,499-24,743)		
Insurance provider				<.001	–
Medicare	7009	16,362	(11,380-22,957)		
Medicaid (reference)	2385	16,222	(11,216-21,365)		
Private insurance	16,328	16,874	(11,274-23,687)		
Self-pay	1605	11,668	(9472-14,825)		
No charge	24	19,916	(18,777-22,138)		
Other (WC, VA, other government)	6625	14,928	(10,295-21,917)		
Worker's compensation	5455	14,674	(10,291-21,235)	<.001	–
NCHS classification				<.001	31 (0.09)
Central counties of ≥1 million (reference)	13,182	16,286	(10,699-23,549)		
Fringe counties of ≥1 million	11,598	19,202	(13,525-27,402)		
Counties 250,000-999,999 population	4958	12,484	(9655-16,057)		
Counties 50,000-249,999 population	1335	15,507	(11,735-20,891)		
Micropolitan counties	2148	11,304	(9029-15,180)		
Not metropolitan/micropolitan counties	755	14,834	(10,545-19,061)		
Accessory services					
Local infiltration ropivacaine	3918	20,090	(13,357-30,421)	<.001	–
Peripheral nerve block	5435	14,782	(11,379-25,802)	<.001	–
Concomitant procedures					
Biceps tenodesis (open)	1708	17,315	(11,833-25,409)	<.001	–
Biceps tenodesis (arthroscopic)	4530	19,423	(13,653-25,192)	<.001	–
Subacromial decompression	28,772	15,940	(11,019-22,766)	<.001	–
Debridement	10,570	16,228	(11,067-25,329)	<.001	–
Distal clavicle excision	8207	15,387	(11,461-22,539)	<.001	–
Capsulorrhaphy	419	20,989	(13,916-29,741)	<.001	–

WC, Worker's Compensation; *VA*, Veteran's Administration; *NCHS*, National Center for Health Statistics; *IQR*, interquartile range.
Reference levels provided for ordinal variables; bivariate *P* values generated from *t* tests for continuous variables and a 1-sample test of proportions for categorical variables; multivariate *P* values generated from 1-way analysis of variance for continuous variables and χ^2 tests for categorical variables.

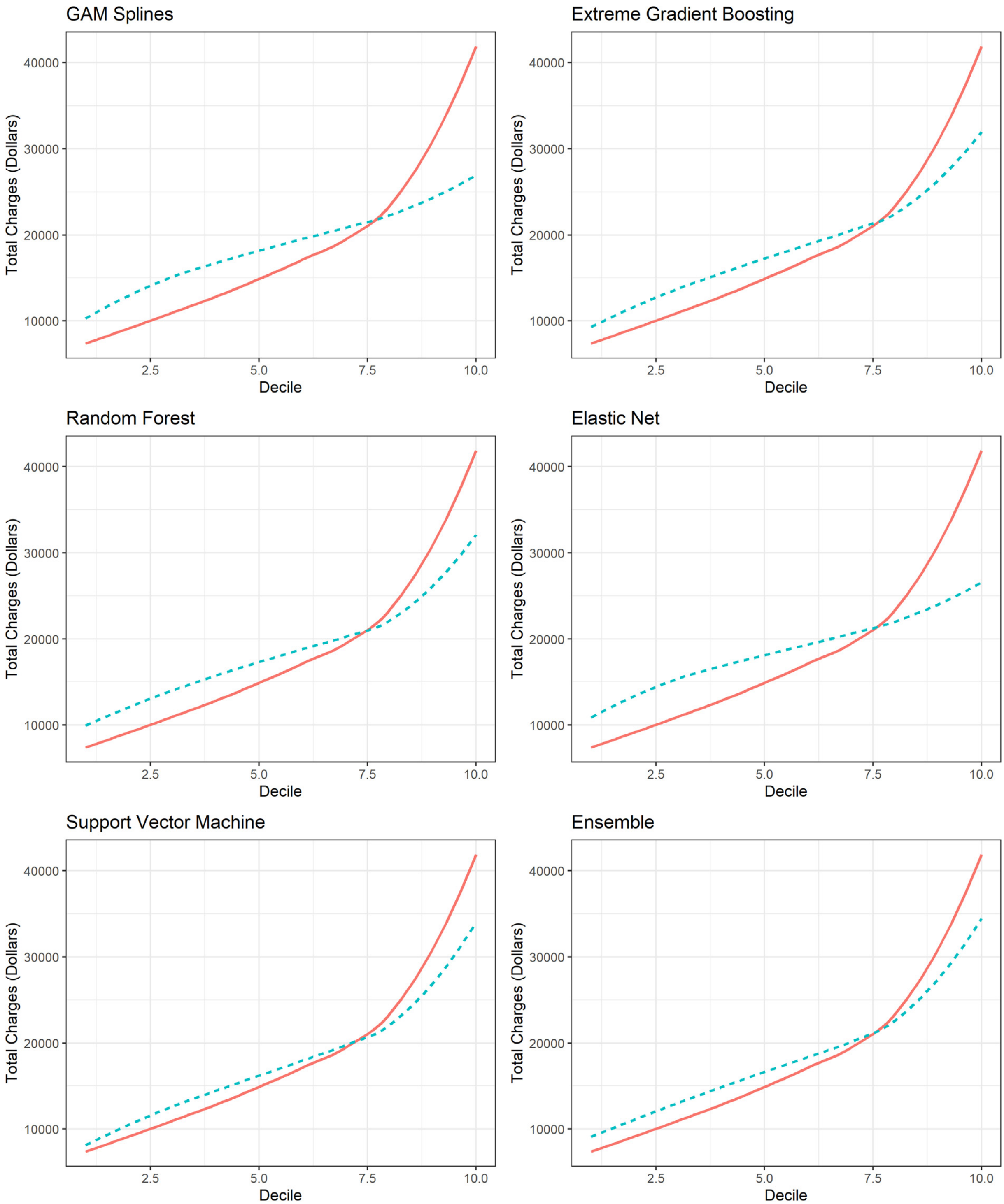

Figure 1 Lift charts produce a graphical assessment of the accuracy of predicted total charges relative to the observed total charges for candidate algorithms. Predicted total charges are divided into 10 equal bins or deciles. Mean predicted total charges ( ) and mean actual total charges ( ) are calculated and plotted for each decile bin.

tenodesis (n = 1708), arthroscopic biceps tenodesis (n = 4530), subacromial decompression (n = 28,772), débridement (n = 10,570), distal clavicle excision (DCE, n = 8207), and capsulorrhaphy (n = 419).

Comparisons to traditional statistics

The GLM model identified the following variables to be significant contributors of the variations in sample means:

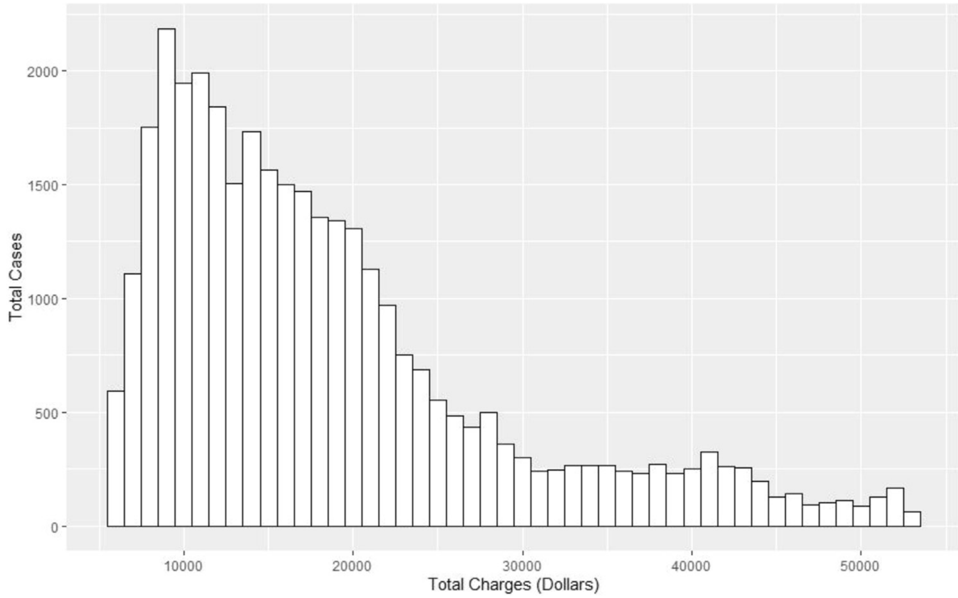

Figure 2 Histogram demonstrating distribution of total charges among cases of ambulatory rotator cuff repairs performed from 2014 to 2016.

Table II Risk factors found to predict RCR costs on multivariate linear regression

Variable	β coefficient ($)	95% CI ($)	*P* value
OR time	1729	(1545.77 to 1912.21)	<.001
General anesthesia only	−564.9	(−795.22 to −334.58)	<.001
Self-pay	−419.7	(−627.66 to −211.7)	<.001
African American race	854.1	(665.7 to 1042.51)	<.001
Hispanic race	854	(467.52 to 1240.49)	<.001
Worker's compensation	−1976.3	(−2375.27 to −1577.27)	<.001
Median income in the fourth quartile	2060.9	(1800 to 2321.86)	<.001
Number of anchors	−794.7	(−978.22 to −611.15)	<.001
Capsulorrhaphy	308.4	(132.63 to 484.18)	<.001
Naropin injection	1413.9	(1234.1 to 1593.8)	<.001
Debridement	1085.1	(905.97 to 1264.22)	<.001
Subacromial decompression	325.7	(148.48 to 502.82)	<.001
Arthroscopic biceps tenodesis	682.9	(505.26 to 860.64)	<.001
Open biceps tenodesis	428.5	(250.91 to 606.05)	<.001
Peripheral nerve block	641.8	(303.23 to 980.45)	<.001

RCR, rotator cuff repair; *OR*, operating room; *CI*, confidence interval.

increased total charges after ambulatory ARCR: 1 hour increase in OR time (β: $1729, 95% CI: $1546-$1712), median income in the fourth quartile of neighborhood (β: $2060, 95% CI: $1800-$2322), capsulorrhaphy (β: $308, 95% CI: $133-$484), ropivacaine HCl injection (β: $1414, 95% CI: $1234-$1594), subacromial decompression (β: $326, 95% CI: $148-$503), arthroscopic biceps tenodesis (β: $683, 95% CI: $505-$861), open biceps tenodesis (β: $429, 95% CI: $251-$606), and utilization of peripheral nerve block (β: $642, 95% CI: $303-$980). Complete breakdown of covariate influences is provided in Table II. Between the two, the generalized additive model outperformed GLM, although the best ML candidate outperformed both (Table III).

Model training and internal validation

The ensemble model demonstrated the best performance assessed via internal validation (RMSE: $7112, 95% CI: 7036-7188; RMSLE: 0.354, 95% CI: 0.336-0.373, R^2: 0.53) (Table III) and identified major drivers of total charges after RCR as increasing OR time, patient income level, use of local infiltration anesthesia/peripheral nerve blocks, non-White race/ethnicity, and concurrent DCE (Fig. 3). Poorly

Table III Model performances on internal validation via 0.632 bootstrap

Metric	RMSE	RMSLE	R^2
GLM	9471 (9370-9572)	0.476 (0.471-0.481)	0.16
GAM splines	9183 (9071-9266)	0.465 (0.46-0.47)	0.22
Elastic net	9258 (9160-9357)	0.495 (0.49-0.5)	0.20
Random forest	7374 (7296-7453)	0.37 (0.366-0.374)	0.50
SVM	7641 (7560-7722)	0.47 (0.469-0.479)	0.46
XGBoost	7409 (7330-7488)	0.373 (0.369-0.377)	0.49
Ensemble	7112 (7036-7188)	0.354 (0.336-0.373)	0.53

GLM, generalized linear model; *GAM,* generalized additive model; *SVM*, support vector machines; *XGBoost*, extreme gradient boosted machine; *RMSE*, root mean squared error in $; *RMSLE*, root mean squared logarithmic error.

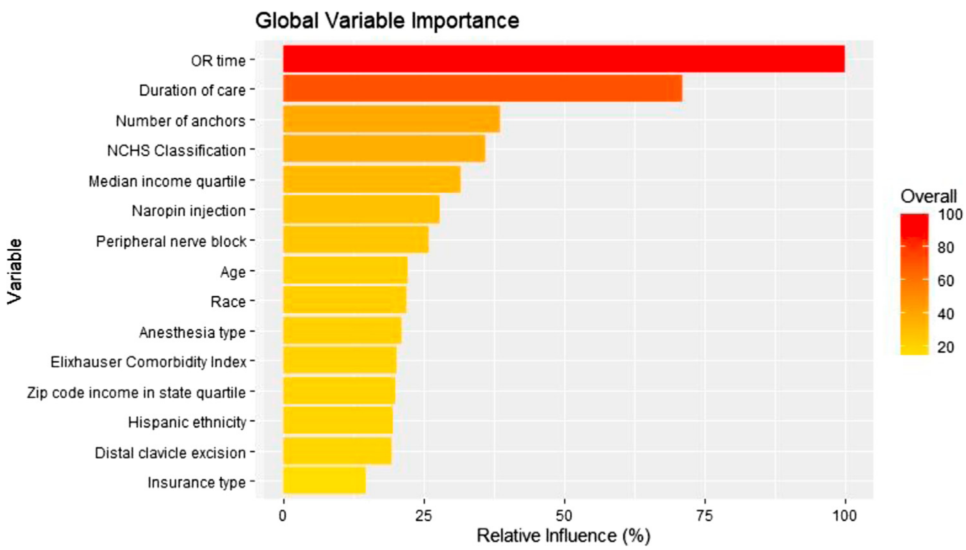

Figure 3 Variable importance plot of the ensemble model, demonstrating the normalized contributions of each input to model predictive performance. Note: an increased relative influence does not necessarily suggest a positive correlation with outcome of interest, but rather a positive correlation with the fidelity of model prediction. *OR*, operating room; *NCHS*, National Center for Health Statistics.

performing models included linear regression (RMSE: $9471, 95% CI: 9370-9572; RMSLE: 0.476, 95% CI: 0.471-0.481, R^2: 0.16), elastic net (RMSE: $9258, 95% CI: 9160-9357; RMSLE: 0.495, 95% CI: 0.49-0.5, R^2: 0.20), and generalized additive model (RMSE: $9183, 95% CI: 9071-9266; RMSLE: 0.465, 95% CI: 0.46-0.47, R^2: 0.22).

Partial dependence curves were used to evaluate the effects of continuous input features on total charges that revealed a sinusoidal dependence of cost on total hours of care and OR time. Age, ECI, and number of anchors demonstrated a linear association with cost of care (Fig. 4, A and B). A comparison of the predictive accuracies of the ensemble model after stratification of the training data by the AHRQ risk of mortality, patients with moderate and severe risk demonstrated a 2% and 10% increase in predictive error, respectively. This translated to a $590 increase in hypothetical reimbursements for moderate-risk patients and a $1249 increase for severe-risk patients in a patient-specific risk-adjusted payment model (Table IV).

A local model-agnostic explanation figure was used to demonstrate the final model's individual patient-specific predictors. The example patient was created using the most important variables in the study, resulting in a predicted total charge after ambulatory RCR of $15,317.3 (Fig. 5). OR time and number of anchors contributed strongly to these total charges, with additional minor contributions from ethnicity, concomitant DCE, and utilization of peripheral nerve block.

The model was integrated into a web-based open-access application capable of providing individual predictions and explanations on a case-by-case basis (https://sportsmed.shinyapps.io/RCR_cost/)

Discussion

The principal findings of our investigation are as follows. First, the ensemble ML algorithm outperformed a trained linear regression model as well as the other candidate ML

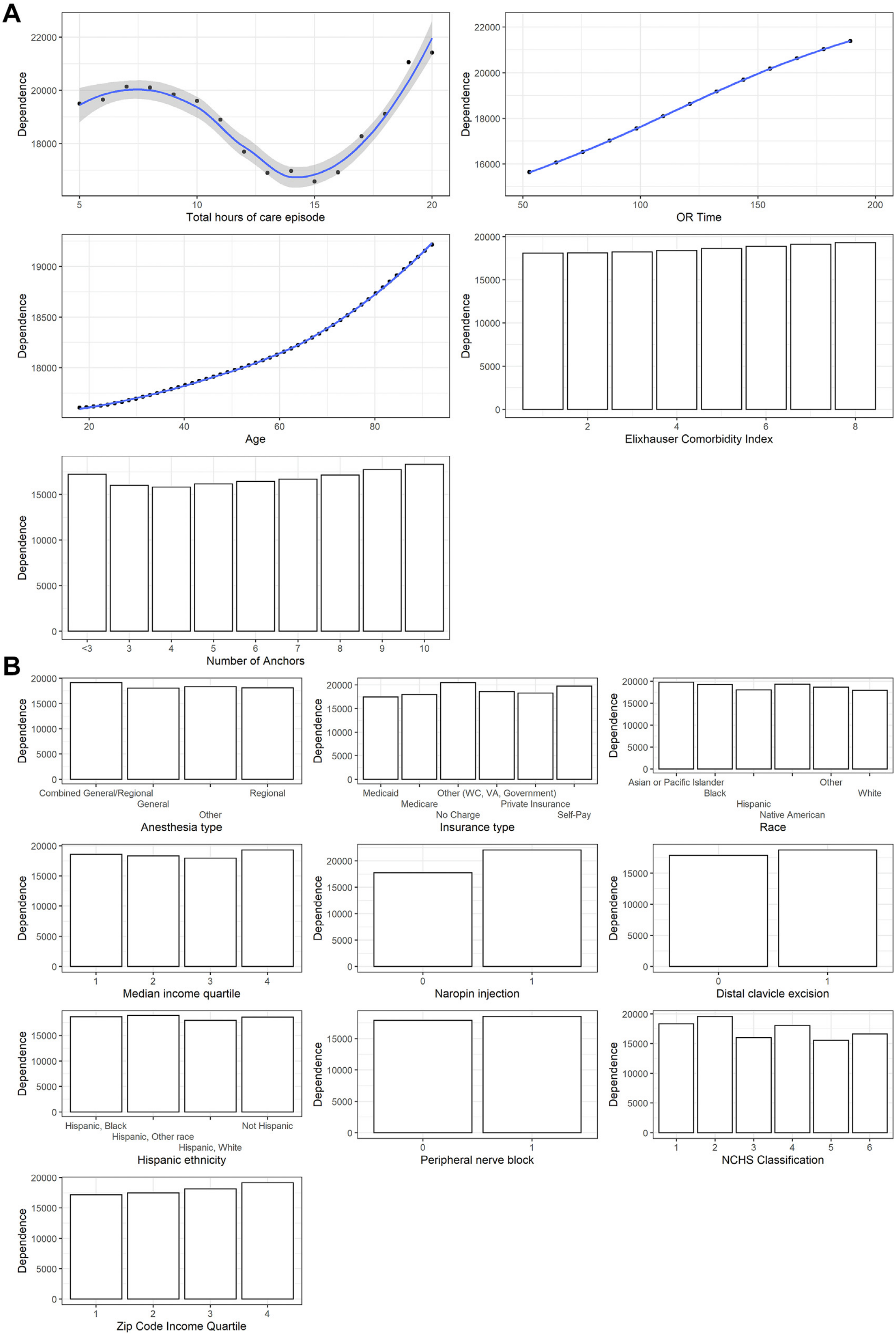

Figure 4 (**A**) Partial dependence plots of continuous input features on total charges after ambulatory rotator cuff repair. There is a sinusoidal dependence of cost on total hours of care episode and operating room time. Cost linearly increases with age, Elixhauser Comorbidity Index, and number of anchors. (**B**) Partial dependence plots of input features on total charges after hip arthroscopy. National Center for Health Statistics classification: 1, "Central" counties of metro areas of $\geq$1 million population; 2, "Fringe" counties of metro areas of $\geq$1 million population; 3, Counties in metro areas of 250,000-999,999 population; 4, Counties in metro areas of 50,000-249,999 population; 5, Micropolitan counties; 6, Not metropolitan or micropolitan counties. *WC*, Worker's Compensation; *VA*, Veteran's Administration.

Table IV Risk-adjusted patient-specific payment model using model's predictive error based on Elixhauser Comorbidity Index

AHRQ comorbidity index	Risk%	Mean payment	Risk-adjusted payment
Mild	100	$18,589	$18,589
Moderate	102	$18,370	$18,960
Severe	110	$19,199	$20,448

AHRQ, Agency for Healthcare Research and Quality.

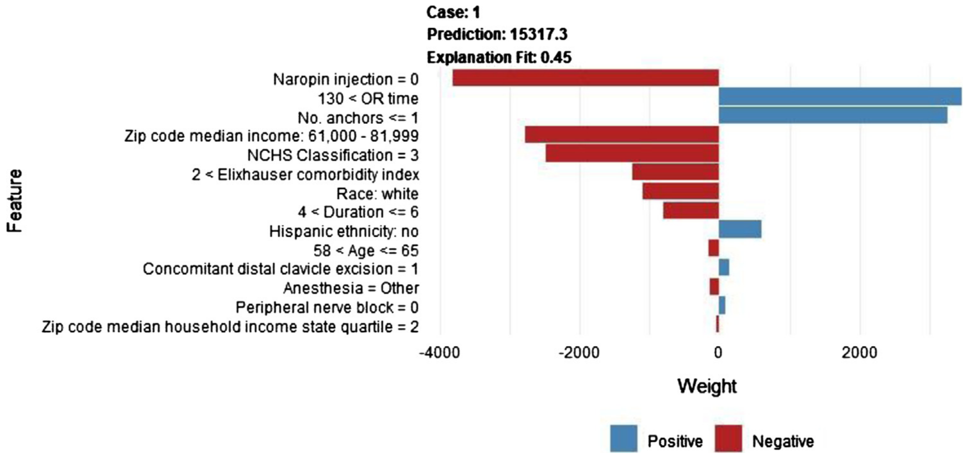

Figure 5 .Example of individual patient-level explanation for predictions made by the gradient boosted ensemble model. This patient had predicted total charges of 15,317.3, with operating room (OR) time and number of anchors contributing strongly to charges. There are minor contributions from ethnicity, concomitant distal clavicle excision, and peripheral nerve blocks as well.

models in minimizing prediction error while maximizing fit. Second, the most significant modifiable drivers of increased total charges included the following modifiable variables: the use of local infiltration anesthesia and peripheral nerve blocks; other nonmodifiable variables included increased OR time and duration of care episode, increased number of anchors used, race, socioeconomic status, and ECI, and concomitant DCE. Third, a hypothetical risk-adjusted payment schedule was established based on the weighted AHRQ ECI for mortality to more appropriately model reimbursements in accordance with relative severity of patient comorbidities.

Several investigations using total charges as a surrogate in the measurement of costs have been performed, which have confirmed the nonmodifiable features identified in the present study as predictors of increased costs. These expectedly included operative time and duration of care, both of which increase the demand and utilization of staff/facility, socioeconomic strata as demonstrated by income and residential characteristics, and overlapping racial/ethnic status, implant fees as indicated by anchor counts, and concomitant DCE.[22,25,29] The ML model further expanded on these findings by modeling the contributions of each continuous variable through the range of their values, demonstrating a sinusoidal relationship between duration of care and cost, with a minimum at approximately 15 hours of stay. Not surprisingly, total stay beyond 15 hours produced increased charges; however, shorter stay times also yielded increases in total charges, possibly secondary to adverse events arising from inadequate post-anesthesia recovery time (pain crisis, nausea, falls, etc.).[14,20] However, we were unable to corroborate this because of the lack of information regarding postoperative complications and emergency department visits. In addition, age and OR time were found to exert a linear relationship with cost. Although race disparities in total charges were present, these findings should not be considered during administrative decision-making for cost minimization, but rather motivate thoughtful analysis to identify the structural drivers for these differences to improve health care inequalities.

The most significant modifiable feature that contributed to total charges was anesthesia type, and this was identified using both a trained linear regression model and the final ensemble model. The linear regression model observed that the performance of a peripheral nerve block and local infiltration with ropivacaine lead to an increase in total charges, a finding corroborated on partial dependence curves. In addition, combined regional and general anesthesia was found to be the most expensive when examining anesthesia type overall. These findings represent considerations during a multidisciplinary, patient-centered, decision-making process; surgeon and anesthesia should outline options and opportunities for cost-saving while optimizing postoperative analgesia and minimizing postanesthesia care unit length of stay and readmission rates. It is well recognized that adequate pain control after ambulatory surgery is directly correlated to reducing readmission rates postoperatively.[40,41] Thus, although ML successfully identified their contribution to the cost of care during outpatient surgery, the nature of their effect on optimizing postoperative analgesia for patients who return home on the same day of surgery cannot be overstated. Although several studies have suggested a negligible improvement in postoperative pain with subacromial ropivacaine or liposomal bupivacaine after rotator cuff surgery when compared with peripheral nerve blocks,[6,36,38] potential future directions may investigate whether the analgesic efficacy of combining nerve block with intra-articular injection is truly additive and identifying patients who may be safely discharged with a nerve block alone. As always, care must be taken to interpret the results of ML models in appropriate clinical context.

An examination of the variable rankings from the global importance plot observed OR time and duration of care to exert the greatest influence on predictive performance, which is unsurprising given that cost likely increased in a predictable proportion with continuous variables. This was likewise true for number of anchors, an integer variable, consistent with previous observations.[22] However, multinomial categorical variables had less predictable contributions to performance. Notably, the performance of a DCE was identified as the sole concomitant procedure to drive total charges, reinforcing the findings of Sabesan et al[33] that concomitant DCE and subacromial decompression correlated with the greatest increase in profit margin when performed with ARCR.

Given the increasing prevalence of episodic payment bundles accepted for ambulatory procedures with well-described postoperative course and low complication profiles, such as ARCR, there has been longstanding interest in accurate cost prediction in order to provide appropriate reimbursements based on patient-specific comorbidities.[10] This is especially true given the highly diverse demographic of patients undergoing ARCR. The present study produced a patient-specific payment model with hypothetical reimbursements based on risk as indicated by model prediction variance. The patient-specific payment model risk-stratified patients based on AHRQ-weighted ECI for the likelihood of mortality and highlighted the significant influence patient comorbidities can have on adjustment of reimbursements, as well as and the potential financial burden assumed by surgeons who electively treat patients with greater comorbidities under a fixed-rate bundle. Declining reimbursements for sports medicine procedures such as ARCR, especially within an aging Medicare population of growing complexity, is a forgone conclusion,[24] and this phenomenon more than ever behooves both physicians and institutions to advocate for appropriate mitigation of the financial risks involved in undertaking complex cases.

Although the mathematical foundations of ML have been established for almost half a century, advancements in modern computing capabilities have ushered in an ‘‘artificial intelligence spring’’ that has seen universal interest in, and application of, learned algorithms. Previous investigations in surgical oncology and neurosurgery have modeled cost as a supervised ML regression problem.[26,39] Yet most cost-modeling studies in orthopedics either use naïve linear models or model cost through discretization,[19] a process subject to potential loss of information and introduction of bias. The algorithm produced herein can generate predictions superior to naïve linear regressions based on all the performance metrics considered, and enhancement of global and local model interpretability with partial dependence curves and model-agnostic explanations, respectively, lends it additional merit. This is unsurprising, as total charges after surgery are derived from a complex interplay of implants and services provided, insurance status, and patient comorbidities. Each input may have varying contributions to the charges incurred by each patient, and linear models, due inherent constraint, will have difficulty accounting for individual nuances. We have included a digital application in adherence to the concept of seamless integration in the introduction of new biomedical technology.[31] This application is for education purposes and has been linked to demonstrate a proof-of-concept user interface accessible from the provider side of the electronic medical record. It is important to note, however, that the inclusion of this application is not an endorsement of its external validity. Indeed, although algorithms can be internally validated based on development data, wide adoption requires assessment of model performance both on distinct geographical and temporal cohorts. Several of those incorporated into the ensemble are prone to overfit on training data; despite mitigation during modeling, this tool will require rigorous external validation on both temporally and geogrpahically distinct populations to ensure safe use, patient privacy, and explainable insights.[30]

Limitations

This study is not without its limitations. Although its primary strength is both its efficiency and accuracy in surveilling vast sums of data from multiple contributing centers, implementing ML algorithms to address the questions put forth by this study was limited to the data provided by the NYSASD and requires further external validation, especially given propensity of certain algorithms, such as random forest, to overfit on the training data or disproportionately fit certain confounders. Given the size of the database and the use of standardized data, it is not possible for the authors to individually vet and validate each entered subject, and despite quality controls by the HCUP, there may remain data points affected human coding errors. However, the ability to analyze thousands of patients across multiple contributing centers while performing internal validation of our data is a strength. Second, the lack of granular data in the NYSASD limited our investigation of additional modifiable risk factors to those presented in this study. Thus, future investigation into additional cost drivers, such as preoperative functional status, level of assistance at home, and distance required to access services and care, is warranted. Third, it should be noted that several nonmodifiable variables identified to drive total charges, such as race, ethnicity, and socioeconomic strata as reflected by level of income, insurance, and residential classifications, should not be used to drive surgical decision-making by administrators in such a way that does not ensure equitable provision of care. The possible disparities highlighted herein should, however, be the subject of root cause analysis. Finally, because ML inherently prioritizes predictive performance, the model may include certain variables such as race and residential area characteristics that can capture overlapping information; however, we attempted to mitigate this through the exclusion of highly collinear variables in the feature selection process.

Conclusion

This study developed an ensemble supervised ML algorithm that outperformed standard linear regression in predicting total charges cost after ambulatory RCR. Important contributors to total charges included OR time, duration of care, number of anchors used, anesthesia, distal clavicle excision, and community and residency characteristics. Generation of a patient-specific payment schedule based on AHRQ risk of mortality highlighted the financial risk assumed by physicians in flat episodic reimbursement schedules given variable patient comorbidities and highlighted the importance of an accurate prediction algorithm to appropriately reward high-value care at low costs.

Disclaimers:

Funding: The authors would like to acknowledge the support from the Foderaro-Quattrone Musculoskeletal-Orthopaedic Surgery Research Innovation Fund.

Conflicts of interest: Jonathan D. Barlow receives IP royalties from and is a paid consultant for Stryker. Joaquin Sanchez-Sotelo is a board or committee member of American Shoulder and Elbow Surgeons; receives publishing royalties and financial or material support from Elsevier and Oxford University Press; is a paid consultant for Exactech, Inc. and Precision OS; is on the editorial or governing board of and receives publishing royalties and financial or material support from the Journal of Shoulder and Elbow Surgery; has stock or stock options in PSI; receives IP royalties and research support from and is a paid presenter and speaker for Stryker; and receives IP royalties from and is a paid consultant and paid presenter or speaker for Wright Medical Technology, Inc. Christopher L. Camp received hospitality payments and educational support from Arthrex and hospitality payments (2015) from Zimmer Biomet. The other authors, their immediate families, and any research foundation with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jse.2022.04.008.

References

1. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt 2014;34:502-8. https://doi.org/10.1111/opo.12131
2. Austin DC, Torchia MT, Lurie JD, Jevsevar DS, Bell J-E. Identifying regional characteristics influencing variation in the utilization of rotator cuff repair in the United States. J Shoulder Elbow Surg 2019;28:1568-77. https://doi.org/10.1016/j.jse.2018.12.013
3. Bayliss L, Jones LD. The role of artificial intelligence and machine learning in predicting orthopaedic outcomes. Bone Joint J 2019;101-B:1476-8. https://doi.org/10.1302/0301-620x.101b12.Bjj-2019-0850.R1
4. Bodenheimer T, Fernandez A. High and rising health care costs. Part 4: can costs be controlled while preserving quality? Ann Intern Med 2005;143:26. https://doi.org/10.7326/0003-4819-143-1-200507050-00007
5. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 2021;7:e623. https://doi.org/10.7717/peerj-cs.623
6. Coghlan JA, Forbes A, McKenzie D, Bell SN, Buchbinder R. Efficacy of subacromial ropivacaine infusion for rotator cuff surgery. A randomized trial. J Bone Joint Surg Am 2009;91:1558-67. https://doi.org/10.2106/jbjs.H.00948
7. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Br J Surg 2015;102:148-58. https://doi.org/10.1002/bjs.9736

8. Colvin AC, Egorova N, Harrison AK, Moskowitz A, Flatow EL. National trends in rotator cuff repair. J Bone Joint Surg Am 2012;94:227-33. https://doi.org/10.2106/JBJS.J.00739
9. Crawford DC, Li CS, Sprague S, Bhandari M. Clinical and cost implications of inpatient versus outpatient orthopedic surgeries: a systematic review of the published literature. Orthop Rev (Pavia) 2015;7:6177. https://doi.org/10.4081/or.2015.6177
10. Cutler DM, Ghosh K. The potential for cost savings through bundled episode payments. N Engl J Med 2012;366:1075-7. https://doi.org/10.1056/NEJMp1113361
11. Darst BF, Malecki KC, Engelman CD. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC Genet 2018;19(Suppl 1):65. https://doi.org/10.1186/s12863-018-0633-8
12. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998;36:8-27.
13. Fraser I, Encinosa W, Glied S. Improving efficiency and value in health care: introduction. Health Serv Res 2008;43(Pt 2):1781-6. https://doi.org/10.1111/j.1475-6773.2008.00904.x
14. Ganter MT, Blumenthal S, Dübendorfer S, Brunnschweiler S, Hofer T, Klaghofer R, et al. The length of stay in the post-anaesthesia care unit correlates with pain intensity, nausea and vomiting on arrival. Perioper Med (Lond) 2014;3:10. https://doi.org/10.1186/s13741-014-0010-8
15. Goldman AH, Kates S. Pay-for-performance in orthopedics: how we got here and where we are going. Curr Rev Musculoskelet Med 2017;10:212-7. https://doi.org/10.1007/s12178-017-9404-9
16. 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; 2014.
17. Jensen AR, Cha PS, Devana SK, Ishmael C, Di Pauli von Treuheim T, D'Oro A, et al. Evaluation of the trends, concomitant procedures, and complications with open and arthroscopic rotator cuff repairs in the Medicare population. Orthop J Sports Med 2017;5. https://doi.org/10.1177/2325967117731310. 2325967117731310.
18. Kadhim M, Gans I, Baldwin K, Flynn J, Ganley T. Do surgical times and efficiency differ between inpatient and ambulatory surgery centers that are both hospital owned? J Pediatr Orthop 2016;36:423-8. https://doi.org/10.1097/bpo.0000000000000454
19. Karnuta JM, Churchill JL, Haeberle HS, Nwachukwu BU, Taylor SA, Ricchetti ET, et al. The value of artificial neural networks for predicting length of stay, discharge disposition, and inpatient costs after anatomic and reverse shoulder arthroplasty. J Shoulder Elbow Surg 2020;29:2385-94. https://doi.org/10.1016/j.jse.2020.04.009
20. Knutsen Glette M, Kringeland T, Røise O, Wiig S. Hospital physicians' views on discharge and readmission processes: a qualitative study from Norway. BMJ Open 2019;9:e031297. https://doi.org/10.1136/bmjopen-2019-031297
21. Legates DR, McCabe GJ Jr. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour Res 1999;35:233-41.
22. Li L, Bokshan SL, Ready LV, Owens BD. The primary cost drivers of arthroscopic rotator cuff repair surgery: a cost-minimization analysis of 40,618 cases. J Shoulder Elbow Surg 2019;28:1977-82. https://doi.org/10.1016/j.jse.2019.03.004
23. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016;18:e323. https://doi.org/10.2196/jmir.5870
24. Malik AT, Kopechek KJ, Bishop JY, Cvetanovich GL, Khan SN, Neviaser AS. Declining trends in Medicare physician reimbursements for shoulder surgery from 2002 to 2018. J Shoulder Elbow Surg 2020;29:e451-61. https://doi.org/10.1016/j.jse.2020.02.005
25. Morris JH, Malik AT, Hatef S, Neviaser AS, Bishop JY, Cvetanovich GL. Cost of arthroscopic rotator cuff repairs is primarily driven by procedure-level factors: a single-institution analysis of an ambulatory surgery center. Arthroscopy 2021;37:1075-83. https://doi.org/10.1016/j.arthro.2020.11.033
26. 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;131:507-16. https://doi.org/10.3171/2018.4.Jns18306
27. Myers TG, Ramkumar PN, Ricciardi BF, Urish KL, Kipper J, Ketonis C. Artificial intelligence and orthopaedics. J Bone Joint Surg Am 2020;102:830-40. https://doi.org/10.2106/jbjs.19.01128
28. Myers TG, Ramkumar PN, Ricciardi BF, Urish KL, Kipper J, Ketonis C. Artificial intelligence and orthopaedics: an introduction for clinicians. J Bone Joint Surg Am 2020;102:830-40. https://doi.org/10.2106/JBJS.19.01128
29. Narvy SJ, Ahluwalia A, Vangsness CT Jr. Analysis of direct costs of outpatient arthroscopic rotator cuff repair. Am J Orthop (Belle Mead NJ) 2016;45:E7-11.
30. Oosterhoff JHF, Thio QCBS, Groot OQ, Bongers MER, Ghaednia H, Karhade AV, et al. Integration of automated predictive analytics into electronic health records: can spine surgery applications lead the way using SMART on FHIR and CDS Hooks? Semin Spine Surg 2021;33:100870. https://doi.org/10.1016/j.semss.2021.100870
31. PCMag. Seamless integration. 2021. Available at: https://www.pcmag.com/encyclopedia/term/seamless-integration. Accessed April 7, 2021
32. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019;380:1347-58. https://doi.org/10.1056/NEJMra1814259
33. Sabesan VJ, Shahriar R, Chatha K, Malone DL, Sherwood A, Peaguda CF, et al. Factors affecting the cost and profitability of arthroscopic rotator cuff repair. Arthroscopy 2019;35:38-42. https://doi.org/10.1016/j.arthro.2018.07.034
34. Stekhoven DJ, Bühlmann P. MissForest–non-parametric missing value imputation for mixed-type data. Bioinformatics 2012;28:112-8. https://doi.org/10.1093/bioinformatics/btr597
35. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 2013;10:e1001381. https://doi.org/10.1371/journal.pmed.1001381
36. Toyooka S, Ito M, Kakinuma A, Kayama S, Watanabe K, Miyamoto W, et al. Periarticular multimodal drug injection does not improves early postoperative analgesia compared with continuous interscalene brachial plexus block after arthroscopic rotator cuff repair: a retrospective single-center comparative study. J Orthop Sci 2020;25:405-9. https://doi.org/10.1016/j.jos.2019.04.013
37. U.S. Bureau of Labor Statistics. Consumer price index archived news releases. 2021. Available at: https://www.bls.gov/bls/news-release/cpi.htm#2015. Accessed January 20, 2021
38. Verdecchia NM, Rodosky MW, Kentor M, Orebaugh SL. Liposomal bupivacaine infiltration in the surgical site for analgesia after rotator cuff repair: a randomized, double-blinded, placebo-controlled trial. J Shoulder Elbow Surg 2021;30:986-93. https://doi.org/10.1016/j.jse.2020.10.035
39. 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 2009;9:161. https://doi.org/10.1186/1472-6963-9-161
40. YaDeau JT, Soffin EM, Tseng A, Zhong H, Dines DM, Dines JS, et al. A comprehensive enhanced recovery pathway for rotator cuff surgery reduces pain, opioid use, and side effects. Clin Orthop Relat Res 2021;479:1740-51. https://doi.org/10.1097/corr.0000000000001684
41. Zhu S, Qian W, Jiang C, Ye C, Chen X. Enhanced recovery after surgery for hip and knee arthroplasty: a systematic review and meta-analysis. Postgrad Med J 2017;93:736-42. https://doi.org/10.1136/postgradmedj-2017-134991