J Shoulder Elbow Surg (2020) 29, 2385–2394

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

The value of artificial neural networks for predicting length of stay, discharge disposition, and inpatient costs after anatomic and reverse shoulder arthroplasty

Jaret M. Karnuta, MS[a,b], Jessica L. Churchill, MD[a,b], Heather S. Haeberle, MD[b,c], Benedict U. Nwachukwu, MD, MBA[b,c], Samuel A. Taylor, MD[b,c], Eric T. Ricchetti, MD[a,b], Prem N. Ramkumar, MD, MBA[a,b,*]

[a]*Department of Orthopedic Surgery, Cleveland Clinic, Cleveland, OH, USA*
[b]*Machine Learning Arthroplasty Lab, Cleveland Clinic, Cleveland, OH, USA*
[c]*Sports & Shoulder Service, Hospital for Special Surgery, New York, NY, USA*

Hypothesis/Purpose: The objective is to develop and validate an artificial intelligence model, specifically an artificial neural network (ANN), to predict length of stay (LOS), discharge disposition, and inpatient charges for primary anatomic total (aTSA), reverse total (rTSA), and hemi- (HSA) shoulder arthroplasty to establish internal validity in predicting patient-specific value metrics.
Methods: Using data from the National Inpatient Sample between 2003 and 2014, 4 different ANN models to predict LOS, discharge disposition, and inpatient costs using 39 preoperative variables were developed based on diagnosis and arthroplasty type: primary chronic/degenerative aTSA, primary chronic/degenerative rTSA, primary traumatic/acute rTSA, and primary acute/traumatic HSA. Models were also combined into diagnosis type only. Outcome metrics included accuracy and area under the curve (AUC) for a receiver operating characteristic curve.
Results: A total of 111,147 patients undergoing primary shoulder replacement were included. The machine learning algorithm predicting the overall chronic/degenerative conditions model (aTSA, rTSA) achieved accuracies of 76.5%, 91.8%, and 73.1% for total cost, LOS, and disposition, respectively; AUCs were 0.75, 0.89, and 0.77 for total cost, LOS, and disposition, respectively. The overall acute/traumatic conditions model (rTSA, HSA) had accuracies of 70.3%, 79.1%, and 72.0% and AUCs of 0.72, 0.78, and 0.79 for total cost, LOS, and discharge disposition, respectively.
Conclusion: Our ANN demonstrated fair to good accuracy and reliability for predicting inpatient cost, LOS, and discharge disposition in shoulder arthroplasty for both chronic/degenerative and acute/traumatic conditions. Machine learning has the potential to preoperatively predict costs, LOS, and disposition using patient-specific data for expectation management between health care providers, patients, and payers.
Level of evidence: Basic Science Study; Computer Modeling

Keywords: Machine learning; shoulder arthroplasty; artificial neural network; artificial intelligence; length of stay; cost; discharge; outcomes

Investigation performed at the Cleveland Clinic, Cleveland, OH.
Institutional review board approval was not required for this basic science study.

*Reprint requests: Prem N. Ramkumar, MD, MBA, Department of Orthopedic Surgery, Cleveland Clinic, 9500 Euclid Ave, Cleveland, OH, USA.
E-mail address: premramkumar@gmail.com (P.N. Ramkumar).

1058-2746/$ - see front matter © 2020 Journal of Shoulder and Elbow Surgery Board of Trustees. All rights reserved.
https://doi.org/10.1016/j.jse.2020.04.009

Shoulder arthroplasty is a common procedure used to treat glenohumeral arthritis and preserve shoulder function after trauma, such as displaced proximal humerus fractures. Both anatomic and reverse total shoulder arthroplasty (aTSA and rTSA) are widely considered safe and effective procedures with low overall morbidity.[6] However, not all patients are at equal risk for complications, and several patient factors have been retrospectively identified that increase the complication rate including advanced age, medical comorbidities, and longer operative time.[6,11,14,17] A predictive model that preoperatively risk stratifies and can quantitatively calculate the increased risk specific to a patient before shoulder arthroplasty would be valuable for all health care stakeholders, specifically in terms of expectation management and insurance preauthorization arbitration to optimally allocate resources and maximize the value of care delivered.[6]

With the current availability of large patient data sets, machine learning (ML) represents a form of artificial intelligence particularly suited for preoperative medical risk stratification and resource allocation. ML may be used to learn and improve from experience with complicated data and nonlinear relationships, and the algorithm becomes more accurate and predictive as additional data sets are presented.[5,22,20] This ability to ''learn'' from complex data relationships separates ML from the more familiar logistic regression analysis and other purely statistical techniques commonly used in previous studies.[3] Unlike ML, regression analysis is static and often relies on predefined relationships (eg, linear), making it less robust for large data sets with complex relationships between patient characteristics and outcomes. Recently, ML has been applied in orthopedics to stratify patient risk before surgery and predict likely hospital costs based on preoperative patient characteristics.[5,6,11,15,22,23] However, limited work has been performed to establish internal validity for value metric in shoulder arthroplasty.

Artificial neural network (ANN) represents a form of ML that ''learns'' through experience rather than static statistical regression.[1,2,12] The ability to ''learn'' makes these models useful for predicting data with nonlinear relationships, which are more frequently encountered in medicine and thus better captured by these algorithms. By using ANN models, providers can input patient metrics preoperatively and predict important postoperative outcomes based on the model's experience with previous patient data. This is particularly useful for the creation of a risk-based, patient-specific payment model to ensure that surgeons are appropriately compensated for the complexity of completed cases. In previous literature, ML has successfully predicted length of stay (LOS), inpatient charges, and discharge disposition after lower extremity joint replacement.[7,21-23] The objective of the current study is to determine if a neural network can be used to predict LOS, discharge disposition, and inpatient charges for primary chronic/degenerative aTSA, primary chronic/degenerative rTSA, primary acute/traumatic rTSA, and primary acute/traumatic hemiarthroplasty (HSA) based on admission diagnosis category (chronic/degenerative vs. acute/traumatic) and arthroplasty type to establish validity in predicting patient-specific value metrics. We hypothesize that an ML ANN model will have the capability to accurately preoperatively predict value metrics for these 4 conditions.

Methods

Data source

This is a retrospective cohort study of TSA using the National Inpatient Sample (NIS) between 2003 and 2014. Published since 1988 and updated yearly, this survey includes demographic, clinical, and resource use data from more than 1000 short-term and nonfederal hospitals, making it the largest publicly available all-payer inpatient discharge database in the United States. The patients included in this database form a heterogeneous, deidentified cohort with varying insurance payers, socioeconomic backgrounds, and geographical locations. All data for each patient stay are entered into NIS with associated diagnosis and procedural International Classification of Diseases, ninth revision (ICD-9) codes.

Data processing

We considered all patients with discharges containing one of the following ICD-9 procedure codes: 8188 (''Reverse total shoulder replacement'' or rTSA), 8181 (''Partial shoulder replacement'' or HSA), and 8180 (''Other total shoulder replacement'' or aTSA). This inclusion criterion resulted in a total of 146,618 patients. This NIS subset was further filtered and processed to arrive at a robust data set for ML, and empty or unknown values for disposition, LOS, sex, age, insurance payer, hospital location, hospital bed size, hospital ownership, elective procedure status, and zip code income quartile were removed. Similarly, patients under the age of 18 were excluded in order to maximize the generalizability of our results. In addition, all revision arthroplasties were excluded (revision arthroplasty exclusion ICD-9 codes are included in Table I). After processing of the data, 111,147 patients remained. This finalized cohort contained 57,069 aTSA, 21,457 rTSAs, and 32,641 HSAs. In order to elucidate any association between preoperative diagnosis with LOS, inpatient charges, and discharge disposition after shoulder arthroplasty, we subdivided this large cohort into patients with chronic/degenerative and acute/traumatic conditions. We divided these patients based on ICD-9 diagnosis codes or external cause of injury Clinical Classifications Software codes corresponding to each type of condition (chronic/degenerative vs. acute/traumatic). These are included in Table I. Patients who were coded as both acute/traumatic and chronic/nontraumatic were classified as acute/traumatic.

To further assess if an ML model could accurately predict inpatient costs for shoulder arthroplasty, we normalized patient costs and charges to 2014 US dollars using the consumer price index supplied by the United States Bureau of Labor Statistics.[24] Patients with LOS, total cost, or total charges greater than the 99th percentile or less than the 1st percentile were

Table I ICD-9 and CCS codes corresponding to each included pathology for chronic/degenerative and acute/traumatic indications for shoulder arthroplasty

Chronic/degenerative			Acute/traumatic		
Code type	Code	Description	Code type	Code	Description
ICD-9 Dx	714	Arthritis	ICD-9 Dx	812[0-3]	Superior humeral fractures
ICD-9 Dx	715	Arthritis	ICD-9 Dx	831[0-1]	Shoulder dislocations
ICD-9 Dx	716	Arthritis and arthropathy	ICD-9 Dx	73311	Pathologic fracture of humerus
ICD-9 Dx	711[1-8]	Arthropathy	ICD-9 Dx	73310	Pathologic fracture unspecified
ICD-9 Dx	6960	Psoriatic arthropathy	ICD-9 Dx	73319	Pathologic fracture of other specified site
ICD-9 Dx	73340	Aseptic necrosis, NOS	CCS E code	2603	Fall
ICD-9 Dx	73341	Aseptic necrosis of head of humerus	CCS E code	2606	Machinery
ICD-9 Dx	73349	Aseptic necrosis of bone, other	CCS E code	2607	Motor vehicle traffic injury
ICD-9 Dx	73381	Malunion of fracture	CCS E code	2608	Pedal cyclist not MVT
ICD-9 Dx	71831	Recurrent dislocation of joint	CCS E code	2609	Pedestrian, not MVT
ICD-9 Dx	73382	Nonunion of fracture	CCS E code	2610	Transport, not MVT
ICD-9 Dx	9052	Late effect of fracture of upper extremities	CCS E code	2614	Struck by object
ICD-9 Dx	71880	Other joint derangement, not elsewhere classified, site unspecified	CCS E code	2619	Other specified injury
ICD-9 Dx	71881	Other joint derangement, not elsewhere classified, shoulder region	CCS E code	2620	Unspecified injury
ICD-9 Dx	72610	Disorders of bursae and tendons in shoulder region, unspecified			
ICD-9 Dx	72761	Complete rupture of rotator cuff			
ICD-9 Dx	8407	Superior glenoid labrum lesion			

ICD-9, International Classification of Diseases, ninth revision; *CCS*, Clinical Classifications Software; *NOS*, not otherwise specified; *MVT*, motor vehicle traffic.

excluded to control for outliers. Similarly, patient disposition was coded as either routine or nonroutine, where routine (as defined by the Agency for Healthcare Research and Quality) is discharge home, discharge to law enforcement, or discharge home with a planned acute care hospital inpatient encounter and nonroutine is all other discharge destinations.[25] For convenience, continuous outcome variables were split into 3 bins (low, medium, and high) based on z-score normalization. A z-score is a statistical normalization technique that subtracts the mean value of a distribution from each inputted value and divides by the standard deviation. The resulting distribution is centered at 0 and has a standard deviation of 1. The use of z-score normalizations allows the model to place equal weight on inputted variables, regardless of the units of the inputted variables. The low bin was z-score ≤ -1, the medium bin was z-score between -1 and 1 exclusive, and the high bin was z-score ≥ 1. Overall, models were created for chronic/degenerative and acute/traumatic indications. Models were further divided into specific clinical indication and arthroplasty type, defined as follows: primary chronic/degenerative aTSA, primary chronic/degenerative rTSA, primary acute/traumatic rTSA, and primary acute/traumatic HSA. Models within a clinical indication were combined (eg, combining the models for aTSA and rTSA within chronic/degenerative) for ease of reporting. Patients used to train combined models still included a variable indicating which type of arthroplasty they received.

Neural network development

A custom Python script (Python Software Foundation, Beaverton, OR, USA) was developed using an open-source neural network toolbox.[19] Variables included in modeling and model parameters can be found in the example model definition file supplied in this project's codebase (https://github.com/JaretK/ShoulderArthroplastyDeepLearning). Briefly, continuous preoperative variables (age) were normalized to their z-score representations before modeling, whereas admission diagnoses, projected procedures, and external cause of injury codes were converted into integers representing each diagnosis before being passed into the network. Year of surgery was passed in without transformation. Categorical and binary variables were passed directly into the network. Zip code, income quartile, All Patients Refined Diagnosis Related Groups risk of mortality, and All Patients Refined Diagnosis Related Groups severity of illness were imputed with the mean value, whereas missing values for elective surgery and transfer status were coded as nonelective and not transferred, respectively. We split the overall data set into separate training, validation, and testing data sets. A total of 70% of the data were used to train the model, 10% to validate the model parameters, and 20% to test the model. This process was repeated 10 times to establish measures of spread for the results. An overview of the algorithm is outlined in Figure 1.

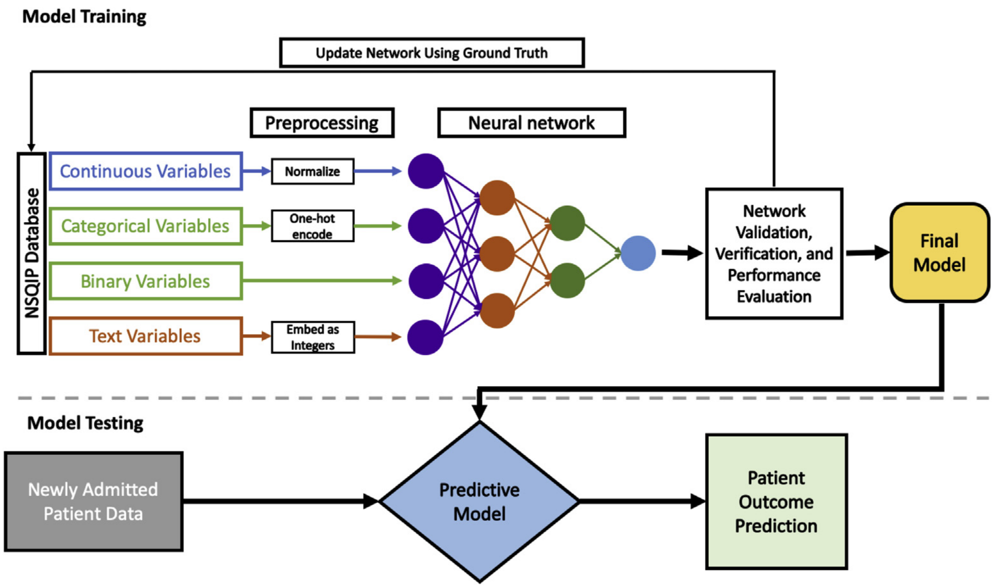

Figure 1 Diagram depicting the pathway of data processing and outcome prediction using machine learning modeling. This study has used this process for final model development, although external testing and outcome prediction remains as a future step.

Subgroup analysis

A subgroup analysis was performed by creating separate models trained on data collected from patients between 2008 and 2014. The accuracy for each model was determined and compared with the accuracy for the 2003-2014 model to ensure that the model is not biased by older data. Comparison was performed using the Mann-Whitney *U*-test.

Statistical analysis

The ML algorithm was validated in terms of accuracy and responsiveness. Responsiveness was established via the receiver operating characteristics (ROC) curve, the area under the curve (AUC) of which was estimated using standard mathematical functions. Accuracy is defined as the percent of correct predictions made during the testing phase of the model. Responsiveness in this paper is defined as the ability of a model to successfully distinguish different outcomes. Responsiveness was graded as excellent (0.9-1.0), good (0.8-0.9), fair (0.7-0.8), poor (0.6-0.7), and fail (0.5-0.6).[13] All statistical analyses were performed using R version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria). A significance threshold of $P < .05$ was chosen for all analyses.

Results

Patient demographics are shown in Table II, and patient comorbidities, current or historical, are shown in Table III.

Chronic/degenerative conditions

A total of 73,162 patients were included in the chronic/degenerative diagnosis cohort (aTSA, rTSA). Of these, 18,051 received rTSA and 55,111 received aTSA. The accuracy for total cost for chronic/degenerative conditions treated with aTSA and rTSA was 76.5% with an AUC of 0.75 (Fig. 2). The accuracy for LOS was 91.8% with an AUC of 0.89. Patient disposition had an accuracy of 73.1% and an AUC of 0.77 (Fig. 3). Total cost, LOS, and discharge disposition for aTSA were predicted with AUCs of 0.75, 0.76, and 0.75, respectively. Chronic/degenerative conditions treated with rTSA had an AUC of 0.74 for total cost, 0.85 for LOS, and 0.78 for discharge disposition (Table IV).

Acute/traumatic injury

A total of 17,630 patients were coded as having an acute/traumatic cause for their inpatient admission. Of these, 13,632 received HSA and 3998 received an rTSA. The accuracy for total cost for acute/traumatic conditions was 70.3% with an AUC of 0.72. The accuracy for LOS was 79.1% with an AUC of 0.78. Patient disposition for traumatic causes of injury had an accuracy of 72.0% and an AUC of 0.79. The ROC curves for these 3 models for both acute/traumatic HSA and rTSA are depicted in Figure 2. In examining rTSA only, the neural network predictor had an AUC of 0.71 for total cost, 0.80 for LOS, and 0.79 for discharge disposition. The predictor of total cost, LOS, and discharge disposition for acute/traumatic HSA alone had AUCs of 0.74, 0.77, and 0.74, respectively (Table IV).

Subgroup analysis on 2008-2014 patients

The combined models for 2008-2014 included a total of 78,404 patients. Within these combined groups, a total of 66,544 patients had a chronic/degenerative diagnosis and

Table II Patient demographics for included patient cohort

Demographics	Chronic/degenerative (n = 73,162)	Acute/traumatic (n = 17,630)	Overall (n = 90,792)
Age (SD)	68.4 (10.7)	71.4 (11.5)	69.0 (10.9)
Sex			
Female	40,385 (55.2)	13,610 (77.2)	53,749 (59.2)
Male	32,777 (44.8)	4019 (22.8)	37,043 (40.8)
Admission month	Available	Available	Available
Weekend admission	Available	Available	Available
Insurance payer			
Medicaid	1829 (2.5)	529 (3.0)	2361 (2.6)
Medicare	48,505 (66.3)	12,252 (69.5)	60,740 (66.9)
Other	3000 (4.1)	812 (4.6)	3813 (4.2)
Private	19,606 (26.8)	3720 (21.1)	23,424 (25.8)
Self-pay	292 (0.4)	317 (1.8)	545 (0.6)
Race			
Asian or Pacific Islander	293 (0.4)	123 (0.7)	454 (0.5)
Black	2634 (3.6)	353 (2.0)	2996 (3.3)
Hispanic	1902 (2.6)	652 (3.7)	2542 (2.8)
Native American	220 (0.3)	53 (0.3)	272 (0.3)
Other	1098 (1.5)	300 (1.7)	1453 (1.6)
Unknown	13,973 (19.1)	3297 (18.7)	17,250 (19.0)
White	53,042 (72.5)	12,870 (73.0)	65,915 (72.6)
Year			
Mean (SD)	2010 (3.25)	2010 (3.44)	2010 (3.30)
Median [Min, Max]	2010 [2000, 2010]	2010 [2000, 2010]	2010 [2000, 2010]
Zipcode income quartile			
1	19,225 (21.0)	4583 (23.2)	23,808 (21.4)
2	24,689 (27.0)	5314 (26.9)	30,003 (27.0)
3	24,294 (26.6)	4884 (24.7)	29,178 (26.3)
4	21,588 (23.6)	4630 (23.4)	26,218 (23.6)
Missing	1599 (1.7)	341 (1.7)	1940 (1.7)
Hospital ownership			
Government, nonfederal	7682 (10.5)	2010 (11.4)	9624 (10.6)
Private invest-own	9438 (12.9)	2345 (13.3)	11,803 (13.0)
Private not-profit	56,042 (76.6)	13,293 (75.4)	69,365 (76.4)
Transfer status	Available	Available	Available
Partial TSA			
Yes	Excluded	13,632 (77.3)	13,632 (15.0)
Anatomic TSA			
Yes	55,111 (75.3)	Excluded	55,111 (60.7)
Reverse TSA			
Yes	18,051 (24.7)	3998 (22.7)	22,049 (24.3)
APR DRG risk of mortality			
1	59,847 (81.8)	11,195 (63.5)	71,362 (78.6)
2	11,413 (15.6)	4989 (28.3)	16,251 (17.9)
3	146 (2.0)	1181 (6.7)	2542 (2.8)
4	220 (0.3)	193 (1.1)	363 (0.4)
APR DRG severity of illness			
1	73 (0.1)	141 (0.8)	182 (0.2)
2	69,723 (95.3)	14,879 (84.4)	84,800 (93.4)
3	2927 (4.0)	2327 (13.2)	5175 (5.7)
4	146 (0.2)	211 (1.2)	363 (0.4)
Disposition routine?			
Yes	51,213 (70.0)	7845 (44.5)	59,377 (65.4)
Total charges			
Mean (SD)	52,800 (25,600)	60,600 (31,400)	54,200 (26,900)
Median [Min, Max]	46,800 [14,900, 182,000]	52,200 [14,900, 182,000]	47,700 [14,900, 182,000]

(*continued on next page*)

Table II Patient demographics for included patient cohort (*continued*)

Demographics	Chronic/degenerative (n = 73,162)	Acute/traumatic (n = 17,630)	Overall (n = 90,792)
Total cost			
Mean (SD)	16,300 (6450)	18,500 (8400)	16,700 (6890)
Median [Min, Max]	15,100 [2210, 102,000]	16,800 [2920, 110,000]	15,300 [2210, 110,000]
Missing	2540 (2.8)	838 (4.2)	3378 (3.0)
Length of stay			
Mean (SD)	2.10 (1.15)	3.69 (2.40)	2.39 (1.57)
Median [Min, Max]	2.00 [0.00, 12.0]	3.00 [0.00, 12.0]	2.00 [0.00, 12.0]

SD, standard deviation; *TSA*, total shoulder arthroplasty; *APR DRG*, All Patients Refined Diagnosis Related Group.
Data are presented as n (%), unless otherwise specified.

Table III Patient comorbidities based on clinical history

	Chronic/degenerative (n = 73,162)	Acute/traumatic (n = 17,630)	Overall (n = 90,792)
Myocardial infarction	3073 (4.2)	793 (4.5)	3904 (4.3)
Congestive heart failure	2780 (3.8)	1181 (6.7)	3904 (4.3)
Peripheral vascular disease	1682 (2.3)	494 (2.8)	2179 (2.4)
Cerebrovascular disease	951 (1.3)	476 (2.7)	1362 (1.5)
Dementia	220 (0.3)	335 (1.9)	545 (0.6)
COPD	12,803 (17.5)	3191 (18.1)	15,979 (17.6)
Rheumatoid arthritis	4609 (6.3)	740 (4.2)	5357 (5.9)
Peptic ulcer disease	438 (0.6)	88 (0.5)	545 (0.6)
Controlled diabetes	951 (1.3)	405 (2.3)	1362 (1.5)
Uncontrolled diabetes	12,730 (17.4)	4143 (23.5)	16,797 (18.5)
Cancer	731 (1.0)	388 (2.2)	1090 (1.2)
Liver disease	73 (0.1)	35 (0.2)	91 (0.1)
Metastatic disease	72 (0.1)	193 (1.1)	272 (0.3)
HIV	73 (0.1)	4 (0.0)	91 (0.1)

COPD, chronic obstructive pulmonary disease; *HIV*, human immunodeficiency virus.
Data are presented as n (%).

11,860 had a traumatic diagnosis. A total of 16,882 rTSAs, 40,724 aTSAs, and 20,809 rTSAs were included. For chronic/degenerative conditions, the combined accuracy was 76.3% (95% confidence interval [95% CI]: 76.0%-76.6%) for cost, 90.6% (95% CI: 89.5%-91.7%) for LOS, and 74.1% (95% CI: 73.0%-75.2%) for discharge disposition. All of the accuracies for the 2008-2014 cohort were statistically similar to those of the overall cohort ($P < .001$ for all).

For acute/traumatic conditions, the combined accuracy for total cost was 70.0% (95% CI: 69.8%-70.2%). The combined accuracy for LOS was 79.2% (95% CI: 79.0%-79.4%). The combined accuracy for discharge disposition was 72.3% (95% CI: 72.2%-72.4%). Similarly, all of these accuracies were statistically similar to the overall cohort ($P < .001$ for all).

Discussion

Over the last decade, shoulder arthroplasty has become an increasingly routine procedure that demonstrably improves function and reduces pain for both traumatic and degenerative conditions.[6,8] In addition to the intensifying climate of value-based care after the Bundled Payments for Care Improvement initiative for lower extremity arthroplasty, there has been increased momentum to perform shoulder arthroplasty as an outpatient procedure at ambulatory surgery centers.[4,6,10] Because of these potential reimbursement changes, several studies have attempted to create an accurate tool to predict patient outcomes after shoulder arthroplasty with varying success.[6,11] After using the NIS to compile a cohort of 111,147 patients who underwent shoulder arthroplasty, we applied ML techniques to assess the viability of a model capable of accurately predicting the specific value-based metrics of LOS, inpatient cost, and discharge disposition for primary aTSA, rTSA, and HSA in both chronic/degenerative and acute/traumatic clinical conditions.

This study does not represent the first attempt to apply ML to TSA outcomes. Biron et al[6] examined 4500 patients undergoing elective TSA and created an ML model with an AUC of 0.77 when predicting LOS less than or equal to 1 day in

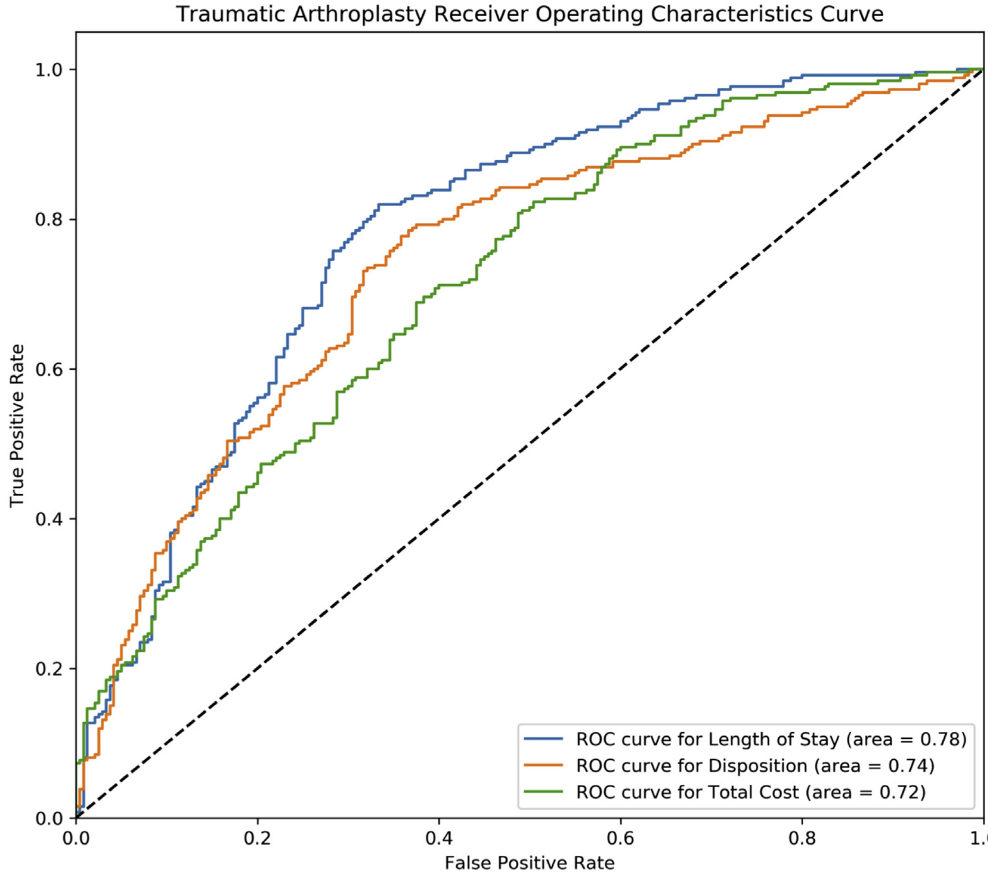

Figure 2 Area under the ROC curve for LOS, discharge disposition, and total cost for the aggregate acute/traumatic injury patient cohort for rTSA and HSA. *ROC*, receiver operating characteristic; *LOS*, length of stay; *rTSA*, reverse total shoulder arthroplasty; *HSA*, hemi-shoulder arthroplasty.

order to determine which patients would be ideal for outpatient TSA. Despite the responsive model, this study failed to delineate between aTSA and rTSA. Similarly, Gowd et al[11] studied 17,119 patients undergoing TSA and created an ML model with an AUC of 0.70 and an accuracy of 82.3% when predicting LOS greater than 3 days; however, this study also did not separately examine aTSA and rTSA. The present study includes a more robust cohort of 90,792 patients than these prior studies. In addition, the strongest predictive ability of the ANN in the current study regarding LOS was good with AUCs of 0.85 and 0.80 for patients undergoing rTSA for chronic/degenerative and acute/traumatic conditions, respectively. In addition to including patient-specific data in the development of the ML model, we examined insurance payer, hospital location, hospital bed size, hospital ownership, elective procedure status, and zip code income quartile. Previous studies using ML models for TSA patients have examined only patient variables.[6,11] With the current economic climate in the wake of the Bundled Payments for Care Improvement initiative, pressure is placed on providers to not only provide a surgical procedure at a low cost, but also discharge patients in an expedient but safe manner. Factors including hospital bed size and geographic location affect these value-based outcomes, making it important to include such variables in a predictive model of these metrics.[9,18] Economic constraints are ever changing; therefore, health care stakeholders need an ML-based predictive tool that can iteratively grow and improve with additional data as they become available.

As the pressure to decrease the cost of care escalates, orthopedic surgeons must adapt and examine their practices for areas of improvement in resource allocation. In order to better allocate resources, health care providers must first understand where resources are being most heavily consumed. As demonstrated in Table IV, our model predicted patients' inpatient costs after TSA with an accuracy ranging from 69% to 77%, consistent with previously published ML models in total hip arthroplasty that have led to the development of risk-adjusted patient-specific payment models.[16,22] Predicting inpatient cost after TSA is an important consideration as physicians explore the possibility of completing these operations in ambulatory surgery centers as outpatient procedures. Similarly, our model predicted patient discharge disposition with an accuracy ranging from 72% to 75% and LOS accuracy ranging from 78% to 92%. The ability to predict discharge disposition and LOS is important to consider when offering patients

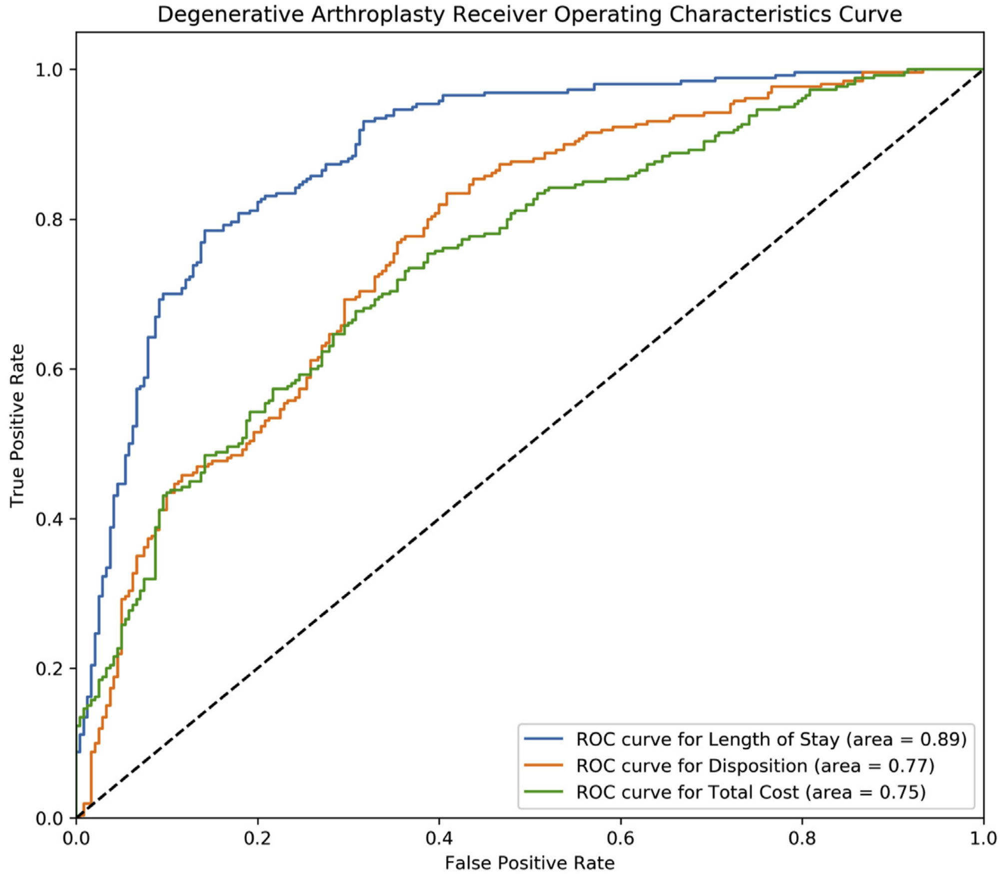

Figure 3 Area under the ROC curve for LOS, discharge disposition, and total cost for the aggregate chronic/degenerative condition patient cohort for aTSA and rTSA. *ROC*, receiver operating characteristic; *LOS*, length of stay; *aTSA*, anatomic total shoulder arthroplasty; *rTSA*, reverse total shoulder arthroplasty.

inpatient vs. outpatient shoulder arthroplasty and for reimbursement arbitration of preauthorization with insurance firms. Overall, our model proved similar in accuracy to ML models in lower extremity total joint arthroplasty and superior to previously published models in TSA, making it a viable option for patient risk stratification before surgery.

Our study has limitations. Artificial intelligence models are a product of the data inputted, just as clinical decision making is a product of experience. As such, the algorithm was trained on a single database and does not represent a global sample. Because ML functions as a "black box," where the connections between the inputs and outputs are unknown, we are unable to determine the strength of each

Table IV Accuracy and AUC for total cost, LOS, and discharge disposition for chronic/degenerative (aTSA and rTSA) and acute/traumatic (rTSA and HSA) indications

Indication	Procedure	Total cost		Length of stay		Discharge disposition	
		Accuracy (%)	AUC	Accuracy (%)	AUC	Accuracy (%)	AUC
Chronic/degenerative	Combined	76.5	0.75	91.8	0.89	73.1	0.77
	aTSA	76.6	0.75	83.3	0.76	73.0	0.75
	rTSA	76.4	0.74	91.4	0.85	75.2	0.78
Acute/traumatic	Combined	70.3	0.72	79.1	0.78	72.0	0.79
	rTSA	69.3	0.71	80.2	0.80	73.2	0.79
	HSA	74.8	0.74	78.3	0.77	72.0	0.74

AUC, area under the curve; *LOS*, length of stay; *aTSA*, anatomic total shoulder arthroplasty; *rTSA*, reverse total shoulder arthroplasty; *HSA*, hemi-shoulder arthroplasty.

variable, although the accuracy and responsiveness are evident. The AUC values did not exceed 0.9, suggesting that the current model is not near-perfect and has room for future improvement. However, the presented preliminary model represents a dynamic framework. In the future, we will apply true cost ratios with Center for Medicare and Medicaid Services (CMS) comorbidity multiplier data, as well as readmissions data, to improve predictive strength in this data-dependent dynamic model. In addition, the patient cohort data set, although large and robust, dates back to 2003. As surgical techniques, patient rehabilitation protocols, and available implants have evolved tremendously since 2003 and the emphasis on value-based care has intensified, it is important that our ML-based models do not skew toward the older data that are less representative of current practices. However, as more data from other institutions are added to the existing algorithm, the inherent weaknesses of the antiquated data are less emphasized in the model's predictive ability.[8] Despite these limitations, our model provided fair to good AUCs in predicting discharge disposition, LOS, and total cost.

Conclusion

The results of this study demonstrate a preliminary ANN model for predicting LOS, inpatient cost, and discharge disposition for shoulder arthroplasty using a large, national database. Our model demonstrates fair to good accuracy and responsiveness. This study should be used to provide a framework for future studies to inform surgeons and hospitals about risk before surgery. If integrated into clinical practice, these models present the predictive opportunity to improve communication and planning for shoulder surgeons in the value-based era and reduce socioeconomic disparities for shoulder arthroplasty patients, thereby more appropriately encouraging health equity and quality of life at the population level one patient at a time.

Disclaimer

This study was funded internally by the Cleveland Clinic, Cleveland, OH, USA.

Benedict U. Nwachukwu received education fees from Smith & Nephew.

Samuel A. Taylor received consulting fees from DJO/Encore Medical, LP, and Mitek.

Eric T. Ricchetti received consulting fees from DJO/Encore Medical and LP.

None of these conflicts are related to the submitted work.

The other authors, their immediate families, and any research foundations 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.

References

1. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. 2016. Available at: http://arxiv.org/abs/1603.04467. Accessed December 6, 2019.
2. Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal 2000;22:717-27.
3. Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor Newsl 2004;6:20-9. https://doi.org/10.1145/1007730.1007735
4. Bean BA, Connor PM, Schiffern SC, Hamid N. Outpatient shoulder arthroplasty at an ambulatory surgery center using a multimodal pain management approach. J Am Acad Orthop Surg Glob Res Rev 2018;2:e064. https://doi.org/10.5435/JAAOSGlobal-D-18-00064
5. Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? J Arthroplasty 2018;33:2358-61. https://doi.org/10.1016/j.arth.2018.02.067
6. Biron DR, Sinha I, Kleiner JE, Aluthge DP, Goodman AD, Sarkar IN, et al. A novel machine learning model developed to assist in patient selection for outpatient total shoulder arthroplasty. J Am Acad Orthop Surg 2020;28:e580-5. https://doi.org/10.5435/JAAOS-D-19-00395
7. Cai X, Perez-Concha O, Coiera E, Martin-Sanchez F, Day R, Roffe D, et al. Real-time prediction of mortality, readmission, and length of stay using electronic health record data. J Am Med Inform Assoc 2016;23:553-61. https://doi.org/10.1093/jamia/ocv110
8. Day JS, Lau E, Ong KL, Williams GR, Ramsey ML, Kurtz SM. Prevalence and projections of total shoulder and elbow arthroplasty in the United States to 2015. J. Shoulder Elbow Surg 2010;19:1115-20. https://doi.org/10.1016/j.jse.2010.02.009
9. Durand WM, Johnson JR, Li NY, Yang J, Eltorai AEM, DePasse JM, et al. Hospital competitive intensity and perioperative outcomes following lumbar spinal fusion. Spine J 2018;18:626-31. https://doi.org/10.1016/j.spinee.2017.08.256
10. Fournier MN, Brolin TJ, Azar FM, Stephens R, Throckmorton TW. Identifying appropriate candidates for ambulatory outpatient shoulder arthroplasty: validation of a patient selection algorithm. J Shoulder Elbow Surg 2019;28:65-70. https://doi.org/10.1016/j.jse.2018.06.017
11. Gowd AK, Agarwalla A, Amin NH, Romeo AA, Nicholson GP, Verma NN, et al. Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty. J Shoulder Elbow Surg 2019;28:e410-21. https://doi.org/10.1016/j.jse.2019.05.017
12. Haeberle HS, Helm JM, Navarro SM, Karnuta JM, Schaffer JL, Callaghan JJ, et al. Artificial intelligence and machine learning in lower extremity arthroplasty: a review. J Arthroplasty 2019;34:2201-3. https://doi.org/10.1016/j.arth.2019.05.055
13. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.
14. Jiang JJ, Toor AS, Shi LL, Koh JL. Analysis of perioperative complications in patients after total shoulder arthroplasty and reverse total shoulder arthroplasty. J Shoulder Elbow Surg 2014;23:1852-9. https://doi.org/10.1016/j.jse.2014.04.008
15. Karnuta JM, Golubovsky JL, Haeberle HS, Rajan PV, Navarro SM, Kamath AF, et al. Can a machine learning model accurately predict

patient resource utilization following lumbar spinal fusion? Spine J 2020;20:329-36. https://doi.org/10.1016/j.spinee.2019.10.007
16. Karnuta JM, Navarro SM, Haeberle HS, Billow DG, Krebs VE, Ramkumar PN. Bundled care for hip fractures: a machine learning approach to an untenable patient-specific payment model. J Orthop Trauma 2019;33:324-30. https://doi.org/10.1097/BOT.0000000000001454
17. Leschinger T, Raiss P, Loew M, Zeifang F. Total shoulder arthroplasty: risk factors for intraoperative and postoperative complications in patients with primary arthritis. J Shoulder Elbow Surg 2017;26:e71-7. https://doi.org/10.1016/j.jse.2016.08.001
18. Menger RP, Kalakoti P, Pugely AJ, Nanda A, Sin A. Adolescent idiopathic scoliosis: risk factors for complications and the effect of hospital volume on outcomes. Neurosurg Focus 2017;43:E3. https://doi.org/10.3171/2017.6.FOCUS17300
19. Molino P, Dudin Y, Miryala SS. Ludwig: a type-based declarative deep learning toolbox. 2019. Available at: http://arxiv.org/abs/1909.07930. Accessed December 6, 2019.
20. 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.
21. Navarro SM, Wang EY, Haeberle HS, Mont MA, Krebs VE, Patterson BM, et al. Machine learning and primary total knee arthroplasty: patient forecasting for a patient-specific payment model. J Arthroplasty 2018;33:3617-23. https://doi.org/10.1016/j.arth.2018.08.028
22. Ramkumar PN, Karnuta JM, Navarro SM, Haeberle HS, Iorio R, Mont MA, et al. Preoperative prediction of value metrics and a patient-specific payment model for primary total hip arthroplasty: development and validation of a deep learning model. J Arthroplasty 2019;34:2228-34.e1. https://doi.org/10.1016/j.arth.2019.04.055
23. Ramkumar PN, Karnuta JM, Navarro SM, Haeberle HS, Scuderi GR, Mont MA, et al. Deep learning preoperatively predicts value metrics for primary total knee arthroplasty: development and validation of an artificial neural network model. J Arthroplasty 2019;34:2220-7.e1. https://doi.org/10.1016/j.arth.2019.05.034
24. CPI Inflation Calculator. Available at: https://www.bls.gov/data/inflation_calculator.htm. Accessed January 3, 2019.
25. NIS Description of Data Elements. Available at: https://www.hcup-us.ahrq.gov/db/nation/nis/nisdde.jsp. Accessed December 6, 2019.