Contents lists available at ScienceDirect

# Arthroplasty Today

Original research

# Machine Learning Model Developed to Aid in Patient Selection for Outpatient Total Joint Arthroplasty

Cesar D. Lopez, MD, Jessica Ding, BA *, David P. Trofa, MD, H. John Cooper, MD, Jeffrey A. Geller, MD, Thomas R. Hickernell, MD

*New York-Presbyterian/Columbia University Irving Medical Center, New York, NY, USA*

A B S T R A C T

*Background:* Patient selection for outpatient total joint arthroplasty (TJA) is important for optimizing patient outcomes. This study develops machine learning models that may aid in patient selection for outpatient TJA based on medical comorbidities and demographic factors.
*Methods:* This study queried elective total knee arthroplasty (TKA) and total hip arthroplasty (THA) cases during 2010-2018 in the American College of Surgeons National Surgical Quality Improvement Program. Artificial neural network models predicted same-day discharge and length of stay (LOS) fewer than 2 days (short LOS). Multiple linear and logistic regression analyses were used to identify variables significantly associated with predicted outcomes.
*Results:* A total of 284,731 TKA cases and 153,053 THA cases met inclusion criteria. For TKA, prediction of short LOS had an area under the receiver operating characteristic curve (AUC) of 0.767 and accuracy of 84.1%; prediction of same-day discharge had an AUC of 0.802 and accuracy of 89.2%. For THA, prediction of short LOS had an AUC of 0.757 and accuracy of 70.6%; prediction of same-day discharge had an AUC of 0.814 and accuracy of 78.8%.
*Conclusion:* This study developed machine learning models for aiding patient selection for outpatient TJA, through accurately predicting short LOS or outpatient vs inpatient cases. As outpatient TJA expands, it will be important to optimize preoperative patient selection and effectively screen surgical candidates from a broader patient population. Incorporating models such as these into electronic medical records could aid in decision-making and resource planning in real time.

## Introduction

Total joint arthroplasty (TJA), including total knee arthroplasty (TKA) and total hip arthroplasty (THA), are reproducible and successful treatments that continue to grow in the number of procedures every year [1], and efforts to reduce the economic impact of TJA has become a recent interest among arthroplasty surgeons. In TJA, the main sources of costs are length of stay (LOS) and complications, but rising costs have recently been reduced by shortening the LOS [2]. Advances in regional anesthesia, minimally invasive surgical techniques, operative duration, early mobilization, and multimodal pain control strategies have proven successful in safely reducing LOS and shifting toward outpatient TJA without significant differences in outcomes or short-term complications compared with inpatient TJA [3-6]. The Centers for Medicaid and Medicare Services (CMS) removed TKA from its inpatient-only list in 2018 and removed THA in 2020 and added TKA and THA to the Ambulatory Surgery Center Payable list in 2020, thus allowing for Medicare and Medicaid reimbursement for same-day procedures performed in hospitals and ambulatory surgery centers [7,8]. As a result, it is predicted that the volume of outpatient TJA will increase at much faster rates than inpatient TJA [9]. Thus, patient selection for outpatient TJA is important for optimizing patient outcomes, and enhanced patient selection tools may further incentivize performing more TJA procedures outside the hospital setting.

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.artd.2021.11.001.
* Corresponding author. 622 W 168th Street, PH-11, New York, NY 10032, USA. Tel.: +1 212 305 8193.
*E-mail address:* jd3628@cumc.columbia.edu

Machine learning (ML) is increasingly reported on in health care, including orthopedics, especially for its applications in predictive analytics. ML is a form of artificial intelligence (AI) that uses algorithms and mathematical models that can learn from data, identify patterns and complex relationships, and make automated decisions—oftentimes with minimal human intervention [10]. Compared with multivariate regression models that determine statistical correlations based on linear relationships between variables, ML can model complex nonlinear relationships and optimize predictive capability. Within orthopedics, AI/ML has been shown to be beneficial in surgical risk stratification and optimization [11], outcome prediction and diagnostics [12], and cost-efficiency analyses [13]. In the field of TJA, ML methods have been used in large database studies to predict LOS, costs, discharge dispositions, and postoperative complications, as well as developing patient-specific payment models based on patient risk levels [13-18].

The use of AI/ML is rapidly expanding in health care and has the potential to improve surgical care and reduce costs, especially hip and knee procedures. The purpose of this study is to develop ML models that may aid in patient selection for outpatient TKA by predicting short LOS and same-day discharge after TKA and THA based on patients' medical comorbidities and demographic factors. We hypothesize that the ML models can accurately predict patients who have a same-day discharge and short LOS (2 days or less) after THA and TKA.

## Material and methods

### Data source

This retrospective cohort study utilizes the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) registry from 2010 to 2018. The NSQIP data set comprises 273 variables collected by trained surgical clinical reviewers at more than 700 U.S. medical centers. Information collected includes deidentified demographical, preoperative, perioperative, and postoperative data, as well as 30-day morbidity and mortality outcomes for surgical patients in both the inpatient and outpatient setting. Data quality is ensured through training for surgical clinical reviewers and the inter-rater reliability audit, with the disagreement rate at 2.3% and up to 5% being considered acceptable [19]. Institutional review board approval was not required for this database study.

TKA and THA cases between the years of 2010 and 2018 were identified with the Current Procedural Terminology (CPT) codes 27,447 and 27,130, respectively. Cases with additional CPT codes, those with LOS greater than 30 days, and patients meeting sepsis, shock, or SIRS criteria in the 48 hours before surgery were excluded. Cases with incomplete data or missing predictive variables were also excluded. Patients who did not stay overnight had an LOS equal to 0 days and were categorized into the same-day discharge/outpatient group. Patients were included in the short LOS cohort if they had LOS less than or equal to 2 days. A total of 334,551 elective TKA cases were identified in the NSQIP registry between 2010 and 2018 (16,139 patients, or 4.8%, had missing data elements and were not used in our analysis), and 284,731 cases met inclusion criteria and were used in our analysis (Fig. 1a). Of 213,346 THA cases in the ACS-NSQIP database from 2010 to 2018 (12,605 patients, or 5.9%, had missing data elements and were not used in our analysis), 153,053 cases met the inclusion criteria and were included in the analyses (Fig. 1a).

### ML model development

Artificial neural network (ANN) models were developed and trained using the NSQIP database of elective TKA cases. All model development and analyses were performed using the TensorFlow Python open-source coding platform (Google Brain, Alphabet Inc., *Mountain View, California, USA*) and shared on the Google Colab cloud network (Google AI, Alphabet Inc., *Mountain View, California, USA*). The data were randomly sorted into a training set (80%) and testing set (20%) using the "random_state" command in Tensor-Flow. The training set data, containing predictive variables, were used to train and develop boosted decision tree and ANN models for predicting the outcome variables (prolonged operating time, return to operation room, and any complication). Predictive patient factors were used as inputs for the model to predict the indicated output, or outcome, variable. For each incorrect prediction, the model self-calibrates and "learns" from the patient data multiple times through a process of reiterative algorithm refinement until optimal model accuracy is achieved. The testing set data were used to evaluate ML model accuracy and performance (Fig. 1b). Categorical nonbinary variables (race, American Society of Anesthesiologists [ASA] class, anesthesia type) were incorporated into the model using one-hot-encoding. Continuous variables, such as age and body mass index (BMI), were converted into logarithmic variables to normalize the data set and minimize bias. Imbalanced data were managed using oversampling of the underrepresented outcome variable.

Model performance was determined by calculating the sensitivity and specificity of the model, which were used to develop a receiver operating characteristic (ROC) curve. The area under the ROC curve (AUC) was calculated, consistent with techniques used in prior studies [16,20]. AUC values range from 0.50 to 1 and measure a prediction models' discriminative ability, with a higher AUC value signifying better predictive ability and overall accuracy of the model correctly placing a patient into an outcome category. A model with an AUC of 1.0 is a perfect discriminator, 0.90 to 0.99 is considered excellent, 0.80 to 0.89 is good, 0.70 to 0.79 is fair, and 0.51 to 0.69 is considered poor [21]. Overall model accuracy (%) was calculated by dividing the sum of true positives (correct prediction of outcome) and true negatives (correct prediction of non-outcome) by the sample size.

### Statistical analysis

Data from the NSQIP registry were extracted, and patient variables included gender, age, race, BMI (calculated from the recorded height and weight available), ASA class, anesthesia type (general, spinal, and other—including intravenous sedation, regional, or epidural anesthesia), operative time, surgery setting (inpatient or outpatient), smoking history, and preoperative diagnosis of diabetes, dyspnea at rest or at moderate exertion, chronic obstructive pulmonary disease, congestive heart failure, or hypertension requiring medication, renal failure, dialysis requirement, disseminated cancer, use of steroids for a chronic condition, bleeding disorder, and need for a preoperative transfusion within 72 hours of surgery. The predicted outcomes included same-day discharge, short LOS, and any complication within 30 days after surgery. Any complication included the incidence of one or more of the following postoperative adverse outcomes within 30 days after surgery: surgical site infection, anemia requiring transfusion, deep vein thrombosis or pulmonary embolism, urinary tract infection, acute renal failure/compensation, sepsis, intubation-related event, pneumonia, myocardial infarction, cerebrovascular event, cardiac arrest, and return to the operation room.

Comparative and descriptive statistics were summarized with unpaired student's t-tests. Linear by linear association test was used to identify significant trends in same-day discharge and LOS during the study period (2010-2018). Multivariate linear regression analysis was conducted to determine factors associated with LOS. The

odds ratios of the predicted outcome variables were analyzed using a binary logistic regression which accounted for multiple preoperative patient characteristics, including medical comorbidities and demographic variables. The logistic regression model accounted for binary predictor variables including age (older than or younger than 70 years), race (White or non-White), BMI (obese or non-obese), ASA classification above 2, general anesthesia, spinal anesthesia, intravenous sedation anesthesia, regional anesthesia, diabetes, smoking history, as well as other collected preoperative characteristics and risk factors presented in the registry. Cases with missing features were excluded from the study. Results of the logistic regressions were reported as odds ratios with 95% confidence intervals (CIs). Confounding and interaction in the regression model were addressed by ensuring that the variance inflation factor, a measure of collinearity, did not exceed a value of 10. All statistical analyses were conducted on Stata, version 16.1 (Stata Corp., *College Station, Texas, USA*). Statistical significance was defined as $P < .05$.

## Results

### Total knee arthroplasty

There were 54,939 patients with a short LOS ($\leq 2$ days) and 5845 same-day discharge TKA cases (LOS = 0). To develop the ANN models, the data were randomly separated into a training set of
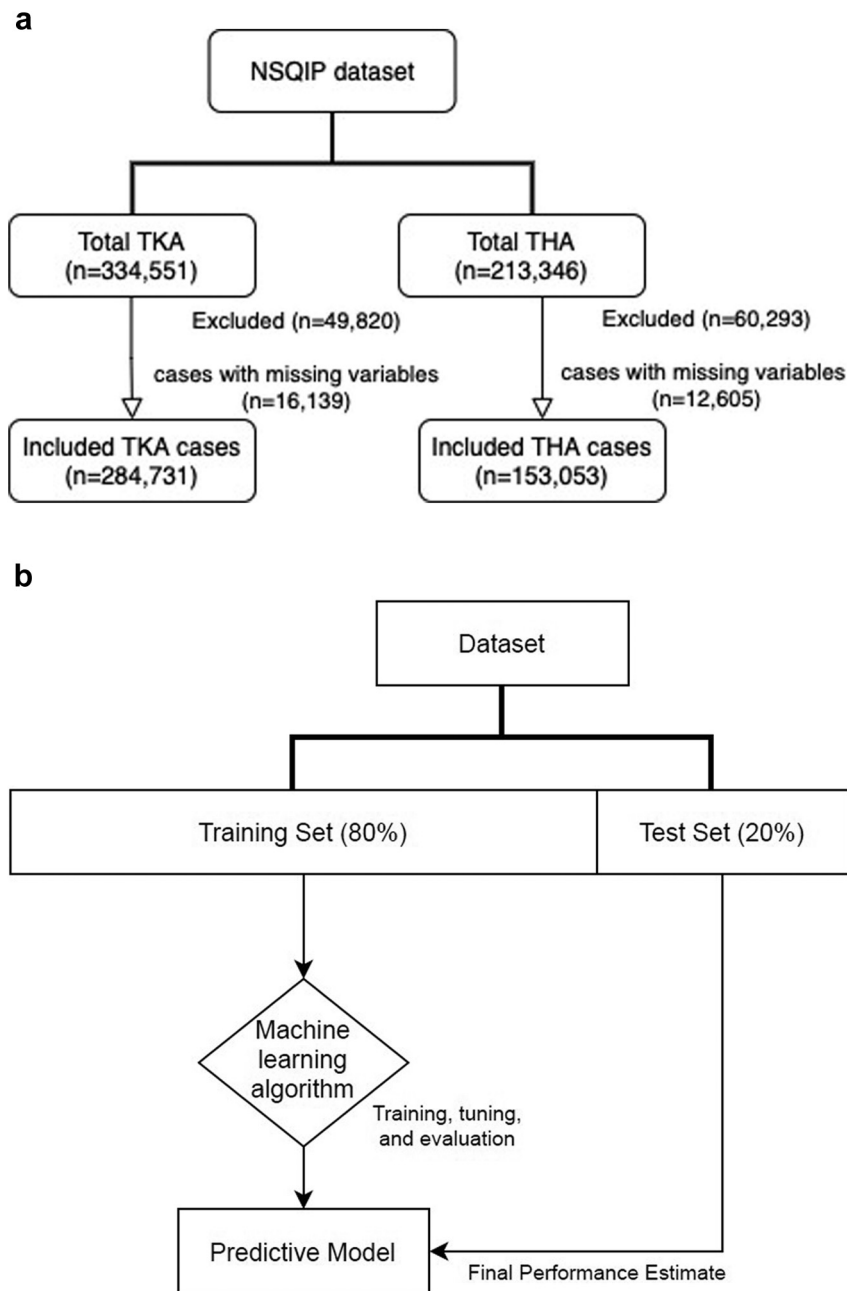


**Figure 1.** Flowcharts showing included TKA and THA cases (a) and machine learning model development (b).
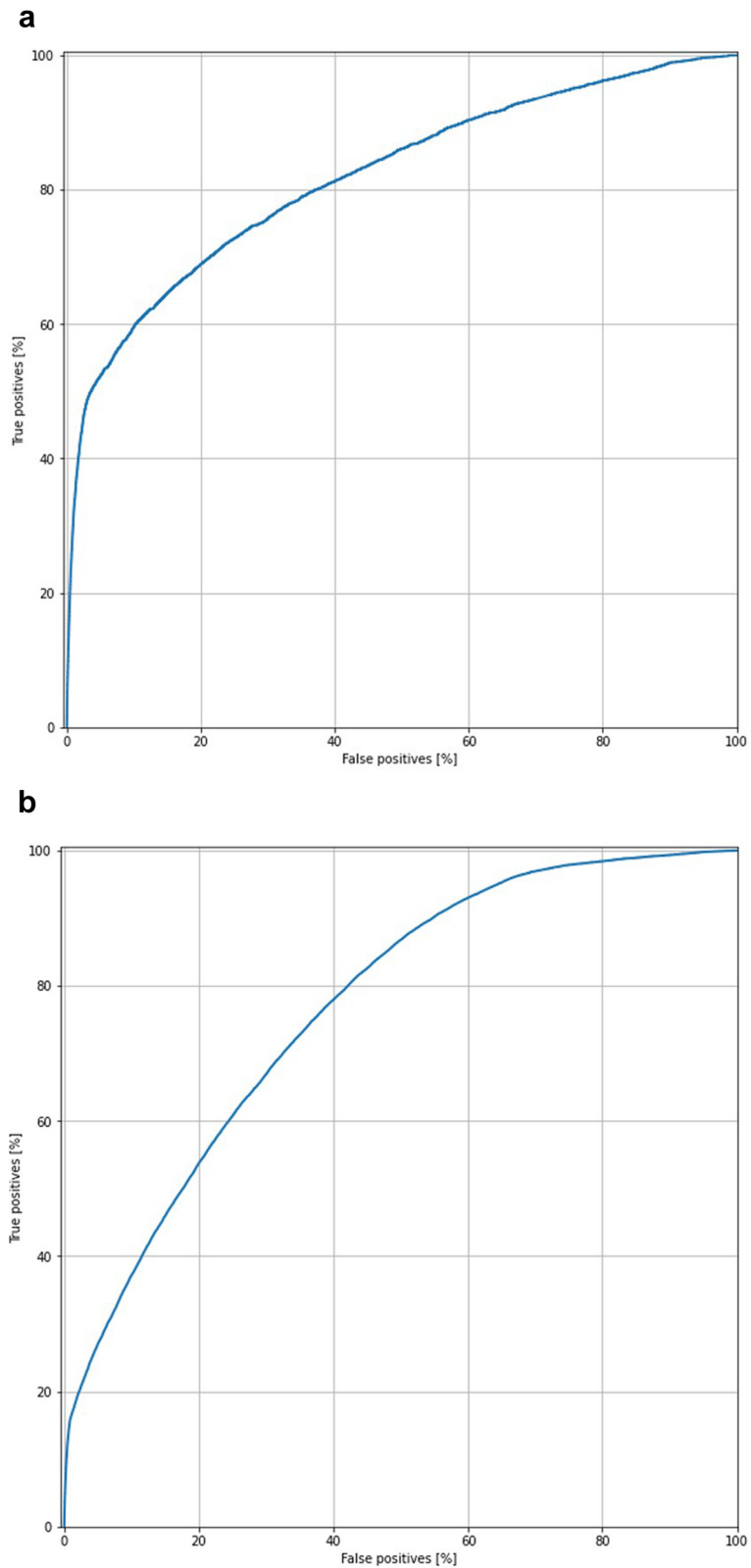
**Figure 2.** Area under the ROC curve of artificial neural network model for total knee arthroplasty (TKA) predicting (a) short LOS and (b) same-day discharge.

**Table 1**
Summary of patient demographics, medical comorbidities, perioperative and postoperative outcomes of ambulatory and nonambulatory TKA cases.

| Predictive factors | Nonambulatory (n = 278,886) | Ambulatory (n = 5845) | *P* value | All TKA (n = 284,731) |
|---|---|---|---|---|
| Female (%) | 61.8% | 53.8% | <.001 | 61.6% |
| White (%) | 87.7% | 87.7% | .956 | 87.7% |
| Avg. age (years) | 66.8 | 65.5 | <.001 | 66.8 |
| Age >70 years (%) | 35.5% | 29.6% | <.001 | 35.4% |
| Avg. BMI | 33.0 | 31.7 | <.001 | 33.0 |
| Obesity (%) | 63.5% | 57.0% | <.001 | 63.4% |
| Diabetes (%) | 18.1% | 14.5% | <.001 | 18.0% |
| Smoke (%) | 8.3% | 7.0% | .002 | 8.3% |
| Dyspnea (%) | 5.7% | 2.1% | <.001 | 5.6% |
| COPD (%) | 3.5% | 2.2% | <.001 | 3.5% |
| CHF (%) | 0.3% | 0.1% | .003 | 0.3% |
| Hypertension (%) | 65.0% | 56.9% | <.001 | 64.8% |
| Renal failure (%) | 0.0% | 0.0% | .224 | 0.0% |
| Dialysis (%) | 0.2% | 0.1% | <.001 | 0.2% |
| Cancer (%) | 0.10% | 0.09% | .845 | 0.1% |
| Wound infection (%) | 0.2% | 0.1% | <.001 | 0.2% |
| Steroid use (%) | 3.5% | 2.3% | <.001 | 3.5% |
| Weight loss (%) | 0.11% | 0.22% | <.001 | 0.1% |
| Bleeding disorder (%) | 2.1% | 1.4% | .001 | 2.1% |
| ASA class >2 (%) | 49.3% | 38.2% | <.001 | 49.1% |
| Outcomes | Nonambulatory (n = 278,886) | Ambulatory (n = 5845) | *P* value | All TKA (n = 284,731) |
| Inpatient (%) | 96.7% | 50.9% | <.001 | 95.8% |
| Avg. operative time (min) | 92.1 | 83.5 | <.001 | 91.9 |
| Prolonged operative time (%) | 16.0% | 8.6% | <.001 | 15.9% |
| Any complication (%) | 6.6% | 2.5% | <.001 | 6.5% |
| Nonhome discharge (%) | 22.0% | 6.0% | <.001 | 21.7% |
| Readmission (%) | 4.0% | 4.7% | .046 | 4.0% |
| Reoperation (%) | 1.2% | 1.0% | .223 | 1.1% |

COPD, chronic obstructive pulmonary disease; CHF, congestive heart failure.

216,960 patients and a testing set of 54,420 patients. After testing, the AUC of the ANN model for predicting same-day discharge was measured to be 0.802, with an overall accuracy of 89.2% (Fig. 2a). The ANN model for predicting short LOS had an AUC of 0.767, with overall accuracy of 84.1% (Fig. 2b).

Direct statistical comparison between same-day discharge and other TKA cases shows significant differences in average patient age (65.5 vs 66.8 years, respectively; *P* < .001), percent of those older than 70 years (29.6% vs 35.5%; *P* < .001), percent female (53.8% vs 61.8%; *P* < .001), average BMI (31.7 vs 33.0; *P* < .001), percent ASA class 3 and above (38.2% vs 49.3%; *P* < .001), and other medical comorbidities (Table 1). Multiple linear regression analysis identified several predictive factors which were significantly associated with total LOS, including female sex, age, BMI, diabetes, smoking and chronic obstructive pulmonary disease history, congestive heart failure, bleeding disorders, steroid use, and ASA classification (Table 2).

Predictive variables associated with same-day discharge included male sex, age less than 70 years, White race, BMI less than 30, no smoking history, and ASA classification below 3 (Fig. 3, Supplementary Table 1). The linear by linear association test reveals a trend in same-day discharge over time from 2010 to 2018, accounting for 0.5% (and average LOS of 3.3 days) of all elective TKAs in 2010 to 2.8% (and average LOS of 2.3 days) in 2018 (*P* < .001).

*Total hip arthroplasty*

When categorized by hospital LOS in the multivariate logistic regression analysis, there were 32,190 patients with short LOS (< 2 days). There were 2621 patients with same-day discharge. The ML model for predicting same-day discharge had an AUC of 0.814 and an accuracy of 78.8%. The ML model for predicting short LOS had an AUC of 0.757 and an accuracy of 70.6% (Fig. 4).

On average, same-day discharge THA patients were younger (61.6 years) than the rest of the patients with THA (64.9 years) (*P* < .001) and had a lower percentage of patients older than 70 years (19.8%) vs the rest of the patients with THA (31.8%) (*P* < .001) (Table 3). The same-day discharge group differed from the rest of the patients in percentage of female patients (45.9% vs 54.8%; *P* < .001), White patients (83.5% vs 76.2%; *P* < .001), and patients

**Table 2**
Factors associated with hospital length of stay (LOS) after TKA, on multivariate linear regression analysis.

| Variable | Coefficient | *P* value | 95% confidence interval | |
|---|---|---|---|---|
| Preoperative factors: | | | | |
| Sex (female) | 0.203 | **<.001** | 0.184 | 0.221 |
| Race (white) | −0.146 | **<.001** | −0.174 | −0.119 |
| Age | 0.016 | **<.001** | 0.015 | 0.017 |
| BMI | 0.010 | **<.001** | 0.009 | 0.011 |
| Diabetes | 0.188 | **<.001** | 0.164 | 0.212 |
| Smoking | 0.095 | **<.001** | 0.061 | 0.128 |
| Dyspnea | 0.214 | **<.001** | 0.173 | 0.254 |
| COPD | 0.334 | **<.001** | 0.285 | 0.383 |
| CHF | 0.646 | **<.001** | 0.483 | 0.810 |
| HTN | 0.057 | **<.001** | 0.037 | 0.077 |
| Renal failure | 0.968 | **.001** | 0.389 | 1.546 |
| Cancer | 0.170 | .240 | −0.114 | 0.454 |
| Steroid use | 0.164 | **<.001** | 0.115 | 0.212 |
| Bleeding disorder | 0.380 | **<.001** | 0.320 | 0.441 |
| ASA class | 0.260 | **<.001** | 0.241 | 0.279 |
| Perioperative factors: | | | | |
| Op. time | 0.006 | **<.001** | 0.006 | 0.006 |
| Anesthesia: general | 0.045 | .095 | −0.008 | 0.097 |
| Anesthesia: spinal | −0.258 | **<.001** | −0.311 | −0.205 |
| Anesthesia: other | −0.323 | **<.001** | −0.380 | −0.265 |

COPD, chronic obstructive pulmonary disease; CHF, congestive heart failure; HTN, hypertension.
Bold values are statistically significant.

with ASA class greater than 2 (26.1% vs 42.6%; $P < .001$). BMI was significantly different between the two groups (29.2 vs 30.3; $P < .001$) in addition to other comorbidities (Table 3). Multiple linear regression identified predictive factors associated with an increased LOS, including age, female sex, BMI, ASA class, smoking, steroid use, operative time, and other comorbidities (Table 4). Among same-day discharge patients, factors found to increase odds of a same-day discharge were White race ($P < .001$), use of spinal anesthesia or other anesthesia techniques ($P < .001$), and wound class above 1 ($P = .024$) (Fig. 5, Supplementary Table 2). The linear by linear association test reveals a trend in same-day discharge over time from 2010 to 2018, accounting for 0.6% (and average LOS of 3.2 days) of all elective THAs in 2010 to 6.7% (and average LOS of 2.0 days) in 2018 ($P < .001$).

## Discussion

This study developed ML models for aiding preoperative patient selection for outpatient TKA and THA, through accurate prediction of outpatient vs inpatient surgical candidates, as well as patients with short LOS (<2 days). The ANN models were more accurate in predicting same-day discharge for both TKA and THA, which both showed good predictive and discriminative ability (AUC >0.80). The models for predicting short LOS for both TKA and THA had fair predictive and discriminative ability (AUC >0.70). Prediction of same-day discharge was more accurate likely because the patients represented within this group have distinctly different characteristics and medical comorbidities from those who may need a longer LOS. Patients within the short LOS cohort are more difficult to differentiate because they may share overlapping features with the patients with same-day discharge and patients who may need longer LOS.

Given the increased cost-consciousness around healthcare spending, the use of ML in helping surgeons identify suitable candidates for outpatient TJA may provide potentially significant cost savings, especially as more insurers are introducing bundled payments for surgical care [22,23]. Previous retrospective studies have evaluated the accuracy of different preoperative calculators or scoring systems of patients to predict if the patients had a same-day discharge or less than two midnight stay after TJA [24—26]. The Outpatient Arthroplasty Risk Assessment was developed with nine comorbidity categories and points assigned to each category based on the presence and severity of each of these comorbidities. Outpatient Arthroplasty Risk Assessment scores within the early discharge range were found to be significantly associated with an actual early discharge in their study population and significantly more accurate than other scoring systems such as the American Society of Anesthesiologists Physical Status or the Charlson Co-morbidity Index [25,26]. Other calculators used to predict LOS include the ACS-NSQIP surgical risk calculator, which uses the type of procedure, demographic variables, and comorbidities to predict surgical risk and LOS, as well as the CMS Diagnosis-Related Group, which uses either the arithmetic or geometric means of the LOS of patients who have undergone the same procedure within the past year. It was found that both the ACS surgical risk calculator and the CMS Diagnosis-Related Group means significantly overestimate the LOS [24].

Our study uses neural networks, which are comprised of multiple "neural layers" that can process large amounts of data and share information using weighted connections which are optimized during the training process [27]. In particular, neural networks contain a hidden layer of nodes, where multiple input data can combine and transform, ultimately reaching the final ML model that predicts outcomes [13]. As a result, they can achieve predictive performance which provides an advantage in accurately modeling complex nonlinear relationships in high-volume data sets. These models have the potential to be utilized in clinical settings and integrated into a hospital electronic medical record, enabling
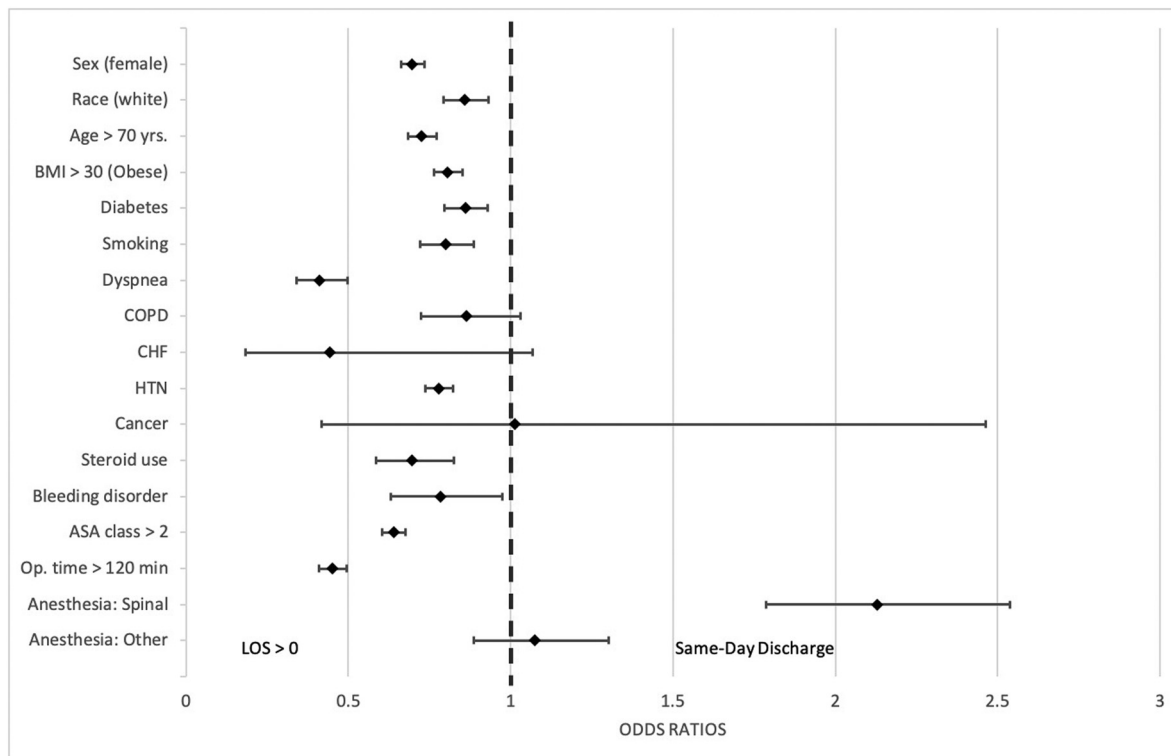


**Figure 3.** Factors associated with greater odds of same-day discharge after TKA, on multivariate logistic regression analysis.

a



b
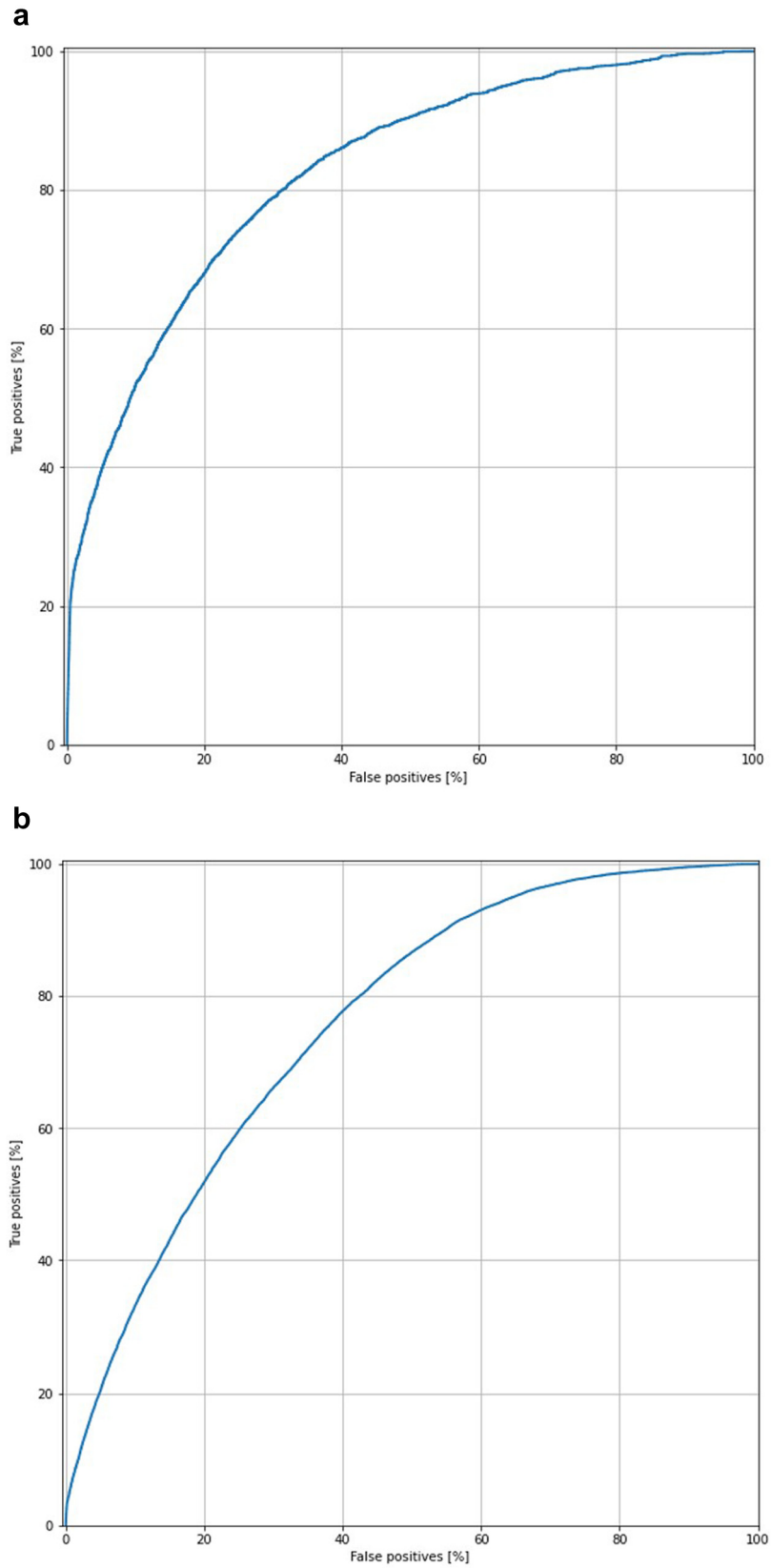


**Figure 4.** Area under the ROC curve of artificial neural network model for total hip arthroplasty (THA) predicting (a) short LOS and (b) same-day discharge.

**Table 3**
Summary of patient demographics and medical comorbidities and perioperative and postoperative outcomes of ambulatory and nonambulatory THA cases.

| Predictive factors | Nonambulatory (n = 150,432) | Ambulatory (n = 2621) | *P* value | All THA (n = 153,053) |
|---|---|---|---|---|
| Average age (years) | 64.9 | 61.6 | **<.001** | 64.9 |
| Age >70 years (%) | 31.8% | 19.8% | **<.001** | 31.6% |
| Female sex (%) | 54.8% | 45.9% | **<.001** | 54.7% |
| White race (%) | 76.2% | 83.4% | **<.001** | 76.4% |
| Average BMI | 30.3 | 29.2 | **<.001** | 30.3 |
| Obesity (%) | 46.8% | 40.3% | **<.001** | 46.7% |
| Diabetes (%) | 12.0% | 8.2% | **<.001** | 12.0% |
| Smoking (%) | 11.2% | 13.0% | **.006** | 13.0% |
| COPD (%) | 2.4% | 3.9% | **.001** | 3.9% |
| Dyspnea (%) | 4.6% | 1.5% | **<.001** | 4.5% |
| CHF (%) | 0.3% | 0.03% | **.016** | 0.3% |
| Hypertension (%) | 56.0% | 44.4% | **<.001** | 55.8% |
| Renal failure (%) | 0.04% | 0.04% | .952 | 0.04% |
| Dialysis (%) | 0.2% | 0.1% | .127 | 0.2% |
| Disseminated cancer (%) | 0.3% | 0.1% | .102 | 0.3% |
| Wound infection (%) | 0.3% | 0.2% | .259 | 0.3% |
| Steroid use (%) | 3.8% | 2.3% | **<.001** | 3.8% |
| Weight loss (%) | 0.2% | 0.2% | .682 | 0.19% |
| Bleeding disorder (%) | 2.1% | 0.1% | **<.001** | 2.1% |
| ASA class >2 (%) | 42.6% | 26.1% | **<.001** | 42.3% |
| Wound class >1 (%) | 0.4% | 0.6% | .152 | 0.4% |
| Outcomes | Nonambulatory (n = 150,432) | Ambulatory (n = 2621) | *P* value | All THA (n = 153,053) |
| Inpatient (%) | 99.1% | 80.7% | **<.001** | 98.8% |
| Average operative time (min) | 92.1 | 82.4 | **<.001** | 92.0 |
| Prolonged operative time (%) | 17.3% | 9.4% | **<.001** | 17.1% |
| Any complication (%) | 8.8% | 2.6% | **< .001** | 8.7% |
| Nonhome discharge (%) | 20.0% | 5.1% | **<.001** | 19.8% |
| Readmission (%) | 4.2% | 3.2% | **.014** | 4.2% |
| Reoperation (%) | 2.3% | 2.1% | .590 | 2.3% |

COPD, chronic obstructive pulmonary disease; CHF, congestive heart failure.
Bold values are statistically significant.

individualized assessment of patients who may be treated in outpatient settings, which are shown to be more cost-efficient with a streamlined model that eliminates inpatient and other ancillary expenses [28–30]. AI/ML models can also balance sensitivity (ie,

**Table 4**
Factors associated with hospital length of stay (LOS) after THA, on multiple linear regression analysis.

| Variable | Coefficient | *P* value | 95% confidence interval | |
|---|---|---|---|---|
| Preoperative | | | | |
| Age | 0.024 | **<.001** | 0.023 | 0.025 |
| Female sex | 0.236 | **<.001** | 0.214 | 0.259 |
| White race | −0.776 | **<.001** | −0.803 | −0.749 |
| BMI | 0.005 | **<.001** | 0.003 | 0.007 |
| Diabetes | 0.070 | **<.001** | 0.034 | 0.107 |
| Smoking | 0.089 | **<.001** | 0.054 | 0.124 |
| COPD | 0.288 | **<.001** | 0.227 | 0.349 |
| Dyspnea | 0.299 | **<.001** | 0.243 | 0.356 |
| CHF | 0.781 | **<.001** | 0.571 | 0.992 |
| Hypertension | 0.002 | .902 | −0.235 | 0.027 |
| Renal failure | 0.993 | **.001** | 0.396 | 1.590 |
| Dialysis | 1.008 | **<.001** | 0.762 | 1.255 |
| Disseminated cancer | 1.003 | **<.001** | 0.792 | 1.215 |
| Wound infection | 0.786 | **<.001** | 0.604 | 1.006 |
| Steroid use | 0.214 | **<.001** | 0.155 | 0.274 |
| Weight loss | 0.264 | **.042** | 0.010 | 0.520 |
| Bleeding disorder | 0.411 | **<.001** | 0.332 | 0.491 |
| ASA class | 0.248 | **<.001** | 0.226 | 0.269 |
| Wound class | 0.321 | **<.001** | 0.213 | 0.428 |
| Perioperative | | | | |
| Operative time | 0.005 | **<.001** | 0.004 | 0.005 |
| Anesthesia: spinal | −0.108 | **<.001** | −0.133 | −0.082 |
| Anesthesia: other | −0.301 | **<.001** | −0.336 | −0.267 |

COPD, chronic obstructive pulmonary disease; CHF, congestive heart failure.
Bold values are statistically significant.

selecting low-risk candidates) and specificity (ie, selecting high-risk candidates) of patient selection [31]. Our study showed that the 30-day complication rate was significantly lower in same-day discharge TKA and THA cases than in inpatient cases, which supports the feasibility of performing outpatient TKA and THA in patients with low risk of postoperative complications.

Recently, more ML techniques have been used in retrospective database studies to develop possible tools for preoperative patient selection for outpatient TKA and THA [28,29,30] Wei et al. developed an ANN model to predict outpatient vs inpatient status after TKA, which demonstrated good predictive and discriminative ability (AUC >0.80) [32]. Zhong et al. developed an ANN model and a random forest model, an ML technique that regresses and analyzes data through many decision trees, based on comorbidities and preoperative laboratory values to determine same-day discharge or an inpatient stay [33]. Their ANN model had fair predictive and discriminative ability (AUC >0.70), whereas the random forest model had good predictive and discriminative ability (AUC >0.80). Furthermore, Kugelman et al. used the extreme gradient boost ML technique, which builds multiple models based on the errors and residuals of previous models to reach an ultimate model which demonstrated good predictive and discriminative ability (AUC >0.80) [28]. Despite the variations on the methods used, our ANN models for both TKA and THA show similar results to the aforementioned studies (AUC >0.80 for same-day discharge for both TKA and THA).
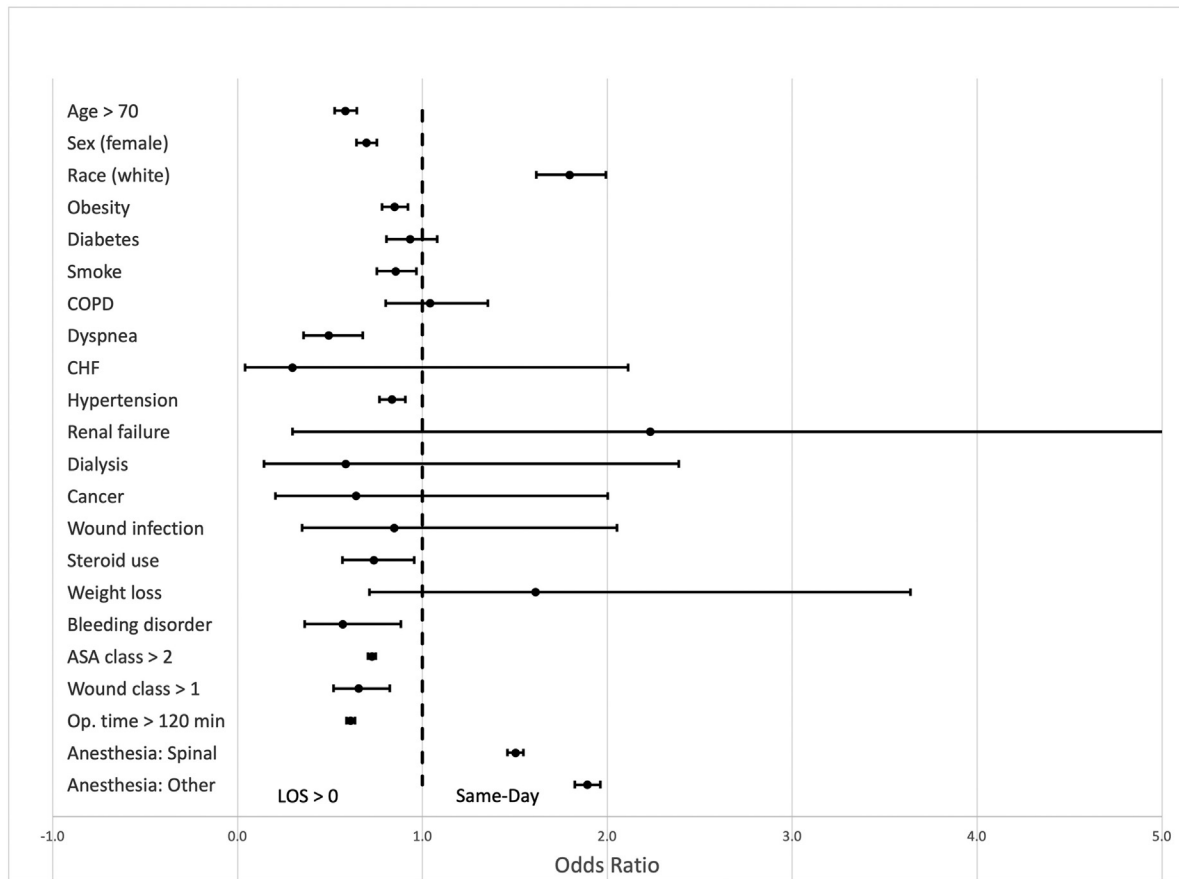
AI/ML models can accurately assess a patient's likelihood for same-day discharge based on multiple factors rather than identifying the individual risk factors or demographic variables that may contribute. However, ML models act as a "black box" because only the input predictive variables and output predictions are known and relationships and strengths of individual variables remain unknown [14]. Although ANN models have better predictive ability,

**Figure 5.** Factors associated with greater odds of same-day discharge after THA, on multivariate logistic regression analysis.

multivariate logistic regression is useful to determine which individual factors are associated with the outcomes [11]. The ANN models in this study used variables found on multiple linear regression to be significantly associated with total LOS and variables found on multivariate logistic regression analyses to be significantly associated with same-day discharge.

Our analysis found that patients with more manageable anesthesia needs are more likely to have shorter LOS, especially with a higher proportion of patients with ASA classification of 2 or less and undergoing spinal anesthesia rather than general anesthesia. Several studies have shown that an effective perioperative anesthetic plan is essential for the success of outpatient joint procedures, including TKA and THA [29,30]. Factors that commonly prolong hospital stay after TJA include inadequately controlled pain, nausea, vomiting, urinary retention, and limited mobility [29,34,35]. Regional anesthetic techniques such as spinal or epidural anesthesia have become widely used for TJA because they have shown to be associated with reduced 30-day mortality, lower blood transfusion frequency, fewer superficial wound infections, and shorter hospital LOS than general anesthesia [36,31]. Given the shift toward shortened LOS and reduced complications risk in outpatient TJA, anesthesia type may play an important role in ensuring cost-efficiency and quality of care associated with these procedures.

Regarding differences in demographics, we found that patients selected for same-day discharge TKA and THA were more likely to be male and younger than 70 years. Although the majority of TKA and THA patients are female (61.6% of nationwide TKA patients and 55.4% of THA patients) [37] there exists underrepresentation of female patients in outpatient TKA and THA studies. In a systematic

review of studies comparing evidence regarding the safety and feasibility of performing TJAs in the outpatient setting, the TKA and THA studies generally included an older population, with mean ages ranging from 55 to 68 years and the majority of studies including a larger proportion of males than female patients [38]. Among same-day discharge TKA and THA feasibility studies in the literature, factors associated with a longer hospital stay include female sex, increasing age, and an ASA classification of 3 or 4 [39]. In addition, our analysis found that White race was the main patient factor found to decrease odds of long LOS and increase odds of same-day discharge. Our results agree with those of a longitudinal analysis of racial disparities in TJA, which found that White patients were more likely to have shorter hospital stays, home discharge, and lower readmission rates, whereas Black patients had longer hospital stays, greater rates of nonhome discharge, more readmissions, and higher mortality rates. Black patients also had significantly lower rates of TJA utilization [40]. These racial disparities are systemic, as previous studies identified that biases exist in screening algorithms for comorbidities and for calculating costs, in addition to the smaller physician referral networks in predominantly Black communities, which may act as a barrier to patients obtaining quality care and good outcomes [41,42].

This study has several potential limitations. The ML tools developed in this study require external validation before conclusions can be made about its efficacy in clinical settings. Our analysis did not factor for surgery dates within the studied time period, especially as more recent surgeries are more likely to utilize new protocols, such as Enhanced Recovery After Surgery protocols and updated multimodal pain control protocols, which were designed

to shorten operative times and decrease LOS. In addition, there are inherent limitations when using a large database such as the NSQIP, which may include coding errors, missing data points, and inaccurate information. Sample bias makes ML model predictions only as reliable as the training data set. As a result, we decided to omit cases from our analysis that had missing data. We also addressed possible group attribution bias which results from training a model with data that contains an asymmetric view of certain groups—for example, if certain racial or gender groups have disproportionate outcomes compared with other groups, the model will be inclined to incorporate these inequities into their predictions as the standard. We considered the effect of societal inequities on ML model development and took these factors into consideration when evaluating the meaning of the model's predictions. Future studies may validate this application using a similar data set, especially as more TJAs are performed and more data become available.

## Conclusions

This study developed ML models for aiding preoperative patient selection for outpatient TKA and THA, through accurate prediction of same-day discharge and short LOS (<2 days) after TKA and THA. ANN models were more accurate in predicting same-day discharge, whereas prediction of short LOS was less accurate. As outpatient procedures become more common, it will be important to optimize preoperative patient selection and to effectively screen surgical candidates from a broader patient population. This study developed ML models to aid in patient selection for outpatient TKA and THA. Incorporating models such as these into electronic medical records could aid in decision-making and resource planning in real time. Given the increased cost-consciousness around healthcare spending, the use of ML to help surgeons identify suitable candidates for outpatient TJA may provide potentially significant cost savings, especially as more insurers are introducing bundled payments for surgical care.

## Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.artd.2021.11.001.

## References

[1] Sloan M, Premkumar A, Sheth NP. Projected volume of primary total joint arthroplasty in the U.S., 2014 to 2030. J Bone Joint Surg Am 2018;100(17): 1455.
[2] Molloy IB, Martin BI, Moschetti WE, et al. Effects of the length of stay on the cost of total knee and total hip arthroplasty from 2002 to 2013. J Bone Joint Surg Am 2017;99(5):402.
[3] Edwards PK, Milles JL, Stambough JB, et al. Inpatient versus outpatient total knee arthroplasty. J Knee Surg 2019;32(8):730.
[4] Isaac D, Falode T, Liu P, et al. Accelerated rehabilitation after total knee replacement. Knee 2005;12(5):346.
[5] Arshi A, Leong NL, Wang C, et al. Outpatient total hip arthroplasty in the United States: a population-based comparative analysis of complication rates. J Am Acad Orthop Surg 2019;27(2):61.
[6] Richards M, Alyousif H, Kim J, et al. An evaluation of the safety and effectiveness of total hip arthroplasty as an outpatient procedure: a Matched-cohort analysis. J Arthroplasty 2018;33(10):3206.
[7] Services CfMM. Ambulatory surgical Center (ASC) payment. 2019.

[8] Services, C.f.M.a.M. CY 2020 hospital outpatient PPS Policy Changes and payment rates and Ambulatory surgical Center payment system Policy Changes and payment rates CMS-1717-P. 2019. https://www.federalregister.gov/documents/2019/11/12/2019-24138/medicare-program-changes-to-hospital-outpatient-prospective-payment-and-ambulatory-surgical-center [accessed 09.07.20].
[9] Bert JM, Hooper J, Moen S. Outpatient total joint arthroplasty. Curr Rev Musculoskelet Med 2017;10(4):567.
[10] Deo RC. Machine learning in medicine. Circulation 2015;132(20):1920.
[11] Lin CC, Ou YK, Chen SH, et al. Comparison of artificial neural network and logistic regression models for predicting mortality in elderly patients with hip fracture. Injury 2010;41(8):869.
[12] Chung SW, Han SS, Lee JW, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop 2018;89(4):468.
[13] Ramkumar PN, Navarro SM, Haeberle HS, et al. Development and validation of a machine learning algorithm after primary total hip arthroplasty: applications to length of stay and payment models. J Arthroplasty 2019;34(4): 632.
[14] Ramkumar PN, Karnuta JM, Navarro SM, 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(10): 2220.
[15] Ramkumar PN, Karnuta JM, Navarro SM, 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(10):2228.
[16] Navarro SM, Wang EY, Haeberle HS, et al. Machine learning and primary total knee arthroplasty: patient Forecasting for a patient-specific payment model. J Arthroplasty 2018;33(12):3617.
[17] Karnuta JM, Navarro SM, Haeberle HS, et al. Predicting inpatient payments prior to lower Extremity arthroplasty using deep learning: which model Architecture is best? J Arthroplasty 2019;34(10):2235.
[18] Harris AHS, Kuo AC, Weng Y, et al. Can machine learning methods Produce accurate and Easy-to-use prediction models of 30-day complications and mortality after knee or hip arthroplasty? Clin Orthop Relat Res 2019;477(2): 452.
[19] ACSNSQIP, User Guide for the 2017 ACS NSQIP Participant Use data file (PUF). https://www.facs.org/-/media/files/quality-programs/nsqip/nsqip_puf_userguide_2017.ashx.
[20] Biron DR, Sinha I, Kleiner JE, et al. A Novel machine learning model developed to assist in patient selection for outpatient total shoulder arthroplasty. J Am Acad Orthop Surg 2019:e580.
[21] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143(1):29.
[22] Huang A, Ryu JJ, Dervin G. Cost savings of outpatient versus standard inpatient total knee arthroplasty. Can J Surg 2017;60(1):57.
[23] Karas V, Kildow BJ, Baumgartner BT, et al. Preoperative patient Profile in total hip and knee arthroplasty: predictive of increased Medicare payments in a bundled payment model. J Arthroplasty 2018;33(9):2728.
[24] Carr CJ, Mears SC, Barnes CL, et al. Length of stay after joint arthroplasty is less than predicted using two risk calculators. J Arthroplasty 2021;36(9): 3073.
[25] Meneghini RM, Ziemba-Davis M, Ishmael MK, et al. Safe selection of outpatient joint arthroplasty patients with medical risk stratification: the "outpatient Arthroplasty risk assessment score. J Arthroplasty 2017;32(8): 2325.
[26] Ziemba-Davis M, Caccavallo P, Meneghini RM. Outpatient joint arthroplasty-patient selection: Update on the outpatient Arthroplasty risk assessment score. J Arthroplasty 2019;34(7S):S40.
[27] Patel JL, Goyal RK. Applications of artificial neural networks in medical science. Curr Clin Pharmacol 2007;2(3):217.
[28] Kugelman DN, Teo G, Huang S, et al. A Novel machine learning predictive tool assessing outpatient or inpatient Designation for Medicare patients undergoing total hip arthroplasty. Arthroplast Today 2021;8:194.
[29] Cullom C, Weed JT. Anesthetic and Analgesic Management for outpatient knee arthroplasty. Curr Pain Headache Rep 2017;21(5):23.
[30] Berger RA, Sanders S, Gerlinger T, et al. Outpatient total knee arthroplasty with a minimally invasive technique. J Arthroplasty 2005;20(7 Suppl 3): 33.
[31] Memtsoudis SG, Sun X, Chiu YL, et al. Perioperative comparative effectiveness of anesthetic technique in orthopedic patients. Anesthesiology 2013;118(5): 1046.
[32] Wei C, Quan T, Wang KY, et al. Artificial neural network prediction of same-day discharge following primary total knee arthroplasty based on preoperative and intraoperative variables. Bone Joint J 2021;103-B(8): 1358.
[33] Zhong H, Poeran J, Gu A, et al. Machine learning approaches in predicting ambulatory same day discharge patients after total hip arthroplasty. Reg Anesth Pain Med 2021;46(9):779.
[34] Vehmeijer SBW, Husted H, Kehlet H. Outpatient total hip and knee arthroplasty. Acta Orthop 2018;89(2):141.

[35] Husted H, Lunn TH, Troelsen A, et al. Why still in hospital after fast-track hip and knee arthroplasty? Acta Orthop 2011;82(6):679.

[36] Johnson RL, Kopp SL, Burkle CM, et al. Neuraxial vs general anaesthesia for total hip and total knee arthroplasty: a systematic review of comparative-effectiveness research. Br J Anaesth 2016;116(2):163.

[37] (AAOS), A.A.o.O.S.. In: A.J.R.R. (AJRR), editor. The American joint Replacement registry Annual Report 2019; 2019. https://www.aaos.org/registries/publications/ajrr-annual-report/.

[38] Pollock M, Somerville L, Firth A, et al. Outpatient total hip arthroplasty, total knee arthroplasty, and Unicompartmental knee arthroplasty: a systematic review of the literature. JBJS Rev 2016;4(12):e4.

[39] Bozic KJ, Wagie A, Naessens JM, et al. Predictors of discharge to an inpatient extended care facility after total hip or knee arthroplasty. J Arthroplasty 2006;21(6 Suppl 2):151.

[40] Singh JA, Lu X, Rosenthal GE, et al. Racial disparities in knee and hip total joint arthroplasty: an 18-year analysis of national Medicare data. Ann Rheum Dis 2014;73(12):2107.

[41] Ghomrawi HMK, Funk RJ, Parks ML. Physician referral patterns and racial disparities in total hip replacement: a network analysis approach. PLoS One 2018;13(2):e0193014.

[42] Obermeyer Z, Powers B, Vogeli C. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447–53.