RESEARCH ARTICLE

Predicting Length of Stay using machine learning for total joint replacements performed at a rural community hospital

Srinivasan Sridhar [1] *, Bradley Whitaker[2], Amy Mouat-Hunter[3], Bernadette McCrory [1]

1 Mechanical and Industrial Engineering, Montana State University, Bozeman, Montana, United States of America, 2 Electrical and Computer Engineering, Montana State University, Bozeman, Montana, United States of America, 3 Bozeman Health, Bozeman, Montana, United States of America

* srinivasansridhar@montana.edu

Abstract

Background

Predicting patient's Length of Stay (LOS) before total joint replacement (TJR) surgery is vital for hospitals to optimally manage costs and resources. Many hospitals including in rural areas use publicly available models such as National Surgical Quality Improvement Program (NSQIP) calculator which, unfortunately, performs suboptimally when predicting LOS for TJR procedures.

Objective

The objective of this research was to develop a Machine Learning (ML) model to predict LOS for TJR procedures performed at a Perioperative Surgical Home implemented rural community hospital for better accuracy and interpretation than the NSQIP calculator.

Methods

A total of 158 TJR patients were collected and analyzed from a rural community hospital located in Montana. A random forest (RF) model was used to predict patient's LOS. For interpretation, permuted feature importance and partial dependence plot methods were used to identify the important variables and their relationship with the LOS.

Results

The root mean square error for the RF model (0.7) was lower than the NSQIP calculator (1.21). The five most important variables for predicting LOS were BMI, Duke Activity Status-Index, diabetes, patient's household income, and patient's age.

Conclusion

This pilot study is the first of its kind to develop an ML model to predict LOS for TJR procedures that were performed at a small-scale rural community hospital. This pilot study

contributes an approach for rural hospitals, making them more independent by developing their own predictions instead of relying on public models.

Introduction

With increasing rates of Total Joint Procedures (TJR) in the United States (US), predicting length of stay (LOS) is vital for hospitals to control costs, manage resources, and prevent post-operative complications [1–3]. LOS is the period between the time when the patient is admitted after the surgery to the time when the patient is discharged from the hospital. A longer LOS has been shown to negatively affect the quality of care and patient satisfaction [4, 5]. Specifically, patients with longer LOS drastically increases surgical expenses due to the high average inpatient costs at hospitals, which is $2,000 to $3,000 per day [6–8]. Moreover, past studies have demonstrated that patients with longer LOS have higher chances of experiencing poor postoperative surgical outcomes such as readmission and discharge to nursing or rehabilitation facility [9, 10].

To better manage surgical costs, allocate resources, and improve patient satisfaction, researchers have identified factors responsible for longer LOS using various analytic tools, including statistical and machine learning (ML) models [2, 11–13]. However, limited work has been done to predict LOS at community hospitals located in rural areas. It is more challenging to predict patients' LOS in rural than urban areas because community hospitals located in rural areas often lack adequate resources–such as data, staff, and expertise–needed to develop an accurate predictive model [14]. Instead, many hospitals use publicly available models that were developed to quantify patient risk before surgery [15].

One such available model is the National Surgical Quality Improvement Program (NSQIP) risk calculator. The NSQIP risk calculator is widely used by hospitals to predict risks for TJR procedures performed on knees, hips, and shoulders [15]. At a single hospital or institution, the NSQIP risk calculator can be useful for surgeons to assess patient risk, but it has been found to be suboptimal when predicting LOS for TJR procedures [8, 15, 16]. Interpretability is also a concern for the NSQIP risk calculator. In NSQIP, the risk factors are not quantified (i.e., it does not let the clinicians know which risk factor is most associated with a particular outcome). NSQIP's lack of factors quantification demonstrates that this predictive tool may not be suitable for evidence-based decision-making patient optimization in PSH care models. Predicting LOS has become more vital for orthopedic clinicians since recently the Center for Medicare and Medicaid services (CMS) removed knee and hip arthroplasty from the inpatient list [17, 18]. Knowing the risk factors and which patient will stay longer in the hospital after surgery are pertinent metrics for clinicians, administrators, and payers to correctly evaluate resource utilization, cost, and severity of illness [19].

Recent research has explored a promising application of ML for predicting surgical outcomes [20]. ML is a part of Artificial Intelligence (AI) which uses algorithms to recognize patterns in data to make predictions [21]. In the past decade, the application of ML in healthcare increased drastically due to wider usage of Electronic Medical Record (EMR) systems, which enabled the generation of 'big data' [22–25]. Big data has motivated many researchers and clinicians to apply ML and predict various health outcomes to improve patient treatment and quality of care [26, 27]. Proportionally, the role of ML surged in the orthopedic field as well [20, 28, 29]. For example, Ramkumar et al. [30] developed an ML using a naïve Bayesian model to forecast LOS and payments for total hip arthroplasty (THA). The authors used it as a

classification problem by grouping the inpatient payments to <$12000, $12100-$24000, and >$24000, and LOS to 1–2, 3–5, and 6+days. The ML model was found valid, reliable, and accurate in its predictions with AUC of 0.87 and 0.71 for LOS and payment respectively. Similarly, Li et al. [13] developed an ML model which accurately predicted the LOS for total knee arthroplasty (TKA) procedure with an AUC greater than 0.73. Gabriel et al. [31] predicted LOS using ML methods to determine patients who do not require prolonged LOS. The developed model was reliable with AUC of more than 0.7 and helped hospital administrators to plan resources to avoid both overcrowding and underutilization of TJR patients. Relatedly, Han et al. [32] predicted LOS for knee patients in China and identified that Random Forest model predicted more accurately than other ML models with AUC of 0.7. Aazad et al. [33] utilized various ML methods to predict the duration of surgery and LOS which significantly contributed to an increase in surgical cost. The study found that PyTorch MLP performed better among other ML models with least Mean Square Error of 0.89 and 0.69 for duration of surgery and LOS, respectively. Klemt et al. [34] used three ML methods to predict LOS for knee revision patients. The authors observed all three ML models performed well with AUC more than 0.8 and decision curve analysis with P-value <0.01. In addition, the authors identified that patients' age, Charles comorbidity index, and body mass index, were strong predictors for predicting LOS. The above examples are some of the recent studies that used ML methods to predict LOS for TJR procedures. Several studies were also performed in predicting TJR outcomes including surgical site infection, readmission, discharge disposition, and other adverse events [20, 35].

Yet, limited research has been performed with ML in rural hospitals. In addition, to the authors' knowledge, no research has been performed in predicting surgical outcomes at a rural hospital that adopted newly emerging coordinated surgical system in orthopedics—the Perioperative Surgical Home (PSH) [14]. The PSH model of care was created by the leaders within the American Society of Anesthesiologists (ASA) to improve surgical outcomes and patient experience [36–38]. Therefore, this research focuses on developing an ML model to predict LOS for TJR procedures performed at a PSH-implemented community hospital. Despite the challenges associated with limited availability of data, this study expects that the developed ML model will perform better in accuracy and interpretation than the NSQIP risk calculator.

Methods

Data collection and preprocessing

A rural community hospital formed an integrated PSH outpatient clinic in November 2018. The hospital was an 83-bed, licensed level-III trauma center primarily serving three rural counties, but often sees patients from more than 10 surrounding rural counties which span 9,000 square miles and approximately 136,000 residents. The PSH outpatient clinic was affiliated with the hospital began seeing patients preoperatively for TJR including hip, knee, and shoulder replacements. A written consent was obtained from the patients before their participation in this study. The consent was documented and attached to patient's EMR for reference. The study had no patients who are younger than 18. The scope of this pilot study focused on elective procedures and excluded any revisions. A total of 158 TJR patients were analyzed retrospectively after visiting the PSH clinic during preoperative surgical assessment from August to December 2020. All preoperative surgical assessments were performed within 30 days before surgery.

A total of 28 independent variables were collected for each patient, which included 20 NSQIP preoperative characteristics and eight additional variables. The NSQIP characteristics were collected and entered manually into the risk calculator to determine NSQIP-predicted

LOS [39]. Additional variables were extracted from the Electrical Medical Record (EMR) such as the Duke Activity Status Index (DASI) [40], living status (whether the patient was living alone or with family), patient's household income, postoperative nausea and vomiting score (PONV) [41], depression (whether the patient was depressed at the time of the assessment), preoperative preparation period (the number of days between assessment and surgery), distance traveled by the patient (in miles, from their residence to the hospital), and patient primary insurance type (private or public payer). These additional variables were included in the analysis as they were found to be closely associated with risk for poor surgical outcomes in past studies [3, 37, 42–45]. After patient's discharge from the hospital, the actual LOS was extracted from the EMR.

Eleven NSQIP categorical variables were excluded as there were no cases observed in those categories: steroid use, ascites, systemic sepsis, ventilator-dependent, emergency case, dissemination of cancer, congestive heart failure, chronic obstructive pulmonary disease, dyspnea, dialysis, and acute renal failure. After exclusion, a total of 17 variables were considered in the analysis (Table 1). The variable distance traveled by the patient was produced by calculating the mileage between their zipcode and the hospital on Google Maps [46]. There were no missing values for the independent variables except for patient's household income. Thirty-six patients (23%) out of 158 declined to provide their household income to clinicians during the assessment. These missing values were imputed using the median income of the patient's zipcode [47]. The 2019 US Census Bureau database was used to input the zipcode median income obtained from Montana Demographics by Cubits [48]. Additionally, the independent variables that were classified as ordinal were ranked accordingly for use in the correlation analysis (Table 1).

Table 1. Variable description.

	Variables	Type	Description
Response Variable	Length of Stay (LOS)	Continuous	Length of stay post surgery in hours
NSQIP Variables [39]	Procedure	Categorical	Total Knee (TKA), Total Hip (THA), Total Shoulder (TSA) Arthroplasty
	Age	Continuous	Patient age in years
	Sex	Dichotomous	Gender: Male, Female
	Functional Status	Ordinal	Fully Independent (1), Partially independent (2), Fully dependent (3)
	ASA Class	Ordinal	Healthy patient (1), Mild systemic disease (2), Severe systemic disease (3), Severe disease with constant threat to life (4)
	Diabetes	Dichotomous	No (1), Yes (2)
	Hypertension	Dichotomous	No (1), Yes (2)
	Current Smoker within 1 year	Dichotomous	No (1), Yes (2)
	BMI	Continuous	Body Mass Index in Kg/m^2
Additional variables	Duke Activity Status Index [40]	Continuous	Functional capacity scaled from 2.74 (low functional activity) to 9.89 (high functional activity) in METs
	Living Status [45]	Dichotomous	Patient's primary household status—lives alone, living with another
	Patient's Household Income [42]	Ordinal	Patient's household income level: <$49K (1), $50-99k (2), $100+k (3)
	Postoperative Nausea and Vomiting Score (PONV) [41]	Ordinal	To estimate nausea after surgery due to anesthesia, scaled 0 (low chances of nausea) to 4 (high chances of nausea)
	Depression [3]	Dichotomous	Depression at the time of preoperative assessment: No (1), Yes (2)
	Preoperative Engagement Period [49]	Discrete-Ordinal	Difference in time period (in days)
	Patient Insurance Type [50]	Dichotomous	Patient's payer—Public, Private type
	Distance Travelled by the patient [44]	Continuous	Distance traveled by the patient from their resident to the hospital (in miles)

https://doi.org/10.1371/journal.pone.0277479.t001

Feature selection

Feature selection was performed to identify the most important features, i.e., independent variables to train a novel ML model, improve accuracy, and reduce computation time [51]. This pilot study used Spearman's rank correlation filter method [52] and Boruta wrapper method [53] to identify the important independent variables to predict LOS.

Spearman's rank correlation. Correlation analysis was performed to identify highly correlated variables (correlation value close to 1 or -1). Independent variables that are highly correlated with one another can act as redundant in the analysis as they do not improve the model performance but increase the model complexity and computation time [54]. The database in this pilot study consisted of both continuous and ordinal variables and Spearman's rank method was used to perform correlation analysis [52]. No highly correlated (correlation more than 0.7) independent variables were observed (Fig 1). Similarly, there were no highly correlated independent variables with dependent variable, LOS. The remaining feature selection was refined using the wrapper method.

Boruta feature selection method. Boruta feature selection is a wrapper method that utilizes the random forest algorithm to rank variable importance [53]. The Boruta uses shadow variables that are obtained by shuffling the original values across objects [53, 55]. The Boruta ranks variable importance using shadow variables as a reference. Any variable that showed higher importance than shadow variables is considered important [53]. The Boruta is known to have comparable, if not superior, ability in independent variable selection than other available feature selection methods [56].

This study simulated Boruta for 100 runs to eliminate random errors in the results. The independent variables that were found important and statistically significant in the binomial distribution (P-value < 0.01) were selected for the prediction modeling. The variables that were found important were Insurance Type, Patient's Household Income, DASI, BMI, Functional Status, Diabetes, Living Status, and Age (Fig 2). The rest of the independent variables were found not important. This study considered the important and excluded the non-important variables from the ML analysis.
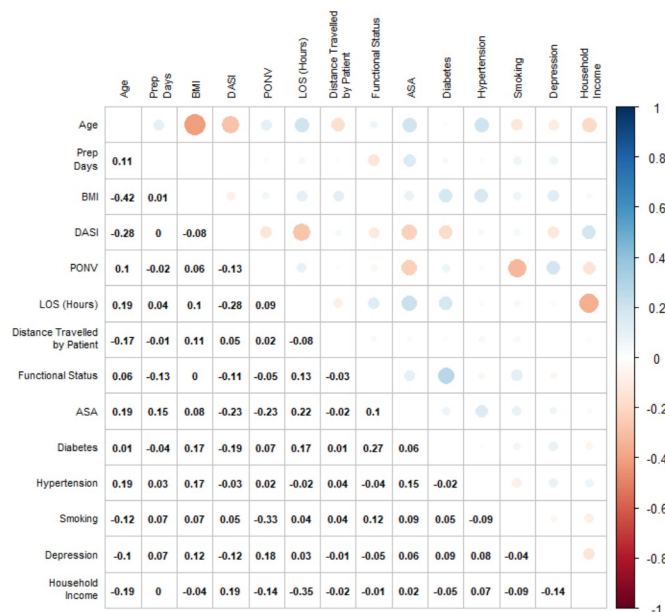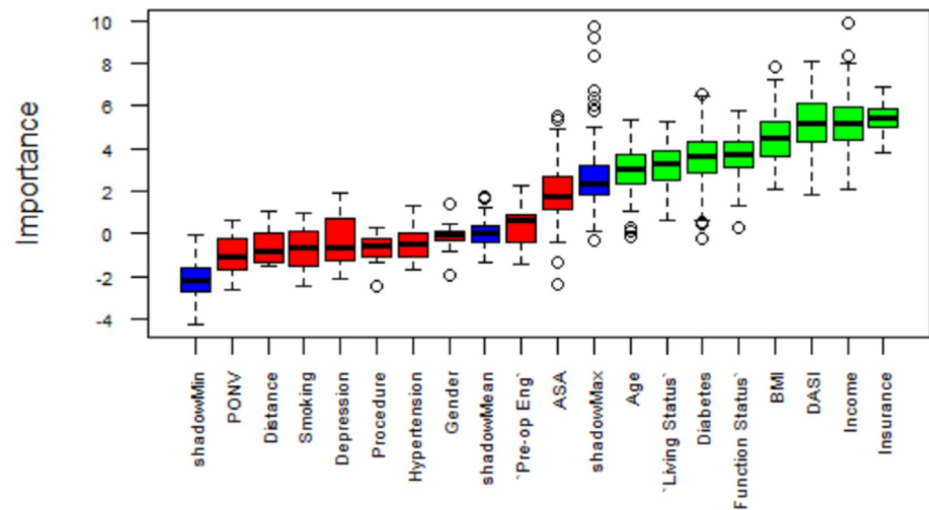

Fig 1. Correlation plot.

https://doi.org/10.1371/journal.pone.0277479.g001

Fig 2. Independent variable selection using Boruta [53].

Model selection

The study found no strong linear correlations between dependent and independent variables indicating that the dependent variable and independent variables were non-linearly related. To identify non-linear patterns, Random Forest (RF) method was used. RF is an ensemble learning where the output of multiple decision trees is combined to deliver the final outcome or prediction [57]. Past studies also have exhibited the effectiveness of RF in predicting surgical outcomes with limited data, similar to this study [58, 59]. Compared to RF, other commonly known ML methods such as Neural Networks, Boosted Trees, and Support Vector Machines have more tuning parameters and often require more data to train [60–62]. Due to the very small sample size and RF being one of the most familiar ML methods in predicting orthopedic surgical outcomes [20], this preliminary study used only RF to predict patient LOS after TJR procedures performed at a community hospital. Other ML methods will be considered in the future upon more data availability (n> 2,000).

Data splitting and tuning the parameter

The data was split into training (80%, n = 127) and testing (20%, n = 31) (Table 2). The data splitting and tuning were performed using the CARET package in R software [62].

The RF has two main tuning parameters, which are the number of trees in the forest (n_{tree}) and the number of variables randomly sampled as candidates at each split (m_{try}). One thousand trees (n_{tree}) were used in the forest, as recommended by past researchers [57, 62]. Having more trees in the RF does not affect the performance of the prediction negatively, but it can increase computation time [63]. The study expects no significant increase in computation time by using 1000 trees as compared to fewer trees (100 to 500 trees) because there were only 127 data points in the analysis. Leave One Out Cross Validation (LOOCV) was used to find the optimal number of variables available for splitting at each tree node–m_{try} [64] (Fig 3). The study chose LOOCV as it was more appropriate to use in smaller datasets [62, 65].

Model validation

The NSQIP predicted LOS in days, segmented into half-day intervals. For appropriate comparison, the actual LOS and RF-predicted LOS, which were in hours, were converted to days.

Table 2. Baseline of training and test sample.

Variables		Train data (n = 127) Mean (SD) [Min, Max] or N (%)	Test Data (n = 31) Mean (SD) [Min, Max] or N (%)
Length of Stay (LOS)		41 (29) [3, 195]	36 (16.5) [5, 77.9]
Age		68.6 (9.63) [43, 91]	70.7 (7.43) [54, 84]
Functional Status			
	Independent	128 (98%)	31 (100%)
Diabetes			
	Yes	13 (10%)	2 (6%)
BMI		30.7 (7.25) [17, 58]	29.9 (4.28) [20, 40]
Duke Activity Score Index (DASI)		7.5 (1.67) [4.4, 9.89]	7.29 (1.5) [4.64, 9.89]
Living Status			
	Alone	26 (20%)	9 (29%)
Patient's Household Income			
	<40K	50 (39%)	14 (45%)
	50-99k	56 (44%)	12 (39%)
Patient Insurance Type			
	Private	34 (27%)	5 (16%)

Conversion was performed in 12 hour-intervals, such that LOS less than or equal to 12 hours was considered 0.5 days, LOS less than or equal to 24 hours was considered 1 day, and LOS less than and equal to 36 hours was considered 1.5 days, and so on. The study was a regression problem as the response variable LOS was numeric and continuous. Therefore, the Root Mean Square Error (RMSE) was used as the cost function to validate the model's prediction performance i.e., the lower the RMSE value, the better the model is performing [66]. Pearson's correlation was also performed to determine the linear relationship between predicted LOS and actual LOS i.e., high correlation demonstrates better model performance. This pilot study expects the developed RF model to have a lower RMSE and higher correlation than NSQIP.

Model interpretation

RF is an ensemble ML model that aggregates many independent decision trees to make a prediction. Though it greatly helps the accuracy of the prediction, the model acts as a black box and the interpretation is complex. In RF, each tree has a different structure as they are built based on the subset of predictors or independent variables that were randomly selected [57]. To understand and explain each tree in the forest is complex and nearly impossible, which makes the interpretability of a RF model difficult. However, past researchers have been able to interpret ML models (such as RF) substantially, if not completely, using different model-agnostic approaches [67, 68]. In ML, unlike model-specific methods, model-agnostic methods can be applied to any ML model for interpretation. This makes the model agnostic approaches more flexible and reliable than model-specific when interpreting different ML models [67]. This pilot study applied two widely used model-agnostic methods; permutation feature importance and partial dependence plots for interpreting and explicating the relationship of the variables in the RF model [69].

Permutation Feature Importance. Permutation Feature Importance (PFI) method measures importance by calculating the increase in prediction error after permuting the independent variable [70]. In other words, the independent variable in the data set is randomly permuted which degrades the relationship between that independent variable and the response variable. The importance of a variable is measured based on the difference in cost function before and after the variable is permuted [67]. The PFI approach uses randomness in the
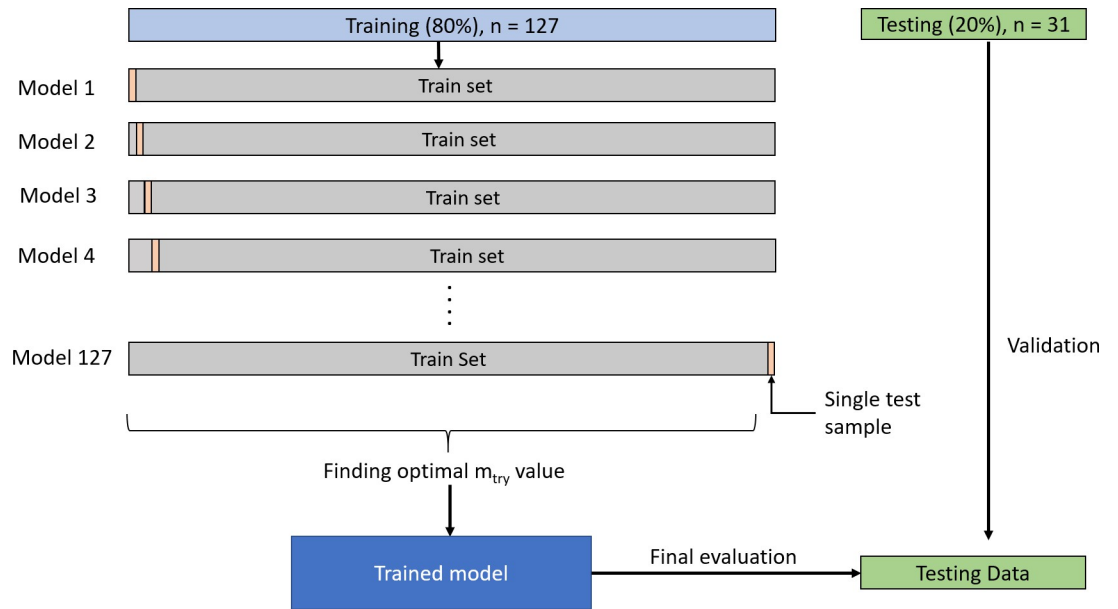
Fig 3. Leave One Out Cross Validation.

procedure to evaluate the importance. Thus, this study simulated this method for 100 times to minimize random errors and ranked the important variables based on the average value. The PFI algorithm used in this study adapted from [67, 71]:

Let j be the total number of independent variables

Let X be the independent variable

Let E be the baseline RMSE value for the trained model

Let F be a two-dimensional matrix of RMSE values after a feature is permuted in the training data

1. For k = 1, 2, 3, ...100: (for simulating 100 times)

    a. For i = 1, 2, 3, ...j:

        i. Permute the values of variable X_i in the training data.

        ii. Recompute the RMSE value on the permuted data–F_{ki}.

2. For i = 1, 2, 3, ...j:

    a. Compute average importance for each variable, $Imp_i = \frac{1}{100}\sum_{k=1}^{100}(E - F_{ki})$

3. Sort the average importance (Imp_i) by descending order.

Partial Dependence Plot. The Partial Dependence Plot (PDP) is an another agnostic method that helps to understand the marginal effect of a variable on the predicted outcome of an ML model [72]. The PDP shows the relationship between a response variable and an independent variable whether they are linear, monotonic, or complex [67]. This demonstrates how the response variable changes as the value of an independent variable while considering the average effect of all the other independent variables in the model [69]. The biggest disadvantage of PDP is that it is effective when the variables are not correlated. However, this study had

no strong correlation between any independent variables. Therefore, the PDP approach was more ideal method as they were easy to implement and more importantly, simple to interpret. In Eq 1, f_S was the partial function which was estimated by calculating the average value in the training data. The x_S was the independent variable that was being plotted in the PDP where $S \subset (1, 2,...j)$. The $x_C^{(i)}$ were the rest of the independent variables in the training data where C was complement of S. The variable n was the total number of data points in the training data which was 127. [67]

$$f_S(x_s) = \frac{1}{n}\sum_{i=1}^{n} f(x_s, x_C^{(i)})$$

(1)

All data handling, visualizations, statistical analysis, ML modeling and interpretations were performed using R (V 4.0.3, Vienna, Austria). The data were recorded and secured in an encrypted database and were accessed only by the authors and the clinicians. The research study was conducted at a single rural community hospital and was approved by Montana State University Institute Review Board (approval number–BM050819 (EX))

Results

Model performance

The m_{try} with the lowest RMSE value was found at 2. The RMSE of RF for the train data (n = 127) and test data (n = 31) were 0.71 and 0.61, respectively. The RMSE of RF for the whole data (n = 158) was 0.7, which was lower than NSQIP which was 1.21 (Fig 4). The Pearson's correlation between predicted and actual for NSQIP and RF were 0.22 and 0.82 (Fig 5). Compared to NSQIP, the RF model had lower RMSE and higher Pearson's correlation, making it a better model for predicting patient LOS after the TJR procedure.
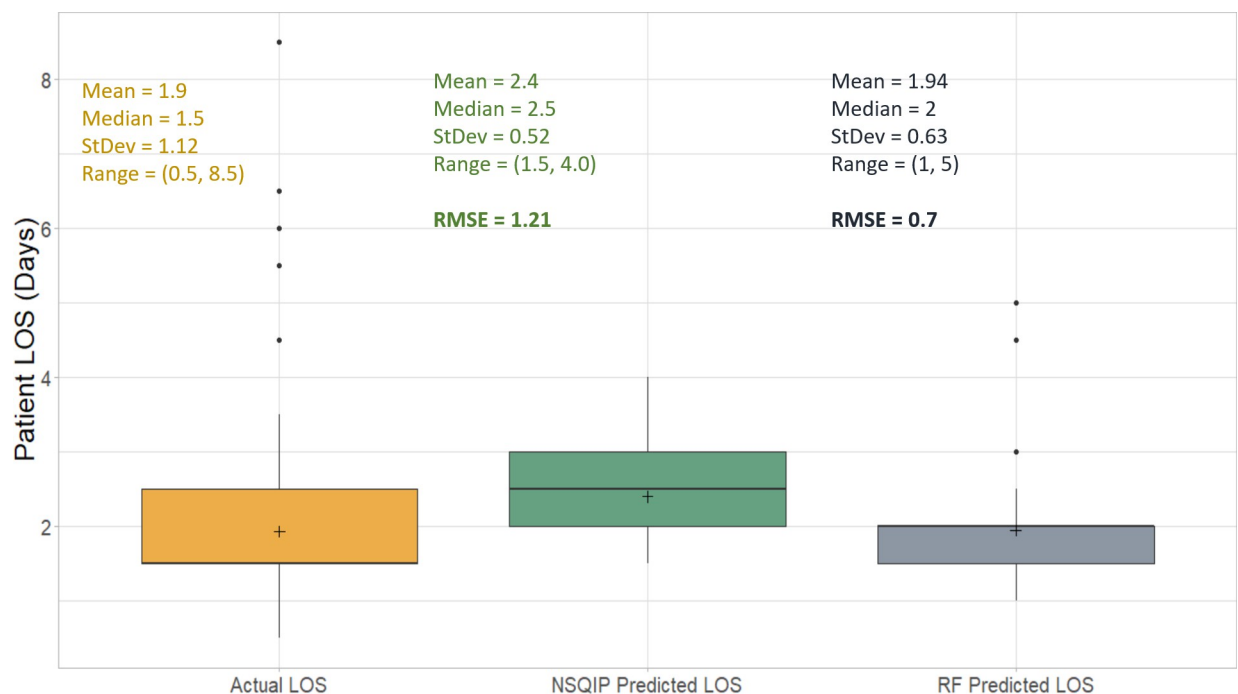

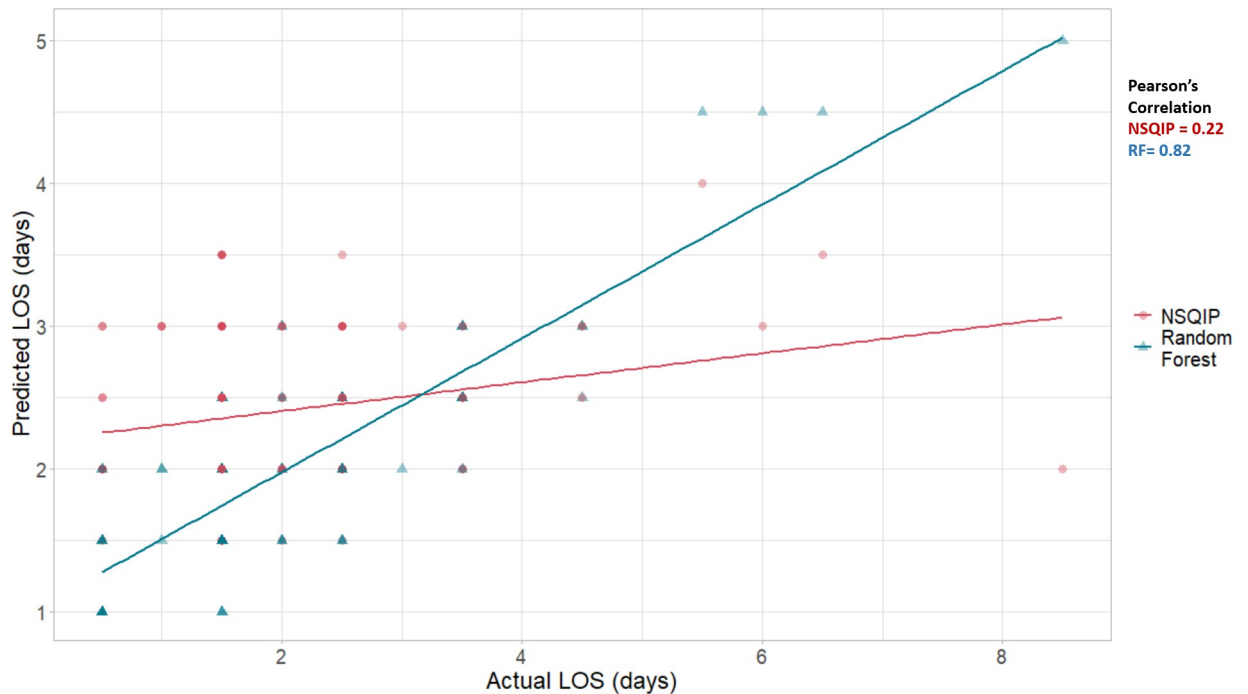**Fig 4. Comparison of NSQIP and Random Forest (RF).**

https://doi.org/10.1371/journal.pone.0277479.g004

Fig 5. Pearson's correlation for NSQIP and random forest.

Model interpretation

The PFI ranked independent variables based on their contributions to an accurate estimation of LOS using the RF training model. For example, BMI contributes most towards accurate prediction of LOS such that, if the values in BMI were randomly permuted, there will be an increase in overall RMSE by 5.1 (Fig 6). Similarly, the variables diabetes, DASI, living status, household income, ASA class, age, insurance type, and functional status were ranked most important to least important based on their average increase in RMSE after permutation (Fig 6). The PDP plots show the relationship between independent variables and the response variable (LOS) (Fig 7). More detailed explanation on their relationship is delineated in the discussion section of this paper.

Discussion

The pilot study developed an ML model to predict LOS for TJR procedures that were performed at a small-scale community hospital (bed size less than 100) located in a rural area. The developed model predicted LOS (RMSE = 0.7) more accurately than the NSQIP risk calculator (RMSE = 1.21). The NSQIP was developed using a cohort of 1.4 million patient data, which were taken from 393 health institutions in the US [39]. Sixty-nine percent of the cohort were collected from a teaching or academic affiliation and 43% of the cohort were collected from large hospitals (i.e., bed size more than 500) [39, 73]. The inaccuracy in patient risk assessment is due to the vast differences between NSQIP cohort and the population (collected from a single hospital) on which NSQIP is used [15]. For instance, the TJR patients collected from a rural community hospital may be vastly different from the NSQIP cohort that was used to build the risk calculator. Moreover, the NSQIP includes insufficient numbers of orthopedic data i.e., only 12% of orthopedic patient data contribute to the analysis [73]. These factors contribute to the NSQIP's suboptimal performance when predicting LOS for TJR procedures
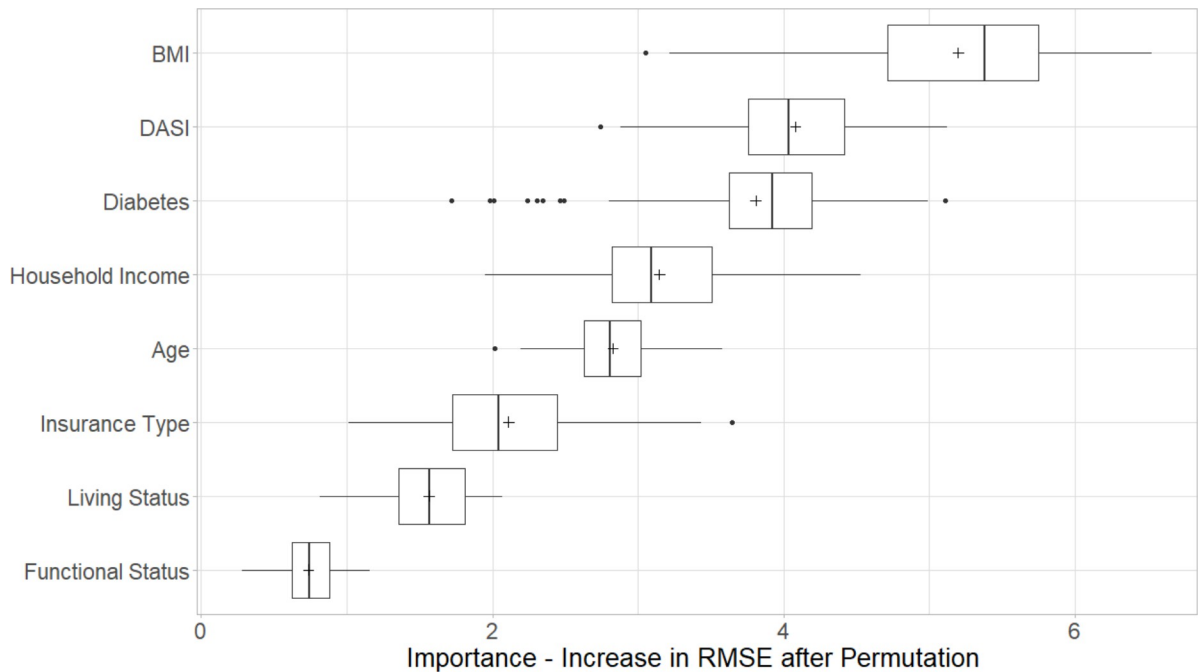
Fig 6. Variable importance using Permuted Feature Importance (PFI) method.

https://doi.org/10.1371/journal.pone.0277479.g006

especially those performed at a community hospital. Another difficulty to why NSQIP poorly predicted LOS was the adoption of the PSH system [36]. Past studies have demonstrated that the implementation of PSH has led to a reduction in hospital LOS for TJR procedures [36, 74, 75]. Thus, similar to another study, it was observed that the NSQIP predicted LOS much

Fig 7. Partial Dependence Plot of independent variables against Length of Stay (LOS).

https://doi.org/10.1371/journal.pone.0277479.g007

higher than the actual LOS for TJR procedures performed at the PSH implemented community hospital [8].

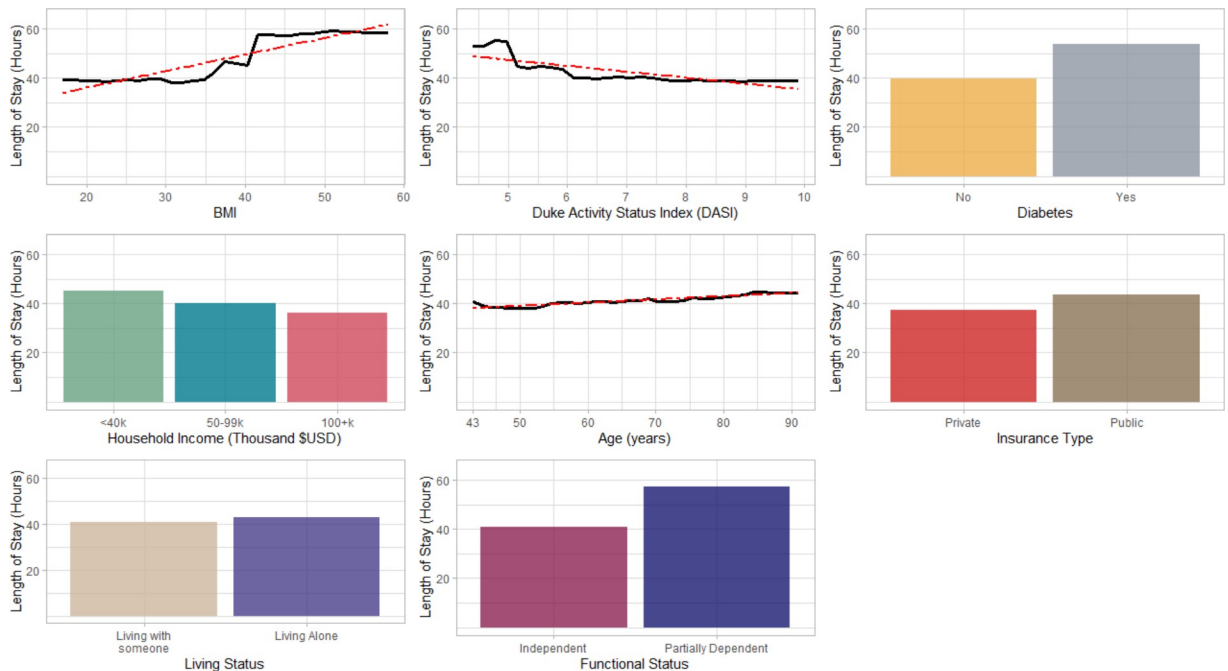To the authors' knowledge, this pilot study is the first of its kind to develop an ML model that exceeds NSQIP risk calculator in predicting a TJR surgical outcome at a community hospital and the first to predict rural patients only. The developed ML model also provides a clearer interpretation compared to the NSQIP risk calculator. The model agnostic methods, PFI and PDP plots, revealed important independent variables and their relationship with LOS. The PFI model agnostic method ranked independent variables that most contributed toward accurate prediction of LOS (Fig 5). Through this, clinicians can prioritize those factors they should address first and design a suitable intervention in the preoperative phase to optimize patients, given the severity of the condition, surgery timing, and comorbidities.

The PDP model agnostic method illustrated the relationship between the independent variables and the response variable (Fig 7). The PDP for BMI indicated that patients with higher BMI (specially more than 40 Kg/m^2) were more likely to stay longer at the hospital after the surgery compared to patients with lower BMI [76, 77]. For DASI, the LOS was found decreasing with an increase in DASI score. DASI assesses patient's heart condition and functional capacity using likert scale questionnaires on daily activities, personal care, and recreation [40]. Patients with a higher DASI score represent they are healthier and more active. Thus, in this study, it was reasonable to observe that patients with lower DASI scores (especially less than 5 METs) had longer LOS compared to patients with higher DASI scores. The PDP for diabetes showed that on average, patients with diabetes had 14 hours longer LOS compared to the patients with no diabetes [78]. The income levels in the PDP plot of household income revealed that patients with lower household income were more likely to have longer LOS than patients with higher household income (100k+). Patients with lower household income (<40K) often delay their pre-operative treatments or even postpone their surgical procedures due to their financial barriers and cost of undergoing TJR procedures. These delays increase the complications at the time of surgery requiring a longer LOS to stabilize them before discharge [42].

Concerning age, like most studies, it was observed that the LOS was higher with an increase in age [11, 79, 80]. The PDP plot for insurance type showed that on average, patients who had public insurance as their payer had six hours longer LOS than patients who had private insurance. The public insurance payers in this study were Medicare and Medicaid. Compared to private insurance patients, Medicare patients are older (aged more than 65) with increased chances of having one or two medical complications [42]. The Medicaid patients are US citizens or legal permanent residents who are mostly from a low-income background with certain disabilities, behavioral health problems, or other complications [81]. Therefore, patients with public insurance are often more medically complicated which results in them staying longer at the hospital [42].

For living status, patients who were living alone, on average stayed three hours longer than patients who were living with someone (spouse, friends, or family). This was because a majority of the patients who lived alone had lower social support causing mental health problems (such as loneliness and sadness) which influenced them to stay longer at the hospital after surgery [45]. Another reason was most patients who lived alone had to rely on the hospital to arrange for a ride before discharge. This often takes longer than patients who get picked up by their family or friends during the discharge. Finally, for functional status, despite the limited sample size in the partially dependent category (less than 3%, Table 2), it was observed that on average, partially dependent patients had 17 hours longer LOS than fully independent patients.

Akin to many studies, variables BMI [76, 77, 82], age [11, 79, 80, 83, 84], and insurance type [42, 83, 84] were found as important predictors in this study well. Conversely, unlike many

studies, the variable DASI was used in this study and was found as an important predictor of LOS. The researchers have used DASI as a preoperative assessment metric to evaluate postoperative risks, especially in colorectal surgeries [85, 86]. To the authors' knowledge, the DASI has not been used in the past TJR studies to predict or measure its association with LOS. It was also observed that the ASA had no relationship in predicting LOS in this study which in contrast, had a significant effect on LOS in other TJR studies [2, 76, 82]. Along with the ASA, other clinical variables such as smoking, depression, and hypertension that were found important in other studies did not contribute to predicting LOS in this study [11, 84]. Instead, demographic variables household income and living status were found more important. This could be due to the implementation of the PSH system which included preoperative assessment and patient education, that helped clinicians to optimize patients with higher ASA, smoking, and hypertension complications [37, 87]. Differences in demographics and factors related to rural Montana location could be another reason for finding demographic variables household income, living status more important than clinical variables ASA, smoking, and hypertension, for predicting LOS.

The need for prediction such as ML is ever more necessary in rural healthcare systems as they do not receive the same attention as urban areas [14]. This research addresses that gap by introducing ML at a rural community hospital and making the community hospital more independent instead of relying on publicly available models/methods. This pilot study is also unique by using global agnostic methods at the rural community hospital for interpretation instead of using traditional interpretable prediction models such as general linear models and decision tree [67]. Future research built from this pilot study should focus on predicting other surgical outcomes such as discharge disposition and 90-days readmission rate [4]. Also, with Medicare's recent removal of TJR surgeries from the inpatient list, future research from this pilot study should focus on developing a LOS prediction model to determine "inpatient" vs "outpatient" status for TJRs performed at rural hospitals [12].

Limitations in this pilot study include using only the RF model for prediction. Further research is on the way to applying different ML models such as Neural Networks, Boosted Trees, and Support Vector Machine (SVM) to discover how well they perform on these surgical data sets compared to RF. Second, the study used only a five-month duration (from August 2020 to December 2020) data with a very small sample size for the ML modeling. Yearly data with a high sample size (n > 1000) could have accounted for a better prediction, validation, and more robust interpretation. Third, the study was retrospective which may contain data collection biases that could alter the results and key findings [88]. Fourth, the testing data contained only 31 patients. More testing and validation data is required to confirm the developed model's validity. Finally, this study was performed at a community hospital located in a microstatistical area (with a population size less than 50,000). The results from this pilot study may not be generalizable to more rural places (e.g., with a population of less than 10,000).

Conclusion

Delivering quality surgical care to TJR patients is a challenge to many US hospitals located in rural areas. Predicting LOS in surgery is an important factor as it helps rural hospitals deliver quality surgical service, ensure patient safety, and plan for resources such as inpatient beds. This research explored how a publicly available model (NSQIP) was not an ideal model to predict LOS after a TJR procedure performed at PSH implemented community hospital. Instead, a customized machine learning model–random forest–delivered more accurate LOS predictions despite the limited data available in rural surgical systems. Further, the random forest model also provided a better interpretation by ranking the important independent variables

and illustrating its relationship against LOS. The pilot study is first of its kind to use ML at PSH incorporated rural surgical system to predict patient LOS. Results from this pilot study will contribute to helping rural surgical care by reducing LOS while improving patient satisfaction.

Acknowledgments

This research would not have been possible without the help of patients and clinicians of Bozeman Deaconess Hospital.

Author Contributions

Conceptualization: Srinivasan Sridhar.

Data curation: Srinivasan Sridhar.

Formal analysis: Srinivasan Sridhar.

Funding acquisition: Bernadette McCrory.

Investigation: Srinivasan Sridhar, Bradley Whitaker, Bernadette McCrory.

Methodology: Srinivasan Sridhar, Bradley Whitaker.

Project administration: Bradley Whitaker, Amy Mouat-Hunter, Bernadette McCrory.

Software: Srinivasan Sridhar, Bradley Whitaker.

Supervision: Bradley Whitaker, Amy Mouat-Hunter, Bernadette McCrory.

Validation: Srinivasan Sridhar, Bradley Whitaker, Amy Mouat-Hunter, Bernadette McCrory.

Visualization: Srinivasan Sridhar.

Writing – original draft: Srinivasan Sridhar.

Writing – review & editing: Bradley Whitaker, Amy Mouat-Hunter, Bernadette McCrory.

References

1. Singh JA, Yu S, Chen L, Cleveland JD. Rates of total joint replacement in the United States: future projections to 2020–2040 using the national inpatient sample. The Journal of rheumatology. 2019; 46 (9):1134–40. https://doi.org/10.3899/jrheum.170990 PMID: 30988126

2. Li G, Weng J, Xu C, Wang D, Xiong A, Zeng H. Factors associated with the length of stay in total knee arthroplasty patients with the enhanced recovery after surgery model. Journal of orthopaedic surgery and research. 2019; 14(1):343. https://doi.org/10.1186/s13018-019-1389-1 PMID: 31694690

3. Ong P, Pua Y. A prediction model for length of stay after total and unicompartmental knee replacement. The bone & joint journal. 2013; 95(11):1490–6. https://doi.org/10.1302/0301-620X.95B11.31193 PMID: 24151268

4. Alper E, O'Malley TA, Greenwald J, Aronson M, Park L. Hospital discharge and readmission. UpToDate Waltham, MA: UpToDate. 2017.

5. Tsai TC, Orav EJ, Jha AK. Patient satisfaction and quality of surgical care in US hospitals. Annals of surgery. 2015; 261(1):2. https://doi.org/10.1097/SLA.0000000000000765 PMID: 24887985

6. Weiss AJ, Elixhauser A. Overview of hospital stays in the United States, 2012: statistical brief# 180. 2014.

7. Candrilli S, Mauskopf J, editors. How much does a hospital day cost. 11th Annual International Meeting of the International Society for Pharmacoeconomics and Outcomes Research; 2006: Citeseer.

8. Carr CJ, Mears SC, Barnes CL, Stambough JB. Length of Stay After Joint Arthroplasty is Less Than Predicted Using Two Risk Calculators. The Journal of Arthroplasty. 2021. https://doi.org/10.1016/j.arth.2021.04.010 PMID: 33933330

9.   London DA, Vilensky S, O'Rourke C, Schill M, Woicehovich L, Froimson MI. Discharge disposition after joint replacement and the potential for cost savings: effect of hospital policies and surgeons. The Journal of arthroplasty. 2016; 31(4):743–8. https://doi.org/10.1016/j.arth.2015.10.014 PMID: 26725136

10. Rachoin J-S, Aplin KS, Gandhi S, Kupersmith E, Cerceo E. Impact of Length of Stay on Readmission in Hospitalized Patients. Cureus. 2020; 12(9):e10669-e. https://doi.org/10.7759/cureus.10669 PMID: 33005555.

11. Johnson DJ, Castle JP, Hartwell MJ, D'Heurle AM, Manning DW. Risk factors for greater than 24-hour length of stay after primary total knee arthroplasty. The Journal of arthroplasty. 2020; 35(3):633–7. https://doi.org/10.1016/j.arth.2019.10.037 PMID: 31757697

12. Kugelman DN, Teo G, Huang S, Doran MG, Singh V, Long WJ. A Novel Machine Learning Predictive Tool Assessing Outpatient or Inpatient Designation for Medicare Patients Undergoing Total Hip Arthroplasty. Arthroplasty Today. 2021; 8:194–9. https://doi.org/10.1016/j.artd.2021.03.001 PMID: 33937457

13. Li H, Jiao J, Zhang S, Tang H, Qu X, Yue B. Construction and comparison of predictive models for length of stay after total knee arthroplasty: regression model and machine learning analysis based on 1,826 cases in a single Singapore center. The Journal of Knee Surgery. 2020.

14. Cecchetti AA. Why introduce machine learning to rural health care. Marshall J Med. 2018; 4:02.

15. Merrill RK, Ibrahim JM, Machi AS, Raphael JS. Analysis and Review of Automated Risk Calculators Used to Predict Postoperative Complications After Orthopedic Surgery. Current Reviews in Musculoskeletal Medicine. 2020; 13(3):298. https://doi.org/10.1007/s12178-020-09632-0 PMID: 32418072

16. Goltz DE, Baumgartner BT, Politzer CS, DiLallo M, Bolognesi MP, Seyler TM. The American College of Surgeons National Surgical Quality Improvement Program surgical risk calculator has a role in predicting discharge to post-acute care in total joint arthroplasty. The Journal of arthroplasty. 2018; 33(1):25–9. https://doi.org/10.1016/j.arth.2017.08.008 PMID: 28899592

17. Medicare Cf, Services M. CY 2020 Medicare hospital outpatient prospective payment system and ambulatory surgical center payment system final rule (CMS-1717-FC). Published; 2020.

18. Iorio R, Barnes CL, Vitale MP, Huddleston JI, Haas DA. Total knee replacement: the inpatient-only list and the two midnight rule, patient impact, length of stay, compliance solutions, audits, and economic consequences. The Journal of Arthroplasty. 2020; 35(6):S28–S32.

19. Stone K, Zwiggelaar R, Jones P, Mac Parthaláin N. A systematic review of the prediction of hospital length of stay: Towards a unified framework. PLOS Digital Health. 2022; 1(4):e0000017.

20. Ogink PT, Groot OQ, Karhade AV, Bongers ME, Oner FC, Verlaan J-J, et al. Wide range of applications for machine-learning prediction models in orthopedic surgical outcome: a systematic review. Acta Orthopaedica. 2021:1–6. https://doi.org/10.1080/17453674.2021.1932928 PMID: 34109892

21. Carbonell JG, Michalski RS, Mitchell TM. An overview of machine learning. Machine learning. 1983:3–23.

22. Bhardwaj R, Nambiar AR, Dutta D, editors. A study of machine learning in healthcare. 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC); 2017: IEEE.

23. Callahan A, Shah NH. Machine learning in healthcare. Key Advances in Clinical Informatics: Elsevier; 2017. p. 279–91.

24. Galetsi P, Katsaliaki K. A review of the literature on big data analytics in healthcare. Journal of the Operational Research Society. 2020; 71(10):1511–29.

25. Srinivasan K. Essays on Digital Health and Preventive Care Analytics: The University of Arizona; 2019.

26. Khanra S, Dhir A, Islam AN, Mäntymäki M. Big data analytics in healthcare: a systematic literature review. Enterprise Information Systems. 2020; 14(7):878–912.

27. Shafqat S, Kishwer S, Rasool RU, Qadir J, Amjad T, Ahmad HF. Big data analytics enhanced healthcare systems: a review. The Journal of Supercomputing. 2020; 76(3):1754–99.

28. Maffulli N, Rodriguez HC, Stone IW, Nam A, Song A, Gupta M, et al. Artificial intelligence and machine learning in orthopedic surgery: a systematic review protocol. Journal of orthopaedic surgery and research. 2020; 15(1):1–5.

29. Lalehzarian SP, Gowd AK, Liu JN. Machine learning in orthopaedic surgery. World Journal of Orthopedics. 2021; 12(9):685. https://doi.org/10.5312/wjo.v12.i9.685 PMID: 34631452

30. Ramkumar PN, Navarro SM, Haeberle HS, Karnuta JM, Mont MA, Iannotti JP, et al. Development and validation of a machine learning algorithm after primary total hip arthroplasty: applications to length of stay and payment models. The Journal of arthroplasty. 2019; 34(4):632–7. https://doi.org/10.1016/j.arth.2018.12.030 PMID: 30665831

31. Gabriel RA, Sharma BS, Doan CN, Jiang X, Schmidt UH, Vaida F. A predictive model for determining patients not requiring prolonged hospital length of stay after elective primary total hip arthroplasty. Anesthesia & Analgesia. 2019; 129(1):43–50.

32. Han C, Liu J, Wu Y, Chong Y, Chai X, Weng X. To predict the length of hospital stay after total knee arthroplasty in an orthopedic center in China: the use of machine learning algorithms. Frontiers in Surgery. 2021; 8:606038. https://doi.org/10.3389/fsurg.2021.606038 PMID: 33777997

33. Abbas A, Mosseri J, Lex JR, Toor J, Ravi B, Khalil EB, et al. Machine learning using preoperative patient factors can predict duration of surgery and length of stay for total knee arthroplasty. International journal of medical informatics. 2021:104670. https://doi.org/10.1016/j.ijmedinf.2021.104670 PMID: 34971918

34. Klemt C, Tirumala V, Barghi A, Cohen-Levy WB, Robinson MG, Kwon Y-M. Artificial intelligence algorithms accurately predict prolonged length of stay following revision total knee arthroplasty. Knee Surgery, Sports Traumatology, Arthroscopy. 2022:1–9.

35. Lopez CD, Gazgalis A, Boddapati V, Shah RP, Cooper HJ, Geller JA. Artificial Learning and Machine Learning Decision Guidance Applications in Total Hip and Knee Arthroplasty: A Systematic Review. Arthroplasty today. 2021; 11:103–12. https://doi.org/10.1016/j.artd.2021.07.012 PMID: 34522738

36. Cline KM, Clement V, Rock-Klotz J, Kash BA, Steel C, Miller TR. Improving the cost, quality, and safety of perioperative care: A systematic review of the literature on implementation of the perioperative surgical home. J Clin Anesth. 2020; 63:109760. Epub 2020/04/15. https://doi.org/10.1016/j.jclinane.2020.109760 PMID: 32289554.

37. Kash B, Cline K, Menser T, Zhang Y. The perioperative surgical home (PSH): a comprehensive literature review for the American Society of Anesthesiologists. Schaumburg (IL): The Society. 2014.

38. Vetter TR, Goeddel LA, Boudreaux AM, Hunt TR, Jones KA, Pittet J-F. The Perioperative Surgical Home: how can it make the case so everyone wins? BMC anesthesiology. 2013; 13(1):6.

39. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. Journal of the American College of Surgeons. 2013; 217(5):833–42. e3. https://doi.org/10.1016/j.jamcollsurg.2013.07.385 PMID: 24055383

40. Hlatky MA, Boineau RE, Higginbotham MB, Lee KL, Mark DB, Califf RM, et al. A brief self-administered questionnaire to determine functional capacity (the Duke Activity Status Index). The American journal of cardiology. 1989; 64(10):651–4. https://doi.org/10.1016/0002-9149(89)90496-7 PMID: 2782256

41. Shaikh SI, Nagarekha D, Hegade G, Marutheesh M. Postoperative nausea and vomiting: A simple yet complex problem. Anesth Essays Res. 2016; 10(3):388–96. https://doi.org/10.4103/0259-1162.179310 PMID: 27746521.

42. El Bitar YF, Illingworth KD, Scaife SL, Horberg JV, Saleh KJ. Hospital length of stay following primary total knee arthroplasty: data from the nationwide inpatient sample database. The Journal of arthroplasty. 2015; 30(10):1710–5. https://doi.org/10.1016/j.arth.2015.05.003 PMID: 26009468

43. Mahajan SM, Mahajan A, Nguyen C, Bui J, Abbott BT, Osborne TF. Predictive models for identifying risk of readmission after index hospitalization for hip arthroplasty: A systematic review. Journal of orthopaedics. 2020; 22:73–85. https://doi.org/10.1016/j.jor.2020.03.045 PMID: 32280173

44. Jackson KL, Glasgow RE, Mone MC, Sheng X, Mulvihill SJ, Scaife CL. Does travel distance influence length of stay in elective pancreatic surgery? HPB. 2014; 16(6):543–9. https://doi.org/10.1111/hpb.12180 PMID: 24245982

45. Krampe H, Barth-Zoubairi A, Schnell T, Salz A-L, Kerper LF, Spies CD. Social relationship factors, preoperative depression, and hospital length of stay in surgical patients. International journal of behavioral medicine. 2018; 25(6):658–68. https://doi.org/10.1007/s12529-018-9738-8 PMID: 30105602

46. Gibson R, Erle S. Google maps hacks: " O'Reilly Media, Inc."; 2006.

47. Erekat A, Servis G, Madathil SC, Khasawneh MT. Efficient operating room planning using an ensemble learning approach to predict surgery cancellations. IISE Transactions on Healthcare Systems Engineering. 2020; 10(1):18–32.

48. Cubits MDB. Montana Demographics 2019. Available from: https://www.montana-demographics.com/.

49. Kash BA, Zhang Y, Cline KM, Menser T, Miller TR. The perioperative surgical home (PSH): a comprehensive review of US and non-US studies shows predominantly positive quality and cost outcomes. The Milbank Quarterly. 2014; 92(4):796–821. https://doi.org/10.1111/1468-0009.12093 PMID: 25492605

50. Leahy I, Johnson C, Staffa SJ, Rahbar R, Ferrari LR. Implementing a pediatric perioperative surgical home integrated care coordination pathway for laryngeal cleft repair. Anesthesia & Analgesia. 2019; 129(4):1053–60. https://doi.org/10.1213/ANE.0000000000003821 PMID: 30300182

51. Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research. 2003; 3(Mar):1157–82.

52. Khamis H. Measures of association: how to choose? Journal of Diagnostic Medical Sonography. 2008; 24(3):155–62.

53. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw. 2010; 36(11):1–13.

54. Hall MA. Correlation-based feature selection for machine learning. 1999.

55. Kursa MB, Jankowski A, Rudnicki WR. Boruta–a system for feature selection. Fundamenta Informaticae. 2010; 101(4):271–85.

56. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. Briefings in bioinformatics. 2019; 20(2):492–503. https://doi.org/10.1093/bib/bbx124 PMID: 29045534

57. Breiman L. Random forests. Machine learning. 2001; 45(1):5–32.

58. Bloomfield RA, Broberg JS, Williams HA, Lanting BA, McIsaac KA, Teeter MG. Machine learning and wearable sensors at preoperative assessments: Functional recovery prediction to set realistic expectations for knee replacements. Medical Engineering & Physics. 2021; 89:14–21. https://doi.org/10.1016/j.medengphy.2020.12.007 PMID: 33608121

59. Elfanagely O, Toyoda Y, Othman S, Mellia JA, Basta M, Liu T, et al. Machine Learning and Surgical Outcomes Prediction: A Systematic Review. Journal of Surgical Research. 2021; 264:346–61. https://doi.org/10.1016/j.jss.2021.02.045 PMID: 33848833

60. Xu H, Kinfu KA, LeVine W, Panda S, Dey J, Ainsworth M, et al. When are Deep Networks really better than Decision Forests at small sample sizes, and how? arXiv preprint arXiv:210813637. 2021.

61. Kuhn M, Silge J. Tidy Modeling with R: " O'Reilly Media, Inc."; 2022.

62. Kuhn M, Johnson K. Applied predictive modeling: Springer; 2013.

63. Probst P, Boulesteix A-L. To tune or not to tune the number of trees in random forest. J Mach Learn Res. 2017; 18(1):6673–90.

64. Berrar D. Cross-Validation. 2019.

65. Wong T-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognition. 2015; 48(9):2839–46.

66. Kuhn M, Johnson K. Feature engineering and selection: A practical approach for predictive models: CRC Press; 2019.

67. Molnar C. Interpretable machine learning: Lulu. com; 2020.

68. Haddouchi M, Berrado A, editors. A survey of methods and tools used for interpreting Random Forest. 2019 1st International Conference on Smart Systems and Data Science (ICSSD); 2019: IEEE.

69. Boehmke B, Greenwell BM. Hands-on machine learning with R: CRC Press; 2019.

70. Fisher A, Rudin C, Dominici F. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. J Mach Learn Res. 2019; 20 (177):1–81. PMID: 34335110

71. Greenwell BM, Boehmke BC, Gray B. Variable Importance Plots-An Introduction to the vip Package. R J. 2020; 12(1):343.

72. Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics. 2001:1189–232.

73. Edelstein AI, Kwasny MJ, Suleiman LI, Khakhkhar RH, Moore MA, Beal MD, et al. Can the American College of Surgeons risk calculator predict 30-day complications after knee and hip arthroplasty? The Journal of arthroplasty. 2015; 30(9):5–10. https://doi.org/10.1016/j.arth.2015.01.057 PMID: 26165953

74. Qiu C, Cannesson M, Morkos A, Nguyen VT, LaPlace D, Trivedi NS, et al. Practice and outcomes of the perioperative surgical home in a California integrated delivery system. Anesthesia & Analgesia. 2016; 123(3):597–606.

75. Alvis BD, King AB, Pandharipande PP, Weavind LM, Avila K, Leisy PJ, et al. Creation and execution of a novel anesthesia perioperative care service at a Veterans Affairs Hospital. Anesthesia & Analgesia. 2017; 125(5):1526–31.

76. Jonas SC, Smith HK, Blair PS, Dacombe P, Weale AE. Factors influencing length of stay following primary total knee replacement in a UK specialist orthopaedic centre. The Knee. 2013; 20(5):310–5. https://doi.org/10.1016/j.knee.2012.07.010 PMID: 22910196

77. Prohaska MG, Keeney BJ, Beg HA, Swarup I, Moschetti WE, Kantor SR, et al. Preoperative body mass index and physical function are associated with length of stay and facility discharge after total knee arthroplasty. The Knee. 2017; 24(3):634–40. https://doi.org/10.1016/j.knee.2017.02.005 PMID: 28336148

78. Akiboye F, Rayman G. Management of hyperglycemia and diabetes in orthopedic surgery. Current Diabetes Reports. 2017; 17(2):1–11. https://doi.org/10.1007/s11892-017-0839-6 PMID: 28265893

79. Newman JM, Szubski CR, Barsoum WK, Higuera CA, Molloy RM, Murray TG. Day of surgery affects length of stay and charges in primary total hip and knee arthroplasty. The Journal of Arthroplasty. 2017; 32(1):11–5. https://doi.org/10.1016/j.arth.2016.06.032 PMID: 27471211

80. Rozell JC, Courtney PM, Dattilo JR, Wu CH, Lee G-C. Preoperative opiate use independently predicts narcotic consumption and complications after total joint arthroplasty. The Journal of arthroplasty. 2017; 32(9):2658–62. https://doi.org/10.1016/j.arth.2017.04.002 PMID: 28478186

81. Casalino LP. Professionalism and caring for Medicaid patients—the 5% commitment? The New England journal of medicine. 2013; 369(19):1775. https://doi.org/10.1056/NEJMp1310974 PMID: 24106910

82. Sephton B, Bakhshayesh P, Edwards T, Ali A, Singh VK, Nathwani D. Predictors of extended length of stay after unicompartmental knee arthroplasty. Journal of clinical orthopaedics and trauma. 2020; 11: S239–S45. https://doi.org/10.1016/j.jcot.2019.09.009 PMID: 32189948

83. Matsen III FA, Li N, Gao H, Yuan S, Russ SM, Sampson PD. Factors affecting length of stay, readmission, and revision after shoulder arthroplasty: a population-based study. JBJS. 2015; 97(15):1255–63.

84. Oh C, Gold H, Slover J. Diagnosis of depression and other patient factors impacts length of stay after total knee arthroplasty. Arthroplasty today. 2020; 6(1):77–80. https://doi.org/10.1016/j.artd.2019.11.010 PMID: 32211480

85. El-Kefraoui C, Rajabiyazdi F, Pecorelli N, Carli F, Lee L, Feldman LS, et al. Prognostic value of the Duke Activity Status Index (DASI) in patients undergoing colorectal surgery. World Journal of Surgery. 2021; 45(12):3677–85. https://doi.org/10.1007/s00268-021-06256-4 PMID: 34448918

86. Pook M, Elhaj H, El Kefraoui C, Balvardi S, Pecorelli N, Lee L, et al. Construct validity and responsiveness of the Duke Activity Status Index (DASI) as a measure of recovery after colorectal surgery. Surgical Endoscopy. 2022:1–8. https://doi.org/10.1007/s00464-022-09145-6 PMID: 35212822

87. McCrory B, Hoge JA, Whiteley R, Wiley JB, Sridhar S, Ma J, editors. Outcomes Following Initial Perioperative Surgical Home Integration at a Rural Community Hospital. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2019: SAGE Publications Sage CA: Los Angeles, CA.

88. Hess DR. Retrospective studies and chart reviews. Respir Care. 2004; 49(10):1171–4. Epub 2004/09/28. PMID: 15447798.