Can Machine Learning Predict Extended Length of Stay After Revision Total Knee Arthroplasty? A National Database Analysis

Ashish Mittal MD[1], Anirudh Buddhiraju MD[1], Michelle Shimizu BSc[1], Tony Lin-Wei Chen MD, PhD[1], Henry Hojoon Seo BA[1], MohammadAmin RezazadehSaatlou MD[1], Ziwei Huang MD, PhD[1], Shane Fei Chen MA[1], Blake M. Bacevich BSc[1], Jona Kerluku BSc[1], Oh-Jak Kwon[1], Christian A. Pean MD[1], John G. Esposito MD[1], Young-Min Kwon MD, PhD[1]

[1]Bioengineering Laboratory, Department of Orthopaedic Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA
ymkwon@mgh.harvard.edu

Disclosures: Ashish Mittal (N), Anirudh Buddhiraju (N), Michelle Shimizu (N), Tony Lin-Wei Chen (N), Henry Hojoon Seo (N), MohammadAmin RezazadehSaatlou (N), Ziwei Huang (N), Shane Fei Chen (N), Blake M. Bacevich (N), Jona Kerluku (N), Oh-Jak Kwon (N), Christian A. Pean (N), John G. Esposito (N), Young-Min Kwon (5- MicroPort; 5- Depuy; 5- Smith & Nephew; 5- Stryker; 5- Zimmer Biomet)

INTRODUCTION: Each additional day of hospital stay can contribute to a 5-8% increase in total healthcare cost, with prolonged length of stay (LOS) increasing average hospital costs by 40%. While length of stay (LOS) can differ regionally and within different hospital systems, literature reports LOS following revision total knee arthroplasty (TKA) to be between 3.6 and 6.2 days. However, there is limited data on the weight of various factors associated with increased risk of prolonged LOS in patients undergoing revision TKA, hindering early recognition and preoperative optimization of this patient population to reduce LOS and subsequent financial burden. Machine learning has emerged as a useful tool to aid in predicting outcomes with greater accuracy than traditional statistical methods. However, previous studies only used institutional datasets for model development, potentially limiting predictive generalizability. Therefore, the aim of this study is to develop and validate four machine learning algorithms for identifying patients at risk for prolonged LOS following revision TKA using a nationally representative patient data.

METHOD: 23,656 patients undergoing revision TKA between 2013 and 2020 were identified using the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) database using common procedural terminology (CPT) codes 27486 and 27487. Patients were divided based on extended LOS as the outcome of interest, with an extended length of stay defined as greater than the 75th percentile of all LOS in the cohort in accordance with previous studies. Four ML algorithms – artificial neural network (ANN), random forest (RF), histogram-based gradient boosting (HGB), and logistic regression (LR)-were developed, optimized, and evaluated based on their (1) ability to distinguish between at-risk and not-at-risk patients, (2) accuracy, (3) calibration, and (4) clinical utility. All statistical analyses were conducted using SPSS (SPSS Version 18.0, IBM Corp., Armonk, NY, USA) and Python (Python Software Foundation, Wilmington, DE, USA).

RESULTS: The study cohort consisted of 23,656 patients, of which 5,705 (24.11%) had a prolonged LOS, while 17,951 (75.88%) did not. All four ML predictive algorithms demonstrated good discrimination, achieving similar AUCs (AUC_{LR} = AUC_{RF} = AUC_{HGB} = 0.75, AUC_{ANN} = 0.74). Likewise, calibration slopes ranged from 0.91 to 1.39, with the RF algorithm demonstrating a near-ideal slope (slope = 0.99) (Fig. 1). Calibration intercepts lay between -0.36 and -0.13, with RF and LR algorithms having the lowest intercept (Intercept: RF = -0.163; Logistic Regression= -0.139). The Brier score of all four models was similar, implying equal accuracy (Brier score: ANN = RF = LR = 0.18, HGB = 0.19) (Table 1). The most important predictors of prolonged LOS after revision TKA were operation time (>178 minutes), a preoperative diagnosis of sepsis, body mass index > 33.8 kg/m2, preoperative serum creatinine (> 1.0 mg/dL), and preoperative blood urea nitrogen (BUN) (>14.52 mg/dL) (Fig. 2).

DISCUSSION: Our findings highlight the importance of utilizing nationally representative patient data for model development. Four generalizable machine learning models developed in this study demonstrated good performance in predicting extended LOS in patients undergoing revision TKA. Prolonged operative time, preoperative sepsis, obesity, and renal disease were significant predictors of prolonged LOS following surgery, which may aid with preoperative patient counseling, effective discharge planning, care coordination, and cost containment with revision TKA.

SIGNIFICANCE/CLINICAL RELEVANCE: Machine learning algorithms developed using a nationally representative patient data demonstrated good performance in identifying patients at risk for prolonged LOS following revision TKA.

Table 1. Discrimination, calibration, and accuracy of the artificial neural network (ANN), random forest (RF), histogram-based gradient boosting (HGB), and logistic regression (LR) in predicting extended length of stay following revision total knee arthroplasty.

Metrics	ANN	RF	HGB	LR
AUC	0.74	0.75	0.75	0.75
Intercept	-0.20	-0.16	-0.36	-0.14
Slope	1.15	0.99	1.39	0.91
Brier Score	0.18	0.18	0.19	0.19

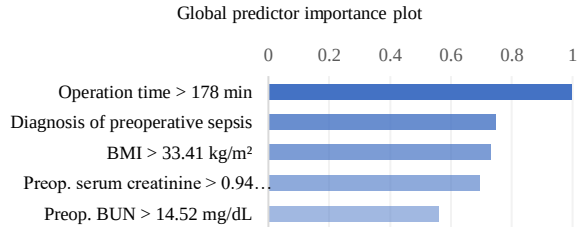

Fig. 2. Relative global predictor importance plot for the prediction of prolonged length of stay following revision total knee arthroplasty.

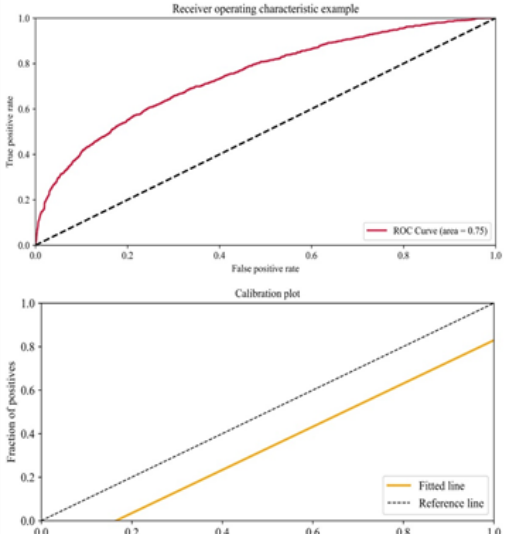

Fig. 1. Discrimination and calibration of the random forest algorithm for the prediction of prolonged length of stay following revision total knee arthroplasty.