Performance Comparison Between Machine Learning Model And The ACS Risk Calculator In Predicting Reoperation Following Primary Total Knee Arthroplasty: ACS-NSQIP National Database Analysis

Tony Lin-Wei Chen MD, PhD[1], Anirudh Buddhiraju MD[1], Henry Hojoon Seo BA[1], MohammadAmin RezazadehSaatlou MD[1], Michelle Shimizu BSc[1], Shane Fei Chen MA[1], Ziwei Huang MD, PhD[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

INTRODUCTION: Reoperation after primary total knee arthroplasty (TKA) is associated with increased healthcare costs, lengthened hospital stay, and poor patient outcomes. Prediction tools that assess the individual patient's probability of reoperation after surgery can help surgeons identify and optimize high-risk patients. Recent studies showed that machine learning models were more powerful than traditional statistic models such as risk calculators in analyzing complex interactions between risk factors for making an accurate prediction of adverse events after surgery. This study aimed to examine and compare the performance of the machine learning model and the American College of Surgeons (ACS) risk calculator in predicting reoperation after primary TKA in a national patient cohort.

METHODS: Data of 365,240 patients who underwent primary TKA between 2013 and 2020 was extracted from the ACS-NSQIP database. The dataset was streamlined to only include those patient variables that were incorporated in the ACS risk calculator and randomly split into training/testing datasets. Random forest (RF) was selected as the representative machine learning model and constructed using the training dataset. The testing dataset was applied to assess and compare the prediction performance of RF and the ACS risk calculator. Performance metrics included discrimination, calibration, and clinical utility.

RESULTS: Random forest (RF) showed a modest level of discrimination and calibration capacities (AUC: 0.68, calibration slope: 0.42, calibration intercept: 0.04, Brier score: 0.012) while outperforming the ACS risk calculator in all metrics (AUC: 0.51, calibration slope: -0.01, calibration intercept: 0.01, Brier score: 0.135). RF also demonstrated clinical utility by producing higher net benefits than the default strategies of treating no or all patients in the decision curve analysis. Age (> 67 years) and BMI (> 34 kg/m2) were the two dominant predictors of reoperation following primary TKA.

DISCUSSION: The study demonstrated the superiority of the machine learning model over the traditionally used ACS risk calculator in predicting patient-specific reoperation following TKA. Integrating additional relevant patient factors, including in-lab tests and surgical details into the algorithms would likely further enhance their predictive accuracy and effectiveness.

SIGNIFICANCE/CLINICAL RELEVANCE: The machine learning models have the potential to be a part of clinical care to improve pre-operative patient-specific stratification and optimization.

Table 1. Discrimination and calibration performance of random forest vs. the ACS risk calculator for the prediction of reoperation following primary TKA on the testing cohort.

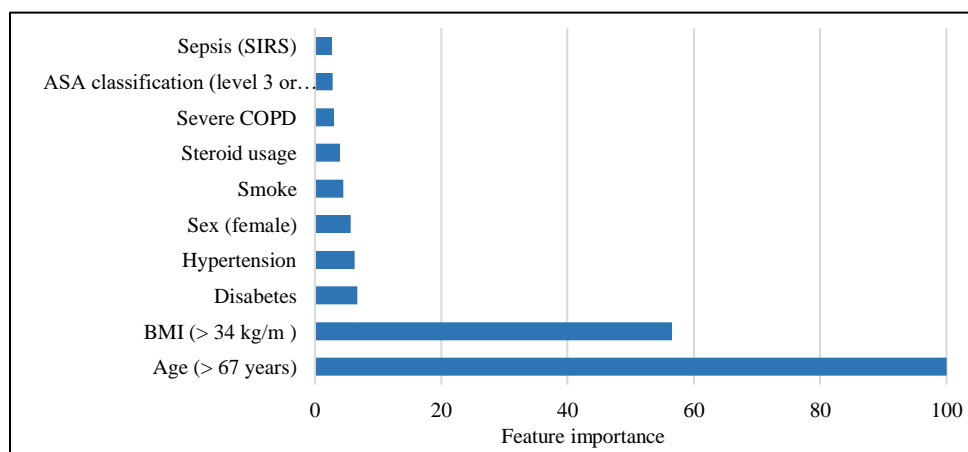| Metric	Random forest	ACS risk calculator
AUC	0.68	0.51
Calibration (slope)	0.42	-0.01
Calibration (intercept)	0.04	0.01
Brier score	0.012	0.135

Figure 1. Feature importance plot for the prediction of reoperation following primary TKA as indicated by RF.