High Specificity of Machine Learning-Based Prediction of Periprosthetic Joint Infection After Primary Total Knee Arthroplasty: Analysis of a Large National Database

Mehdi Sina Salimy[1], Anirudh Buddhiraju[2], Tony Linwei Chen, Ashish Mittal[3], Michelle Riyo Shimizu[1], Oh-Jak Kwon, Hojoon Seo, Blake Bacevich, Young-Min Kwon[4]
[1]Massachusetts General Hospital, [2]The Johns Hopkins University School of Medicine, [3]Palo Alto Medical Foundation, [4]Massachusetts General Hospital/Harvard Medical Sch

INTRODUCTION:

The volume of primary total knee arthroplasties (TKA) is projected to grow by 85% to reach 1.26 million annually by 2030, resulting in a consequent increase in the incidence of revision TKA by 182% within the next decade, with infection as the second leading indication for revision. Treating a periprosthetic joint infection (PJI) places considerable strain on patients and the hospital system due to the high incidence of morbidity and mortality, frequent reoperations, and increased cost of care. Although efforts to identify variables related to the incidence of PJI utilizing artificial intelligence and machine learning (ML) have increased recently, these studies were limited to relatively small sample sizes. Therefore, this study aimed to determine if ML models could accurately predict PJI following primary TKA using a large national database.

METHODS:

Patients who underwent primary TKA between 2013 and 2020 were identified by Current Procedural Terminology (CPT) code 27447 from the American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP) database, which consists of preoperative and procedure-related variables and outcomes from more than 600 hospitals in the United States. All patients over the age of 18 were included in this study. The primary outcome of interest was the development of a deep surgical site infection (SSI), referred to as a PJI, as defined by the Centers for Disease Control and Prevention (CDC) National Healthcare Safety Network (NHSN to include deep incisional and organ/space related outcomes. Data collected included demographic, preoperative, intraoperative, and immediate postoperative outcomes. Two ML models were then used to predict PJI following primary TKA—histogram-based gradient boosting (HGB) and random forest (RF), both of which are conventionally used algorithms for predictive modeling. Receiver operating characteristic (ROC) curves and the corresponding area under the curve (AUC) were determined to assess the discriminatory ability of each ML model. The ML models were then calibrated and assessed using calibration plots with corresponding slopes (target value of 1, with slopes less than 1 showing high variance, and slopes greater than 1 showing little variance), calibration intercepts (target value of 0, with negative values suggesting overestimation and positive values suggesting underestimation), and Brier scores (perfect predicted probability is indicated by a Brier score of 0, compared to the no predictability score of 1).

RESULTS:

After applying exclusion criteria, the final cohort consisted of 77,948 patients, with a PJI prevalence of 0.3%. Both ML predictive algorithms demonstrated excellent discriminatory ability for identifying PJIs, as evidenced by their high AUCs (AUC_{HGB}: 0.93; AUC_{RF}: 0.95). Similarly, the calibration curves ($Slope_{HGB}$: 0.88; $Slope_{RF}$: 1.04), calibration intercepts ($Intercept_{HGB}$: 0.28; $Intercept_{RF}$: 0.22), and Brier scores (Brier $score_{HGB}$: 0.006; Brier $score_{RF}$: 0.006) all demonstrate model validation (Figure 1). The sensitivity and specificity were 36.2% (95% confidence interval (CI): 28.6%-44.4%) and 99.9% (95% CI: 99.9%-100%) for the HGB model and were 23.0% (95% CI: 16.6%-30.5%) and 99.9% (95% CI: 99.9%-100%) for the RF model, respectively (Figure 1). The top five predictors of the development of a PJI following primary TKA were the length of stay (> 3 days), preoperative INR (> 1.04), preoperative albumin (< 4.06 g/dL), preoperative sodium (< 139.52 mEq/L), and preoperative platelet count (< 241,500/mm^3).

DISCUSSION AND CONCLUSION: The two machine learning models developed in this study demonstrated excellent performance for predicting PJI following primary TKA. The top three predictors for PJI were the length of stay after surgery, preoperative INR, and preoperative albumin, showing that both preoperative and postoperative factors influence infection likelihood. Given the model's high-performance metrics, strong considerations may be placed on the aforementioned preoperative laboratory tests to identify patients with clotting disorders or malnutrition at risk of developing an infection. The high specificity also shows promise as an adjunct for clinicians in confirming the diagnosis of PJI for patients who may present with a mixed clinical picture of infection.

Fig 1. Receiver operating characteristic curve and calibration plot of the histogram-gradient boosting (left) and random forest (right) algorithms in predicting periprosthetic joint infection following primary total knee arthroplasty.