

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

Mehdi S. Salimy BSc¹, Anirudh Buddhiraju MD¹, Tony-Lin Wei Chen MD, PhD¹, Ashish Mittal MD¹, Michelle Shimizu BSc¹, Henry Hojoon Seo BA¹, Blake M. Bacevich BSc¹, Shane Fei Chen MA¹, Jona Kerluku BSc¹, Oh-Jak Kwon¹, Ziwei Huang MD, PhD¹, Christian A. Pean MD¹, John G. Esposito MD¹, Young-Min Kwon MD, PhD¹

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

Disclosures: Mehdi S. Salimy (N), Anirudh Buddhiraju (N), Tony Lin-Wei Chen (N), Ashish Mittal (N), Michelle Riyo Shimizu (N), Henry Hojoon Seo (N), Blake M. Bacevich (N), Shane Fei Chen (N), Jona Kerluku (N), Oh-Jak Kwon (N), Ziwei Huang (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: Periprosthetic joint infections (PJI) are the second and third leading indication for revision knee and hip total joint arthroplasty (TJA). Treating a 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 associated with institutional databases. Therefore, this study aimed to determine if ML models could accurately predict PJI following primary TJA using a large national database.

METHODS: Patients who underwent primary TJA between 2013 and 2020 were identified by Current Procedural Terminology (CPT) codes 27130 and 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 PJI. Data collected included demographic, preoperative, intraoperative, and immediate postoperative outcomes. Histogram-based gradient boosting (HGB) and random forest (RF) were used to predict PJI after primary total knee arthroplasty (TKA), and a separate HGB model was developed for primary total hip arthroplasty (THA). 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 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: The final cohort comprised 77,948 TKA and 51,053 THA patients, with a PJI prevalence of 0.3% and 0.5%, respectively. HGB demonstrated excellent discriminatory ability for identifying PJIs following TJA, as evidenced by their high AUCs (TKA: 0.93; THA: 0.88). Calibration curve slopes (TKA: 0.88; THA: 0.79), intercept (TKA: 0.28, THA: 0.32), and Brier scores (TKA: 0.006, THA: 0.007) demonstrated model validation. The sensitivity and specificity of the HGB models were 36.2% (95% CI: 28.6%-44.4%) and 99.9% (95% CI: 99.9%-100%) for TKA, and 18.6% (11.4%-27.7%) and 99.9% (99.9%-100%) for THA, respectively. Test-specific metrics all supported RF performance and validation in predicting PJI after TKA, as shown by the AUC (0.95), calibration slope (1.04), intercept (0.22), and Brier score (0.006). The sensitivity and specificity of the RF model were 23.0% (95% CI: 16.6%-30.5%) and 99.9% (95% CI: 99.9%-100%). The top predictors shared among THA and TKA models were the length of stay (>3.3 vs. 3 days), preoperative sodium (<139.18 vs. 139.52 mEq/L), and preoperative platelet count (<249,260 vs. 241,500/mm³). In addition, preoperative INR (>1.04) and albumin (<4.06g/dL) were influential factors in the development of a PJI following primary TKA. Patients' weight (> 202.8 lbs) and ASA class (4) were among the top predictors following primary THA.

DISCUSSION: The two machine learning models developed demonstrated excellent performance for predicting PJI following primary TJA. The important predictors for PJI following both TKA and THA were the length of stay, preoperative sodium and platelet count, showing that both preoperative and postoperative factors influence the patient-specific risk of infection. The high performance of machine learning models supports considerations to utilize preoperative laboratory tests to identify patients with clotting disorders or malnutrition at risk of developing a PJI following primary TJA.

SIGNIFICANCE/CLINICAL RELEVANCE: The models' high specificity 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.

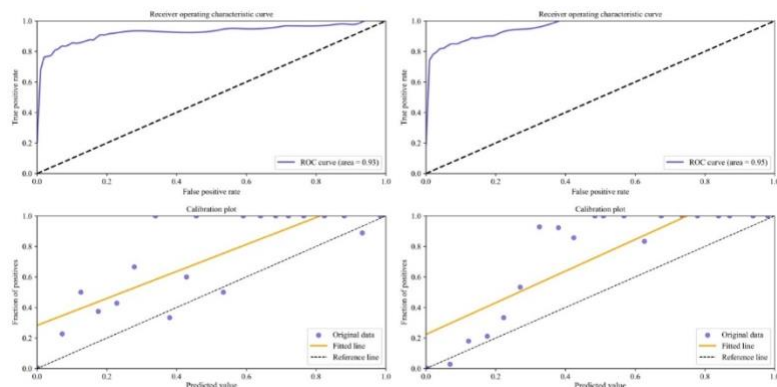


Fig 1. Receiver operating characteristic curve and calibration plot of the HGB (left) and RF (right) algorithms in predicting PJI following primary TKA

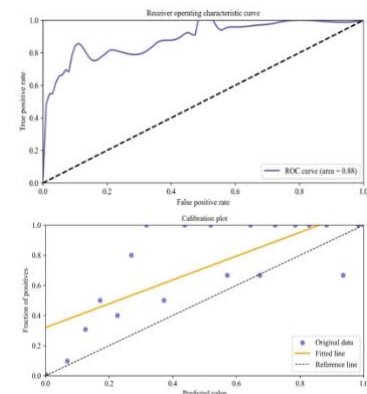


Fig 2. Receiver operating characteristic curve and calibration plot of the HGB algorithm in prediction PJI following primary THA