Machine Learning Algorithms are Superior to Traditional Risk-Assessment Indices in Predicting 30-Day Mortality Following Revision Total Hip or Total Knee Arthroplasty: Analysis of 22,836 Patients in a National Cohort Database

Christian A. Pean MD[1], Anirudh Buddhiraju MD[1], Tony Lin-Wei Chen MD, PhD[1], Henry Hojoon Seo BA[1], Michelle Shimizu BSc[1], MohammadAmin RezazadehSaatlou MD[1], Shane Fei Chen MA[1], Ziwei Huang MD, PhD[1], Blake M. Bacevich BSc[1], Oh-Jak Kwon[1], Jona Kerluku BSc[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: Christian A. Pean (N), Anirudh Buddhiraju (N), Tony Lin-Wei Chen (N), Henry Hojoon Seo (N), Michelle Shimizu (N), MohammadAmin RezazadehSaatlou (N), Shane Fei Chen (N), Ziwei Huang (N), Blake M. Bacevich (N), Oh-Jak Kwon (N), Jona Kerluku (N), John G. Esposito (N), Young-Min Kwon (5- MicroPort; 5- Depuy; 5- Smith & Nephew; 5- Stryker; 5- Zimmer Biomet)

INTRODUCTION: Machine learning (ML) has emerged as a promising technology for patient risk stratification in orthopaedic surgery. Although risk calculators and scores are currently used to prognosticate postoperative outcomes following revision total hip and knee joint arthroplasty (TJA), machine learning-based predictive tools demonstrate the potential to replace traditional risk-assessment indices in the clinical setting. Given the anticipated rise in demand for revision TJA, these innovative predictive tools have the potential to aid clinicians in identifying patients who may be at higher risk of experiencing adverse events, including mortality. Revision TJA has up to a fourfold higher risk of mortality when compared with primary TJA. As such, accurate identification of patients at higher risk could enhance the preoperative shared decision-making process. Therefore, this study aimed to evaluate the predictive ability of three ML models for risk-assessment indices to that of the traditionally used risk-assessment indices CARDE-B score, 5-item (5MFI), and 6-item modified frailty index (6MFI), using data from a large national database.

METHODS: A total of 22,836 patients over the age of 18 years who underwent revision THA and TKA between 2013 and 2020 in the United States were identified from the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) database to comprise the overall cohort. This patient cohort was subsequently split 80:20 in a random fashion to compose the training and validation cohorts. Three ML models: extreme gradient boosting (XGB), random forest (RF), and elastic-net penalized logistic regression (NEPLR), were developed and evaluated based on their discrimination, calibration, and accuracy. Subsequently, discrimination of CARDE-B, 5MFI, and 6MFI scores was assessed using the area under the receiver operating characteristic curves and compared to that of the ML models.

RESULTS: Risk-assessment index CARDE-B had the highest discrimination among the scores (AUC= 0.89), followed by 6MFI (AUC= 0.80) and 5MFI (AUC= 0.68) (**Fig 1**). Both ML models were equally accurate (Brier score: XGB=RF=NEPLR= 0.005) and demonstrated outstanding discrimination with similar areas (AUC) under the receiver operating characteristic curve (XGB= 0.94, RF= 0.93, NEPLR= 0.93). NEPLR was the best-calibrated model overall (slope= 0.54, intercept= -0.004) (**Table 1**) (**Fig. 2**). Hypoalbuminemia (albumin <3.5mg/dL) was found to be the most important predictor of 30-day mortality following revision TJA.

DISCUSSION: ML models outperform traditional risk-assessment indices such as CARDE-B, 5MFI, and 6MFI in predicting postoperative 30-day mortality after revision total hip and knee joint arthroplasty. The identification of hypoalbuminemia, BMI, and WBC count as prognostic markers of 30-day mortality following surgery may allow improved perioperative optimization strategies to mitigate patient-specific adverse outcomes following revision TJA.

SIGNIFICANCE/CLINICAL RELEVANCE: Our study provides supportive evidence to the superior utility of machine learning models compared to the traditionally used risk-assessment indices for predicting patient-specific risk stratification in a large national database analysis.

Table 1. Discrimination and calibration of extreme gradient boosting (XGB), elastic-net penalized logistic regression (NEPLR), and random forest (RF) algorithms in predicting 30-day mortality after revision total hip or knee joint arthroplasty.

Parameter	XGB	NEPLR	RF
Area under the curve (AUC)	0.94	0.93	0.93
Calibration (slope)	0.37	0.54	0.24
Calibration (intercept)	0.08	-0.004	0.04
Brier score	0.005	0.005	0.005

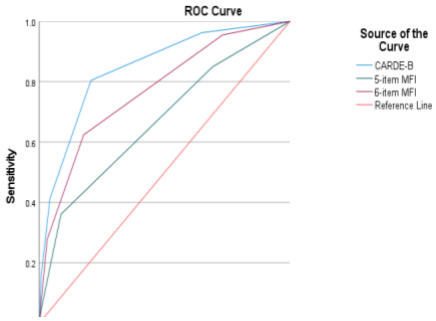

Fig 1. Receiver operating characteristic curves of all three scores (CARDE-B, 5-item, and 6-item Modified Frailty Index) in predicting 30-day mortality after revision total hip or knee arthroplasty.

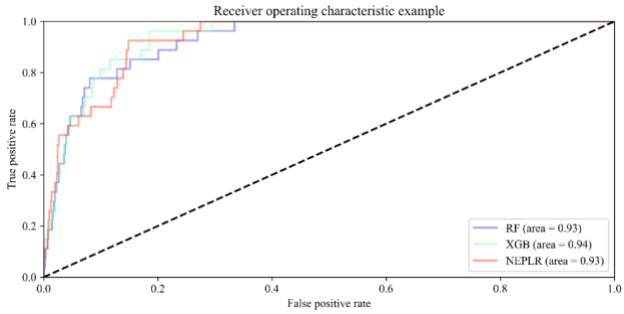

Fig 2. Receiver operating characteristic curves of all three ML models (extreme gradient boosting (XGB), random forest (RF), and elastic-net penalized logistic regression (NEPLR)) for the assessment of the discrimination ability of all three ML models in predicting 30-day mortality after revision total hip or knee arthroplasty.