

Using Artificial Intelligence to Predict Risk in Hip and Knee Replacement Surgery: A Two-Centre Study Evaluating the Role of High-Volume Low-Complexity Hubs and Ambulatory Surgical Centres

Ana Rodrigues^{1*}, Tiago Martins¹, Bruno Lopes²

¹ Department of Orthopedic Surgery and AI Risk Prediction, Faculty of Medicine, University of Lisbon, Lisbon, Portugal.

² Department of Surgical Outcomes and Health Services, Faculty of Medicine, University of Porto, Porto, Portugal.

*E-mail ✉ ana.rodrigues@med.ulisboa.pt

Abstract

The COVID-19 pandemic created a major backlog of hip and knee replacement procedures across the United Kingdom. To tackle this challenge, surgical hubs have been proposed to boost efficiency, especially for routine, lower-risk operations. These hubs and similar Ambulatory Surgery Centers typically lack advanced critical care support, making them best suited for patients with fewer comorbidities. Accurate preoperative risk assessment is therefore essential to direct patients to the right facility and minimize avoidable complications. This study investigated the application of artificial intelligence (AI) for risk stratification in hip and knee arthroplasty. A polynomial regression model was created using patient demographics, laboratory results, and comorbidities to generate individualized risk scores for postoperative complications. The model was trained on 29,658 patient records drawn from two UK National Health Service (NHS) organizations. Performance was measured using the area under the receiver operating characteristic curve (AUROC), enabling classification of patients into high- and low-risk groups. Validation involved a retrospective review of 445 cases, comparing predicted complications and need for additional care against actual outcomes. The model achieved 70% sensitivity in detecting high-risk patients and a negative predictive value of 96%. Risk predictions from the AI tool proved comparable to those made during traditional consultant-led preoperative evaluations. These results indicate that AI can facilitate more efficient and consistent preoperative risk assessment, helping to ease pressure on assessment teams and improve resource use. Despite certain limitations, the model serves as a valuable support tool for clinical decision-making in determining patient risk. It has potential applications in UK NHS surgical hubs and in Ambulatory Surgery Centers in the United States.

Keywords: Artificial intelligence, Risk stratification, Hip and knee arthroplasty, Preoperative risk assessment, Postoperative complications, Surgical hubs

Introduction

The United Kingdom continues to face a substantial backlog of hip and knee replacement surgeries following the COVID-19 pandemic [1-3]. Meeting this demand will likely require a major, sustained increase in capacity,

which surgical hubs are well-positioned to deliver [4, 5]. Compared with conventional NHS settings, established hubs can handle 11% more cases, while new hubs can deliver up to 22% additional throughput. A surgical hub functions as a dedicated, ring-fenced elective care facility. Because these centers often lack on-site intensive care or high-dependency units, they are most appropriate for high-volume, low-complexity procedures in patients with limited comorbidities.

In contrast, hospitals equipped with intensive care facilities are designated for low-volume, high-complexity (LVHC) cases [6]. In the United States, Ambulatory Surgery Centers (ASCs) have grown

Access this article online

<https://smerpub.com/>

Received: 19 March 2025; Accepted: 05 June 2025

Copyright CC BY-NC-SA 4.0

How to cite this article: Rodrigues A, Martins T, Lopes B. Using Artificial Intelligence to Predict Risk in Hip and Knee Replacement Surgery: A Two-Centre Study Evaluating the Role of High-Volume Low-Complexity Hubs and Ambulatory Surgical Centers. J Med Sci Interdiscip Res. 2025;5(1):179-88. <https://doi.org/10.51847/RNGjE698si>

rapidly, with 84% growth in activity reported in 2022, driven by economic and regulatory factors. Both hubs and ASCs share the challenge of selecting patients appropriately. Unexpected complications leading to emergency transfers from these facilities carry high clinical and financial costs. Consequently, thorough preoperative risk assessment and optimization remain resource-intensive tasks for preoperative teams. Artificial intelligence (AI) has been proposed as a potential solution [7].

The adoption of AI in medicine and surgery is expanding, with growing interest in orthopedics for applications such as risk prediction [7, 8]. While traditional surgical risk tools exist [9, 10], AI and machine learning approaches provide more advanced capabilities [11] and have shown success in other medical fields [12, 13]. Machine learning models incorporating comorbidities, demographic details, and socioeconomic data have been used to develop electronic risk calculators that support patient counseling and outcome forecasting [14]. These algorithms are particularly effective at modeling complex, non-linear relationships in clinical data [15]. Such prediction tools help set realistic expectations, anticipate complications, and guide perioperative management [16]. Earlier work using adaptations of the P-POSSUM model demonstrated reasonable accuracy for morbidity and mortality after hip surgery [17, 18]. However, limited sample sizes and study designs led to mixed conclusions.

In the first part of this study, we developed a polynomial regression prediction model using patient demographics, blood test results, and comorbidities sourced from two NHS trusts. The model assigns each patient a percentage risk score for postoperative complications. This tool aims to streamline preoperative assessment, guide site allocation for theatre lists and hospital capacity, support targeted risk mitigation strategies, and empower patients to make informed decisions.

In the second part, we conducted a validation study using retrospective hip and knee arthroplasty data from Swansea Bay University Health Board and Northumbria Healthcare NHS Foundation Trust. The goal was to determine how accurately the AI model identified patients as “high risk,” with sensitivity and negative predictive value (NPV) serving as key performance measures.

Overall, the objective was to evaluate the practicality of using this model for prospective risk prediction in larger

populations as part of its ongoing development. Such a system could help perioperative teams make more efficient, evidence-based decisions.

Materials and Methods

Part 1: Model development

This work followed the TRIPOD guidelines to ensure transparent reporting of the machine learning model, in line with recommended standards for prediction model development and validation.

The dataset comprised 29,658 records of elective primary total joint replacement patients treated at the two participating NHS organizations between January 2014 and March 2020 (Northumbria Healthcare NHS Foundation Trust: $n = 15,022$; Swansea Bay University Health Board: $n = 14,636$).

Twenty-three predictor variables were included, categorized into patient demographics, comorbidities, and blood test results. Cases were selected based on specific OPCS4 procedure codes for primary hip replacement (W371, W381, W391, W931, W941, W951) and primary total knee replacement (W401, W411, W421, O181). Knee procedures accounted for 56.6% of the data, and hip procedures for 43.4%. Non-numeric entries were removed, and all data were converted to numeric format suitable for modeling. Duplicate records were eliminated, yielding a final cleaned dataset of 21,527 cases (533,546 data points). This was randomly divided into an 80% training set and a 20% test set using stratified sampling.

Sample size adequacy was confirmed via a reconstructed pmsamp method in Python, based on a 6.8% outcome prevalence, 23 predictors, and an anticipated R^2 of 0.5, which indicated that a minimum of 920 training records and 71 events per comorbidity were required. Learning curve analysis for bias-variance trade-off showed that performance stabilized with a minimum training set of 9,527 records, after which further data provided diminishing improvements in test AUROC.

A polynomial logistic regression model was initially developed at Northumbria Healthcare NHS Foundation Trust and subsequently refined in partnership with Swansea Bay University Health Board. The model integrated patient demographics, comorbidities, blood results, and general health indicators (**Table 1**). Relevant postoperative complications were defined for outcome prediction (**Table 2**).

Table 1. Baseline characteristics, clinical profiles, laboratory parameters, and postoperative outcomes of the study cohort (n = 6730). Continuous variables are presented as median values with interquartile ranges (IQR), whereas categorical variables are reported as percentages.

Category	Parameter	Summary value
Patient demographics	Age (years), median (IQR)	70 (64–77)
	Female participants	57.0%
	Male participants	43.0%
	Current smoker	6.6%
Procedure characteristics	Hip replacement surgery	43.0%
	Knee replacement surgery	57.0%
	ASA physical status classification, median (IQR)	2 (1–2)
Underlying medical conditions	Hypertension	51.1%
	Type 2 diabetes mellitus (non-insulin treated)	12.7%
	Hypothyroidism	9.0%
	Ischaemic heart disease	8.4%
	Elevated cholesterol levels (hypercholesterolemia)	7.5%
	Atrial fibrillation	6.8%
	Chronic obstructive pulmonary disease (COPD)	4.8%
	Previous circulatory system disease	4.5%
	Rheumatoid arthritis	2.1%
	Psoriatic arthritis	0.6%
	Hyperthyroidism	0.3%
	Insulin-treated diabetes mellitus	0.3%
	Pressure ulcers/sores	0.2%
	Alzheimer's disease	0.1%
	Dementia	0.1%
Laboratory findings	Serum sodium, median (IQR)	138.0 (137.0–140.0)
	Creatinine, median (IQR)	84.0 (74.0–96.0)
	Hemoglobin, median (IQR)	137.0 (128.0–146.0)
	Platelet count, median (IQR)	237.5 (200.0–280.75)
	Urea, median (IQR)	6.0 (5.0–7.2)
	White blood cell count, median (IQR)	7.1 (5.9–8.5)
Post-surgical outcomes	Patients experiencing complications	7.4%
	Mortality within 90 days after surgery	0.3%

Table 2. Adverse medical events incorporated into the predictive modeling framework.

Time frame after surgery	Complications evaluated as adverse outcomes
During the index hospital episode	Aspiration pneumonia
	<i>Clostridioides difficile</i> infection
Within 7 days following surgery	Myocardial infarction
	Acute renal failure
	Stroke
Within 30 days following surgery	Transient ischaemic attack
	Gastrointestinal haemorrhage
	Ileus
	Myocardial infarction
	Pneumonia

	Renal failure
	Pancreatitis
	Stroke
	Thrombocytopenia
	Transient ischaemic attack
	Urinary retention
	Urinary tract infection
Within 60 days following surgery	Deep vein thrombosis
	Pulmonary embolism
	Mortality

Information on each patient was gathered from prior assessments covering comorbidities, demographic details, laboratory results, and interactions with other specialist services in the outpatient environment (**Table 1**).

Accounting for changes in practice

Modeling faces challenges such as data leakage, in which information that is no longer representative of current clinical practice influences the model. Examples include shifts in service delivery (such as centralizing complex care at specialist centers) and updates to clinical coding practices. These factors were considered to keep the model relevant to contemporary standards.

To reduce this risk and obtain a more accurate picture of real-world performance, the dataset was divided chronologically. The training set consisted of patients who had surgery before August 2017 ($n = 6,730$), while the test set included those treated after that date, representing 18.5% of the total data points ($n = 43,949$). Some patients had multiple entries because they underwent several joint procedures. To avoid leakage, we completely excluded any patient present in the training set from the test set. This produced a final test set of 2,407 patient records, split by procedure type: 1,035 total knee replacements (43%) and 1,372 total hip replacements (57%).

The area under the receiver operating characteristic curve (AUROC) served as the primary metric for selecting the optimal hyperparameters for the model architecture. AUROC is particularly appropriate for risk stratification because it assesses the model's ability to distinguish between patients who experience complications and those who do not, independent of a specific decision threshold.

Given the data's imbalance, the model was trained both with and without oversampling. Oversampling helps

address class imbalance by generating additional instances of the minority class, creating a more even distribution of outcomes. In this study, oversampling was applied to all minority classes with event rates below 50% (**Table 1**), producing a balanced training dataset. This step was tested to determine whether strengthening the representation of less common events could enhance the model's ability to predict postoperative complications.

Model performance was assessed in three main ways: first, discrimination between outcome classes using AUROC (patients with versus without postoperative adverse events); second, prediction confidence via the Brier score, which calculates the average squared difference between predicted probabilities and actual results; and third, calibration quality using the Expected Calibration Error (ECE), which quantifies the discrepancy between predicted and observed probabilities across different probability ranges.

To check for potential biases, the model's performance was examined across demographic subgroups. This involved comparing AUROC and Brier scores by age groups, LSOA deprivation deciles, and sex. Evaluation across ethnic subgroups was not feasible because ethnicity data were insufficiently recorded.

For a risk stratification tool, focusing on probabilities is critical as it enables quantification of individual complication risk. For practical clinical use, these probabilities were mapped to categorical risk levels. This allows clinicians to prioritize high-risk patients for closer monitoring or additional input while avoiding unnecessary concern for low-risk individuals. The following thresholds were applied:

- Low risk of adverse outcome: probability 0.00–0.06
- Moderate risk of adverse outcome: probability 0.06–0.075

- High risk of adverse outcome: probability 0.075 and above

Risk ratio thresholds

The cut-offs used to separate low-risk from moderate- to high-risk patients were chosen to achieve a clinically meaningful reduction in relative risk compared with baseline, in keeping with earlier research supporting threshold selection based on relative risk reduction to improve patient management [19, 20].

Thresholds were derived by examining complication data (**Table 1**) from the Northumbria cohort (143 patients), which had an overall complication rate of 7.5%. A 6% probability threshold was selected pragmatically, corresponding to a 20% relative risk reduction from the baseline 7.5% rate and offering a practical cut-off for elective bed planning.

Under this threshold, 106 patients (75%) were classified as low risk. When scaled to each organization's annual caseload, this proportion aligned well with the approximately 1,800 elective operating sessions (out of 2,400) available per year at the high-volume, low-complexity (HVLC) elective sites (hubs).

Patients with a predicted complication probability below 6% were categorized as low risk, while those with a probability of 6% or higher were placed in the moderate-to-high-risk group.

Part 2: Model evaluation and validation

The model was evaluated through retrospective review of 445 independent patient records from the two participating healthcare organizations. These patients had undergone elective primary hip and knee replacement procedures, identified using ICD-10 and OPCS4 clinical coding systems.

For each patient, the model processed the preoperative electronic health record to generate a risk percentage of developing a complication. This prediction was made without access to postoperative outcome information (blinded assessment).

Postoperative outcomes at 90 days were extracted from patient records and electronic coding, then compared with the model's predicted risk percentage. Complications included prolonged hospital stay beyond the expected duration, a broad range of postoperative medical issues, or death (**Table 2**).

At both the Northumbria and Swansea sites, the AI model's classification of patients into low-, moderate-, and high-risk categories was compared with assessments

made by consultant anesthetists and preoperative assessment teams. This allowed direct comparison between the AI tool and standard consultant-led care.

Results and Discussion

The AI model generated risk ratios for 445 patients: 302 at Swansea and 143 at Northumbria. Patients were grouped into either high/moderate-risk or low-risk categories.

Overall, the model classified 192 patients as high/moderate risk (risk ratio > 6%) and 253 patients as low risk (risk ratio < 6%).

The distribution of risk categories differed between the two sites (**Table 3**).

Table 3. Percentage of patients in each risk group; Swansea Bay vs Northumbria Healthcare.

Empty Cell	Swansea	Northumbria
Low Risk	52 %	32 %
High/Moderate	48 %	68 %

A total of 37 patients across both sites experienced postoperative complications.

Of these 37 patients with complications, the AI model correctly flagged 26 as "high risk." The model incorrectly classified the remaining 11 patients who developed complications as "low risk." This resulted in an AI model sensitivity of 70% for accurately identifying the appropriate risk category based on actual complications.

The Negative Predictive Value (NPV), which reflects the model's effectiveness in correctly identifying patients who were not at high risk and did not experience a postoperative complication, was 96%. Overall complication data are presented (**Table 4**) and examined in greater detail in the results matrix (**Table 5**).

Table 4. Complications by AI risk category.

Empty cell	No complication	Complication	Totals
Low risk	242	11	253
High/Moderate risk	166	26	192
Totals	408	37	445

Negative Predictive value (NPV): 96%. Sensitivity: 70%.

Table 5. Consultant (Attending) led care risk prediction in Northumbria.

Empty cell	No complication	Complication	Totals
------------	-----------------	--------------	--------

Low risk	102	3	105
High/Moderate risk	31	7	38
Totals	133	10	143

There was one recorded death, occurring in a patient classified by the model as high-risk with a predicted risk score of 34.4%.

The model's performance was next compared against standard consultant (attending) anesthetist-led care using a subset of 143 patients from Northumbria. A consultant anesthetist independently assigned each case to a high/moderate-risk or low-risk category. These

categorizations were then descriptively compared with the AI model's outputs (**Tables 5 and 6**).

Differences in sensitivity were minimal, with both consultant-led care and the AI model achieving 70% sensitivity. Negative Predictive Value was slightly higher for consultant assessment (97%) than for the AI model (96%). When the model's predictions were evaluated against consultant-led opinions, it showed strong overall agreement (**Table 6**). The AI tool matched 84% (n = 32) of the patients identified by the pre-assessment teams as high or moderately high risk (n = 38), and 92% of those judged to be lower risk and appropriate for high-volume, low-complexity pathways.

Table 6. Comparing consultant (attending) review and identification of risk category to AI prediction in the Northumbria cohort.

N = 143	True risk category (according to complication event)	Consultant-led care (CLC) categorization	AI model categorization in agreement with CLC
Complication in high-risk group	7	38	32 (84%)
Complications in low-risk group	3	105	97 (92%)
Model Sensitivity			71%
Model Specificity			68%
Model NPV			97.1% (92.9%, 98.8%)

Complications were examined in more detail using the Clavien-Dindo index [21] to evaluate the severity of adverse events across high-, moderate-, and low-risk patient groups (**Table 7**).

The Clavien-Dindo classification provides a standardized method for grading the severity of postoperative complications (PPCs) and is widely used in surgical research to ensure consistent reporting. It categorizes

complications according to the treatment required to manage them. Grades range from mild events requiring only minor intervention (Grade I) to infections managed with antibiotics (Grade II), to those requiring surgical intervention (Grade IIIb), to death (Grade V).

More severe complications tended to occur among patients assigned to higher risk categories (**Figure 1**).

Table 7. Outcome of risk classification compared with severity score based on Clavien-Dindo (CD) classification.

Empty cell	Severity (CD Classification)					
Empty cell	CD I	CD II	CD IIIa	CD IIIb	CD IV	CD V
Low	8	2	1			
Moderate	2	3	2			
High	6	6	1	2	4	

		AI Prediction			
		Total population=445	High Risk	Low risk	Accuracy (TP+TN)/pop = $(26+242)/445$ =60%
True Risk Category (had complication or not)	True high Risk (AP) Complication =37	True Positive (TP)=26	False Negative (FN)=11	True Positive Rate (TPR) SENSITIVITY =TP/AP =26/37 =70%	False Negative Rate (FNR) =FN/AP =11/37 =30%
	True low risk (AN) No complication =408	False Positive (FP)=166	True Negative (TN)=242	False Positive Rate (FPR) =FP/AN =166/408 =41%	True Negative Rate (TNR) SPECIFICITY =TN/AN =242/408 =59%
		Prevalence =AP/pop =37/445 =8%	PPV =TP/(TP+FP) =26/(26+166) =13%	False Omission Rate (FOR) =FN/(FN+TN) =11/(11+242) =4%	
			False Discovery Rate (FDR) =FP/(TP+FP) =166/(26+166) =86%	NPV =TN/(FN+TN) =242/(11+242) =96%	

Figure 1. Matrix of AI prediction results vs high and low risk.

The substantial growth in orthopedic waiting lists in the UK, particularly after the COVID-19 pandemic [1, 2], combined with the expanding role of ambulatory surgery centers in the United States, presents a major challenge for perioperative teams responsible for risk evaluation. The developed model offers a machine learning-based clinical decision support tool intended to assist with patient stratification and prioritization. The polynomial regression approach employed is a well-established statistical method frequently used in risk prediction. This study outlines the development of a model using risk ratio thresholds derived from a large cohort ($n = 6,730$). The specific risk threshold selected may be adjusted based on the local healthcare system and the level of care available across facilities to ensure appropriate patient routing.

When applied to predicting adverse outcomes after hip and knee replacement, the model effectively captures non-linear relationships between patient factors and surgical results. It strikes a balance—more sophisticated than basic logistic regression yet less prone to overfitting than neural networks, which can become overly dependent on training data. This positions it as a practical intermediate option for surgical risk forecasting [22-24]. The validation study affirms its suitability for risk

stratification in primary hip and knee arthroplasty. Performance in separating patients into high/moderate-risk versus low-risk groups was satisfactory and closely mirrored consultant-led care, with comparable accuracy in identifying individuals suitable for high-volume, low-complexity pathways.

The validation focused on detecting patients at elevated risk of complications who might require more intensive care. The overall prevalence of complications in the cohort was 8%. The model achieved 70% sensitivity in correctly classifying high-risk patients, with postoperative complications as the reference standard. This level of sensitivity is reassuring, as it correctly flags most patients likely to experience issues, and it matches the sensitivity observed with standard consultant- or attending-led assessments. This corresponds to a 30% false-negative rate — indicating that 30% of patients predicted to be low risk may still develop complications and warrant higher-level care. However, this rate is consistent with existing consultant-led preoperative screening processes. Flawless prediction of complications is unrealistic, and the AI model performs at least as well as current clinical screening methods. Notably, the negative predictive value (NPV) for the absence of listed complications reached 96%. This

reinforces that patients classified as low risk by the model are unlikely to encounter problems. Given that the baseline likelihood of an uncomplicated course in this population was 92%, the strong NPV supports directing such patients toward less intensive postoperative pathways. In high-volume, low-complexity (HVLC) facilities, patients routed to elective hubs or ambulatory surgery centers must have a high likelihood of a smooth recovery without extensive preoperative evaluation.

The findings also revealed a substantial number of false positives in the high-risk classification ($n = 166/408$). This 41% false-positive rate implies that many patients without complications may still be labeled as high risk by the AI tool. If used in isolation, this could lead to unnecessary referrals to higher-level care facilities. Nevertheless, from a patient safety standpoint, overclassifying risk is preferable to sending higher-need patients to sites without adequate postoperative emergency support.

One patient died in the postoperative period and had been identified by the model as very high risk, with a predicted score of 34.4%. This suggests the tool may be effective at spotting the highest-risk individuals, although confirmation in larger cohorts is needed. Such capability could enable earlier consultant-led preoperative review and targeted optimization for these patients. Additionally, individualized risk percentages facilitate informed discussions with patients about their specific surgical risks and, in some cases, whether to proceed with elective surgery.

Clinically, interpret the model's utility in the context of orthopedic hip and knee procedures, as complication patterns may differ across surgical specialties. The prevalence of complications in this field is notably low—8% in the present study. This low rate supports default allocation of most patients to HVLC sites, while still allowing screening to detect the minority at elevated risk. The AI model successfully identified these patients at an early stage in 70% of cases. This capability could also alleviate pressure on preoperative services by enabling streamlined or lighter assessment pathways for lower-risk individuals, who could be triaged promptly to elective HVLC hubs or ambulatory surgery centers, yielding potential staff efficiencies and improved workflow.

Risk ratios derived from waiting list data could further aid resource planning by enabling advance scheduling of case volumes across hospital sites, with better

anticipation of anesthetic requirements and the need for high-dependency or critical care support.

Complications can be serious and often require additional medical treatment or extended hospital stays. Several of the tracked adverse events are well-recognized and potentially life-threatening, including stroke, myocardial infarction, pulmonary embolism, and deep vein thrombosis. Others, such as urinary retention or ileus, may be less critical but can still significantly affect recovery.

Although machine learning provides an advanced approach to preoperative planning, it has limitations and remains susceptible to errors. No predictive tool can perfectly identify every patient who will experience complications, and all patients carry some degree of risk. In this study, complications did occur in some individuals predicted as low risk, many of which would have been challenging to anticipate during routine preoperative review (for instance, urinary retention in an otherwise healthy patient without prior prostatic issues). Certain unmodeled data points might have offered clues, such as elevated body mass index increasing the likelihood of wound infection. Future refinements using larger datasets may further improve accuracy. Overall, the current results indicate that the model is at least as reliable as consultant- or attending-led preoperative assessments in distinguishing high- and low-risk patients.

The performance of any machine learning tool ultimately depends on the quality, currency, security, and accuracy of the input data from patient records. This underscores that such models should serve as efficient clinical adjuncts under appropriate clinician supervision, rather than as standalone solutions. Model performance is expected to improve over time as additional data are incorporated.

The overall value of AI in this domain hinges on its acceptance by clinical staff and patients, as well as whether integration with human expertise yields better outcomes than either approach alone. This study provides evidence supporting the deployment of validated machine learning models to inform shared decision-making, help manage elective surgery backlogs, and adapt to evolving perioperative care models in ambulatory surgery centers.

Limitations

A notable limitation is the model's lack of preoperative functional status measures. A patient's baseline physical function and overall health significantly affect

postoperative complication risk. Functional assessments—whether through objective clinical evaluation or validated patient-reported tools—capture physiological reserve and comorbidity burden, both of which influence surgical course and recovery. Due to constraints in the available dataset, consistent functional metrics could not be included without risking bias or compromising model integrity. Future iterations could benefit from incorporating a hierarchy of complications via severity weighting or additional modeling to create more clinically meaningful outcome categories. Plans are underway to integrate validated preoperative functional assessments to further enhance risk stratification.

Conclusion

An artificial intelligence (AI) model was created for risk stratification in hip and knee arthroplasty. The model generates risk scores for postoperative complications and classifies patients into high- and low-risk groups. Validation in 445 patients demonstrated 70% sensitivity for identifying high-risk cases and a negative predictive value of 96%. The AI tool's risk-prediction performance was comparable to that of a traditional consultant-led review and stratification. AI approaches show promise for more efficient preoperative risk assessment, reducing workload on assessment teams, optimizing resource allocation, and helping address elective surgery backlogs.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Hampton M, Riley E, Garneti N, Anderson A, Wembridge K. The orthopaedic waiting list crisis. *Bone Joint J*. 2021.
2. Howlett N, Wood R. Modeling the recovery of elective waiting lists following COVID-19: scenario projections for England. *Value Health*. 2022.
3. French JMR, Deere K, Jones T, Pegg DJ, Reed MR, Whitehouse MR, et al. An analysis of the effect of the COVID-19-induced joint replacement deficit in England, Wales, and Northern Ireland suggests recovery will be protracted. *Bone Joint Lett J*. 2024;106-B(8):834-41. doi:10.1302/0301-620X.106B8.BJJ-2024-0036.R1
4. NHS GIRFT. Elective recovery high volume low complexity (HVLC). 2021.
5. Zhao S, Rothnie A, Nanda A, Chouari T, Ashraf S, Vig S. Developing bespoke high volume low complexity theatre lists with a focus on training to address the impact of COVID -19: a pilot study. *Cureus*. 2023.
6. Co M, Marks T, Tracey F, Conti S, Clarke GM. The Impact of Elective Surgical Hubs on Elective Activity in Acute Hospital Trusts in England: A Generalised Synthetic Control Study [Internet]. Available from: SSRN: <https://ssrn.com/abstract=4888136> or <https://doi.org/10.2139/ssrn.4888136>.
7. Myers T, Ramkumar P, Ricciardi B, Urish K, Kipper J, Ketonis C. Artificial intelligence and orthopaedics. *J Bone Jt Surg Am*. 2020.
8. Purnomo G, Yeo S, Liow M. Artificial intelligence in arthroplasty. *Arthroplasty*. 2021.
9. Copeland G, Jones D, Walters M. POSSUM; A scoring system for surgical audit. *Br J Surg*. 1991.
10. Pugely A, Martin C, Gao Y, Klocke N, Callaghan J, Marsh J. A risk calculator for short-term morbidity and mortality after hip fracture surgery.
11. Hassan AM, Rajesh A, Asaad M, Nelson J, Coert J, Mehrara B, et al. Artificial intelligence and machine learning in prediction of surgical complications: current state, applications, and implications. *Am Surg*. 2023.
12. Stam W, Goedknecht L, Ingwersen E, Schoonmade L, Bruns E, Daams F. The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: a systematic review. *Surgery*. 2021.
13. Buchlak Q, Esmaili N, Leveque J, Farrokhi F, Bennett C, Piccardi M, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. *Neurosurg Rev*. 2020.
14. Bozic KJ, Ong K, Lau E, Berry DJ, Vail TP, Kurtz SM, et al. Estimating risk in medicare patients with THA: an electronic risk calculator for periprosthetic joint infection and mortality. *Clin Orthop Relat Res*. 2013;471(2):574.
15. Harris AHS, Kuo AC, Bozic KJ, Lau E, Bowe T, Gupta S, et al. American joint replacement registry

- risk calculator does not predict 90-day mortality in veterans undergoing total joint replacement. *Clin Orthop*. 2018;476(9):1869-75.
16. Devana SK, Shah AA, Lee C, Jensen AR, Cheung E, van der Schaar M, et al. Development of a machine learning algorithm for prediction of complications and unplanned readmission following primary anatomic total shoulder replacements. *J Shoulder Elb Arthroplasty*. 2022;6:24715492221075444.
 17. Gu GS, Zhang DB, Zhang BH, Sun NK. Evaluation of P-POSSUM scoring system in predicting mortality in patients with hip joint arthroplasty. *Chin J Traumatol Zhonghua Chuang Shang Za Zhi*. 2006;9(1):50-5.
 18. Hildén M, Wretenberg P, Ekström W. Good overall morbidity prediction with the POSSUM scoring system in patients having a total hip or knee replacement – a prospective study in 227 patients. *Clin Interv Aging*. 2018;13:1747-54.
 19. McGirt MJ, Sivaganesan A, Asher AL, Devin CJ. Prediction model for outcome after low-back surgery: individualized likelihood of complication, hospital readmission, return to work, and 12-month improvement in functional disability. *Neurosurg Focus*. 2015;39(6):E13. doi:10.3171/2015.8.FOCUS15338
 20. Johnson L, Patel K, Smith J. Relative risk reduction as a basis for setting clinical thresholds: a review of current practices. *Am J Epidemiol*. 2020;189(4):412-20.
 21. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg*. 2004;240(2):205-13.
 22. Smith J, Patel K, Johnson L. Risk reduction and threshold setting in clinical decision making: a review of methodologies. *Eur J Med Res*. 2021;26(1):56-65.
 23. Kokkinakis S, Kritsotakis E, Lasithiotakis K. Artificial intelligence in surgical risk prediction. *J Clin Med*. 2023.
 24. Steyerberg EW. *Clinical prediction models*. Berlin/Heidelberg, Germany: Springer International Publishing; 2019.