American College of Rheumatology
Empowering Rheumatology Professionals

**Arthritis Care & Research**
Vol. 76, No. 5, May 2024, pp 636–643
DOI 10.1002/acr.25287
© 2023 American College of Rheumatology

# Development of a Machine Learning Model to Predict the Use of Surgery in Patients With Rheumatoid Arthritis

Natalie B. Baxter,[1] Ching-Heng Lin,[2] Beth I. Wallace,[3] Jung-Sheng Chen,[2] Chang-Fu Kuo,[2] and Kevin C. Chung[1]

**Objective.** One in five patients with rheumatoid arthritis (RA) rely on surgery to restore joint function. However, variable response to disease-modifying antirheumatic drugs (DMARDs) complicates surgical planning, and it is difficult to predict which patients may ultimately require surgery. We used machine learning to develop predictive models for the likelihood of undergoing an operation related to RA and which type of operation patients who require surgery undergo.

**Methods.** We used electronic health record data to train two extreme gradient boosting machine learning models. The first model predicted patients' probabilities of undergoing surgery ≥5 years after their initial clinic visit. The second model predicted whether patients who underwent surgery would undergo a major joint replacement versus a less intensive procedure. Predictors included demographics, comorbidities, and medication data. The primary outcome was model discrimination, measured by area under the receiver operating characteristic curve (AUC).

**Results.** We identified 5,481 patients, of whom 278 (5.1%) underwent surgery. There was no significant difference in the frequency of DMARD or steroid prescriptions between patients who did and did not have surgery, though nonsteroidal anti-inflammatory drug prescriptions were more common among patients who did have surgery ($P = 0.03$). The model predicting use of surgery had an AUC of 0.90 ± 0.02. The model predicting type of surgery had an AUC of 0.58 ± 0.10.

**Conclusions.** Predictive models using clinical data have the potential to facilitate identification of patients who may undergo rheumatoid-related surgery, but not what type of procedure they will need. Integrating similar models into practice has the potential to improve surgical planning.

## INTRODUCTION

Over 20% of patients with rheumatoid arthritis (RA) undergo orthopedic procedures to reduce pain or restore joint function.[1–3] However, there is wide variation in the surgical management of RA. Although guidelines exist to inform medication selection among patients with RA, there are few published recommendations for nonpharmacologic interventions.[4] No tools exist to pinpoint which specific patients may benefit from surgical procedures along RA clinical courses. Consequently, many patients do not reap the benefits of prophylactic procedures or early joint replacements.

Although research has identified predictors of orthopedic surgery for RA,[1–3] these variables have not been meaningfully incorporated into clinical decision-making. Using predictive models to determine the likelihood that a patient will undergo surgery could augment shared decision-making and increase access to effective procedures. One attractive option uses electronic health record (EHR) data to train predictive models built using machine learning (ML). Predictive models applying ML to EHR data have previously been used to predict disease activity and clinical flares in RA and are appealing because all data used to build the model are readily available to clinicians.[5,6]

Supported by joint Chang Gung Memorial Hospital-University of Michigan Medical Center (grants CORPG3L0161 and CORPG3N0191 to Dr Kuo and Dr Chung). Dr Wallace's work was supported by the National Center for Advancing Translational Sciences (grant 5KL2TR002241) and the VA Office of Research & Development (grant CX002430-01).

[1]Natalie B. Baxter, BSE, Kevin C. Chung, MD, MS: University of Michigan, Ann Arbor; [2]Ching-Heng Lin, PhD, Jung-Sheng Chen, MS, Chang-Fu Kuo, MD, PhD: Chang Gung Memorial Hospital, Taipei, Taiwan; [3]Beth I. Wallace, MD, MS: University of Michigan and VA Ann Arbor Healthcare System, Ann Arbor.

Ms Baxter and Dr Lin are co-first authors and contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002.acr.25287).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25287.

Address correspondence via email to Chang-Fu Kuo, MD, PhD, at zandis@gmail.com; Kevin C. Chung, MD, MS, at kecchung@med.umich.edu.

Submitted for publication September 10, 2023; accepted in revised form December 20, 2023.

**SIGNIFICANCE & INNOVATIONS**

- A machine learning model can accurately predict the distant use of surgery among patients with rheumatoid arthritis.
- Electronic health record data, including demographic information and medication use, is useful for the prediction of rheumatoid arthritis outcomes.

We aimed to use retrospective EHR data at a large academic medical center to develop ML models predicting the likelihood of undergoing an RA-related operation and the type of procedure a patient requiring surgery would undergo. We hypothesized that the first model would have high discrimination in predicting whether a patient would undergo surgery. We expected the second model to clearly distinguish between patients who underwent major joint replacements and those who had more minor joint-sparing soft-tissue procedures. Considering that the success of ML models requires representative data, we also compared the study cohort to prior analyses of patients undergoing RA-related procedures.

## METHODS

There were several steps to developing the ML models. First, data were collected and sorted into predictors and outcomes. Next, the ML models—specifically gradient-boosted decision tree models—were applied to the data and iteratively modified until the prediction was optimized. Model 1 predicted whether patients would undergo surgery. Model 2 predicted what type of surgery patients would have. We followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis reporting guidelines (Supplementary Table 1).[7] This study was approved by the Michigan Medicine Institutional Review Board (HUM00183488).

**Study cohort.** We identified patients treated for RA at Michigan Medicine between January 1, 2006, and June 27, 2021 (Figure 1). Patients were classified as having RA if they had at least 1 *International Classification of Disease, Ninth Revision* (ICD-9) and/or *Tenth Revision* (ICD-10) diagnostic code for RA, and ≥1 prescription for a conventional synthetic disease-modifying antirheumatic drug (csDMARD) (Supplementary Tables 2 and 3).[8–10] For each patient, the index date was defined as the last clinic visit at least five years prior to surgery (Figure 2). Patients were required to have hospital encounters spanning at least five years to avoid unequal time between the index date and the date of surgery. No data from this five year period were included in the analysis, so that predictions were made solely based on data from long before the patients' surgeries. The period of five years was selected to theoretically provide enough time to modify a patient's treatment regimen based on model predictions before surgery is inevitable. Patients who lacked data on sex, race, or ethnicity were excluded from the model.

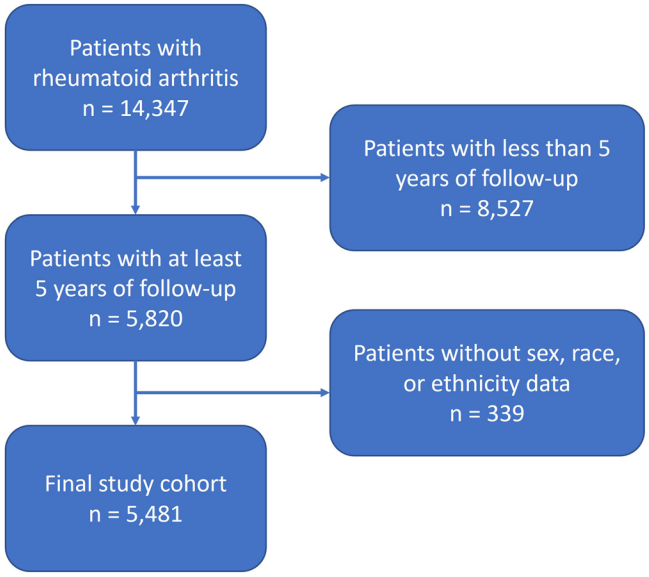



**Figure 1.** Number of patients included and excluded from the study based on criteria such as follow-up duration and missing data.

**Data preprocessing.** We considered 26 variables with clinical relevance to RA when developing the prediction models (Table 1). Demographic variables included sex, race, ethnicity, and age. Clinical variables included height, body mass index (BMI), and rate of change in weight over time. Comorbidities relevant to RA were captured using ICD-9 and ICD-10 codes (Supplementary Table 4) and are listed in Table 1. Laboratory data included glucose level, anticyclic citrullinated peptide level, C-reactive protein level, Westergren red blood sedimentation rate, and rheumatoid factor status (coded as positive or negative). In addition, we incorporated frequency of csDMARD, biologic or targeted synthetic DMARD (bDMARD or tsDMARD), glucocorticoid, and nonsteroidal anti-inflammatory drug (NSAID) use prior to the index date, measured in prescriptions per year.

If data on a particular variable were missing from at least 30% of patients after data sorting, that variable was excluded from model development. If <30% of patients were missing a specific variable, then we performed median imputations to replace the missing data. For example, if an encounter lacked BMI data, we used the median BMI of the entire population as the input for that variable. Ultimately, 21 features were included in model development (Table 1), including the aforementioned demographic and clinical characteristics, comorbidities, and medications. Laboratory tests were excluded during the data sorting phase because >30% of values were missing from the dataset.

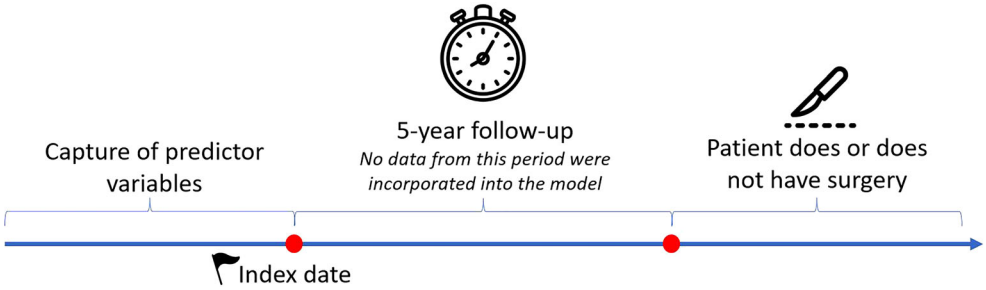



**Figure 2.** Timeline demonstrating the longitudinal set up of the study. There were at least five years between the index date and date of surgery. Only data from prior to the index date were used to make the prediction. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25287/abstract.

**Table 1.** Predictor variables*

| Variables | Definition |
|---|---|
| Demographic variables | |
| Sex | Male, female |
| Race | White, African American, Asian, American Indian or Alaska Native, other. |
| Ethnicity | Hispanic, non-Hispanic. |
| Age | Continuous variable based on closest record before index encounter. |
| Comorbidities | Binary variables (yes/no) based on closest record before index encounter. |
| Alcohol misuse | |
| Currently smokes | |
| Lung disease | |
| Rheumatoid nodules | |
| Eye disease | |
| Felty syndrome | |
| Heart disease | |
| Amyloidosis | |
| Sjogren syndrome | |
| Rheumatoid vasculitis | |
| Clinical | |
| Height | Continuous variables based on closest record before index encounter. |
| BMI | |
| Weight change | Continuous variable defined as change in kg divided by observation time in years. |
| Medications | |
| Conventional synthetic DMARDs | Continuous variables defined as number of prescriptions divided by the observation time in years. |
| Biologic DMARDs | |
| Corticosteroids | |
| NSAIDs | |
| Laboratory tests[a] | Binary variables (normal/abnormal) based on data closest to index date |
| Glucose | Binary variables (normal/abnormal) based on data closest to index date |
| Cyclic citrullinated peptide | |
| C-reactive protein | |
| Westergren red blood cell sedimentation | |
| Rheumatoid factor status | |

* BMI, body mass index; DMARDs, disease-modifying antirheumatic drugs; NSAIDs, nonsteroidal anti-inflammatory drugs.
[a] Excluded during the data sorting phase because of the number of missing values.

Model 1 predicted whether a patient would undergo an elective reconstructive procedure for joint damage related to RA at least five years after the index date (Figure 2). Model 2 predicted whether the subset of patients undergoing surgery would undergo a major versus minor procedure (Table 2). We conducted a comprehensive literature review to identify reconstructive procedures that patients receive for RA. For Model 1, we used common procedural terminology codes to identify individuals in the larger patient cohort who underwent RA-related surgery (Supplementary Table 5). For Model 2, we designated procedures as major versus minor based on prior categorization in the literature.[11,12] Major procedures included total large joint replacement and cervical spine fusion. Minor procedures included small joint surgeries, synovectomy, soft tissue procedures, and preventive osteotomies.

**Model development.** Both models separately underwent 5-fold cross-validation (Supplementary Figure 1). Data underwent preprocessing within each fold of cross-validation to standardize the data. Specifically, we used min–max scaling to normalize the data ranges of continuous variables. Unlike other scaling methods, min–max scaling preserves the distribution and relative relationships between data points. One-hot encoding was used to convert nominal data into numerical values and label encoding was applied to sort the outcomes.[13] Additionally, we used the Tomek links undersampling method to address data imbalance.[14] Tomek links are data points from different classes, such as the surgical versus nonsurgical cohorts, that are close to each other in value. The data point from the majority data set is removed to reduce noise and balance the amount of data from different subgroups.

After executing all preprocessing procedures, we used the algorithm to build the prediction models. XGBoost is an extreme gradient boosting model based on supervised ML, decision tree, and ensemble learning principles. XGBoost outperforms other ML algorithms in various medical studies.[15,16] It combines the predictions of multiple decision trees to create a more robust final prediction model. The algorithm determines the best

**Table 2.** Surgical procedures*

| Categories | Procedures |
|---|---|
| Major procedures | |
| Large joint arthroplasty | Shoulder, elbow, hip, knee |
| Spine fusion | Cervical, thoracic, lumbar |
| Intermediate procedures | |
| Small joint arthroplasty | Wrist, MCP, IP, CMC, ankle, MTP |
| Joint fusion | Wrist, CMC, MCP, IP, pantalar, subtalar, midtarsal, tarsometatarsal, metatarsophalangeal, subtalar, digit tendon sheath, shoulder, hip, knee |
| Synovectomy | Ankle, wrist, CMC, MCP, IP, finger tendon sheath, elbow, feet joints/ tendons, shoulder, hip, knee |
| Osteotomy | Ulna, MTP, calcaneus, talus, other tarsal bones, proximal phalanx |
| Tendon transfer | Forearm/wrist, finger, CMC/dorsum of hand, palm, thumb, cross intrinsic transfer |
| Tendon repair | Hand tendon repair, ECU subluxation procedures |
| Minor procedures | |
| Boutonniere or swan neck treatment | |
| Carpal tunnel release | |
| Nodule removal | |

* CMC, carpometacarpal; ECU, extensor carpi ulnaris; IP, interphalangeal; MCP, metacarpophalangeal; MTP, metatarsophalangeal.

hyperparameters of the algorithm by minimizing a loss function through gradient descent.

**Statistical analysis.** We performed descriptive analyses of the study population. We reported percentages for categorical data and means plus SDs for continuous variables. We used *t*-tests and chi-square tests to compare continuous and categorical data, respectively. We evaluated the average duration between the index date and surgery to ensure consistency across the study population.

Discrimination, a measure of how well a model can differentiate between patients who did and did not experience the outcome, was measured by averaging the area under the receiver operating characteristic curve (AUC) of all folds of the cross-validation. We also compared the XGBoost model outcomes to standard logistic regression models. In addition, feature importance was measured using XGBoost to identify which variables contributed the most to model predictions. XGBoost relies on the Gini index to calculate feature importance.[17] We also generated Shapley Additive explanations (SHAP) analyses to portray the impact of higher versus lower variable values on the model output. Finally, we generated precision-recall (PR) curves to evaluate the quality of the results. Precision relates to the false positive rate, and measures whether a model can return accurate results. Recall is a measure of the false negative rate, thus indicating whether a model can return the majority of positive results.

**Software.** All statistical analyses were performed using SAS statistical software (version 9.4), scikit-learn (version 1.0.2), py-xgboost-cpu (version 1.7.0), and Python (version 3.7). The code and hyperparameters for XGBoost are provided in the Supplementary Materials.

## RESULTS

**Cohort characteristics.** The study cohort included 5,481 patients (4,283 female [78.2%]; 4,594 White [83.8%]; 5,296 non-Hispanic [96.6%]), of whom 278 (5.1%) underwent surgery (Table 3). The average ± SD age of patients in the surgery versus nonsurgery cohorts was 53.4 ± 14.1 years versus 56.3 ± 16.5 years ($P < 0.01$). Prescription frequency for csDMARDs, bDMARD or tsDMARD, NSAIDs, and steroids were more common in the surgery than the nonsurgery cohort, with only NSAIDs being significantly more common ($P = 0.03$). A breakdown of which specific csDMARDs, bDMARDs, tsDMARDs, steroids, and NSAIDs that were prescribed is displayed in Supplementary Table 3. Methotrexate (42.1%) and hydroxychloroquine (40.7%) accounted for the majority of csDMARD prescriptions, whereas adalimumab (36.7%) and etanercept (34.7%) accounted for the majority of bDMARD and tsDMARD prescriptions. Osteoarthritis was the most prevalent comorbidity in 16.5% versus 33.5% of patients who underwent surgery versus patients who did not ($P < 0.01$). The presence of RA-related ocular involvement was also significantly more common among patients not undergoing surgery ($P = 0.04$). For patients who underwent surgery, the average duration between the index date and date of surgery was 5.9 ± 1.4 years. The average duration between patients' first and last encounter was 9.7 ± 2.9 years for the surgery cohort and 10.4 ± 3.3 years for the nonsurgical cohort.

With respect to surgery type, 165 patients (59.4%) underwent major procedures, and 113 patients (40.6%) underwent minor procedures. Prescriptions for bDMARDs and tsDMARDs were more common in the minor procedure cohort (9.8 vs 0.05 prescriptions per year, $P = 0.03$) and prescriptions for csDMARDs were less common in the minor procedure cohort compared to the major procedure cohort (6.9 vs 29.3 prescriptions per year, $P = 0.03$). Patients who underwent minor procedures were also younger (50.3 vs 55.6 years old, $P < 0.01$).

**Prediction model results.** Figure 3 shows the AUC for the XGBoost models. Model 1, which predicted the use of surgery, demonstrated an AUC of 0.90 ± 0.02, meaning that the model could accurately predict whether a patient would have surgery 90% of the time. Model 2, which predicted the type of surgery, demonstrated an AUC of 0.58 ± 0.10. Predictors with

**Table 3.** Cohort characteristics*

| Characteristics | Use of surgery prediction | | | Surgery type prediction | | |
|---|---|---|---|---|---|---|
| | Surgery | Nonsurgery | *P* value | Major | Minor or intermediate | *P* value |
| Number of patients | 278 | 5,203 | | 165 | 113 | |
| Age, mean ± SD, y | 53.4 ± 14.1 | 56.3 ± 16.5 | <.01 | 55.6 ± 13.5 | 50.3 ± 14.4 | <.01 |
| Sex, n (%) | | | 0.91 | | | 0.06 |
| Female | 218 (78.4) | 4,065 (78.1) | | 123 (74.5) | 95 (84.1) | |
| Male | 60 (21.6) | 1,138 (21.9) | | 42 (25.5) | 18 (15.9) | |
| Race, n (%) | | | 0.12 | | | 0.29 |
| African American | 21 (7.6) | 528 (10.1) | | 14 (8.5) | 7 (6.2) | |
| American Indian or Alaska Native | 0 (0) | 32 (0.6) | | 0 (0) | 0 (0) | |
| Asian | 3 (1.1) | 125 (2.4) | | 2 (1.2) | 1 (0.9) | |
| White | 249 (89.6) | 4,345 (83.5) | | 148 (89.7) | 101 (89.4) | |
| Native Hawaiian and Pacific Islander | 0 (0) | 8 (0.2) | | 0 (0) | 0 (0) | |
| Other | 5 (1.8) | 165 (3.2) | | 1 (0.6) | 4 (3.5) | |
| Ethnicity, n (%) | | | 0.14 | | | 0.37 |
| Hispanic | 5 (1.8) | 180 (3.5) | | 2 (1.2) | 3 (2.7) | |
| Non-Hispanic | 273 (98.2) | 5,023 (96.5) | | 163 (98.8) | 110 (97.3) | |
| Height, mean ± SD cm | 165.5 ± 6.9 | 165.7 ± 9.6 | 0.51 | 165.7 ± 6.3 | 165.1 ± 7.7 | 0.54 |
| BMI, mean ± SD | 29.2 ± 5.0 | 29.9 ± 7.8 | 0.03 | 29.6 ± 4.7 | 28.6 ± 5.4 | 0.10 |
| Slope of weight, mean ± SD | 0.001 ± 0.042 | 0.007 ± 0.050 | 0.88 | −0.004 ± 0.016 | 0.003 ± 0.063 | 0.55 |
| Alcohol abuse, n (%) | 1 (0.4) | 89 (1.7) | 0.08 | 1 (0.6) | 0 (0) | 0.41 |
| Current smoker, n (%) | 11 (4.0) | 365 (7.0) | 0.05 | 6 (3.6) | 5 (4.4) | 0.74 |
| Medication, Rx/year* (SD)* | | | | | | |
| NSAID | 18.6 (79.9) | 8.1 (55.7) | 0.03 | 24.5 (91.3) | 9.9 (58.9) | 0.11 |
| Biologic | 4.0 (37.8) | 1.8 (22.9) | 0.34 | 0.05 (0.31) | 9.8 (58.9) | 0.03 |
| Oral steroid | 16.5 (91.7) | 10.4 (59.5) | 0.27 | 9.3 (56.3) | 27.1 (126.5) | 0.16 |
| csDMARD | 20.2 (93.4) | 11.1 (62.3) | 0.11 | 29.3 (113.8) | 6.9 (48.3) | 0.03 |
| Comorbidities, n (%) | | | | | | |
| Felty syndrome | 2 (0.7) | 12 (0.2) | 0.12 | 2 (1.2) | 0 (0) | 0.24 |
| Ocular involvement | 0 (0) | 76 (1.5) | 0.04 | 0 (0) | 0 (0) | N/A |
| Heart involvement | 0 (0) | 6 (0.1) | 0.57 | 0 (0) | 0 (0) | N/A |
| Pulmonary involvement | 0 (0) | 57 (1.1) | 0.08 | 0 (0) | 0 (0) | N/A |
| Amyloidosis | 0 (0) | 12 (0.2) | 0.42 | 0 (0) | 0 (0) | N/A |
| Osteoarthritis | 46 (16.5) | 1,743 (33.5) | <.01 | 33 (20.0) | 13 (11.5) | 0.06 |

* BMI, body mass index; csDMARD, conventional synthetic antirheumatic drug; N/A, not applicable; NSAID, nonsteroidal anti-inflammatory drug; Rx/year, prescriptions per patient per year; SD, standard deviation.

the greatest feature importance in Model 1 included csDMARD use, oral steroid use, and change in weight, at 17%, 14%, and 11%, respectively. More specifically, the SHAP analyses demonstrated that patients with more frequent use of csDMARDs and oral steroids were more likely to undergo surgery. Additional feature importance values and SHAP analyses are included in

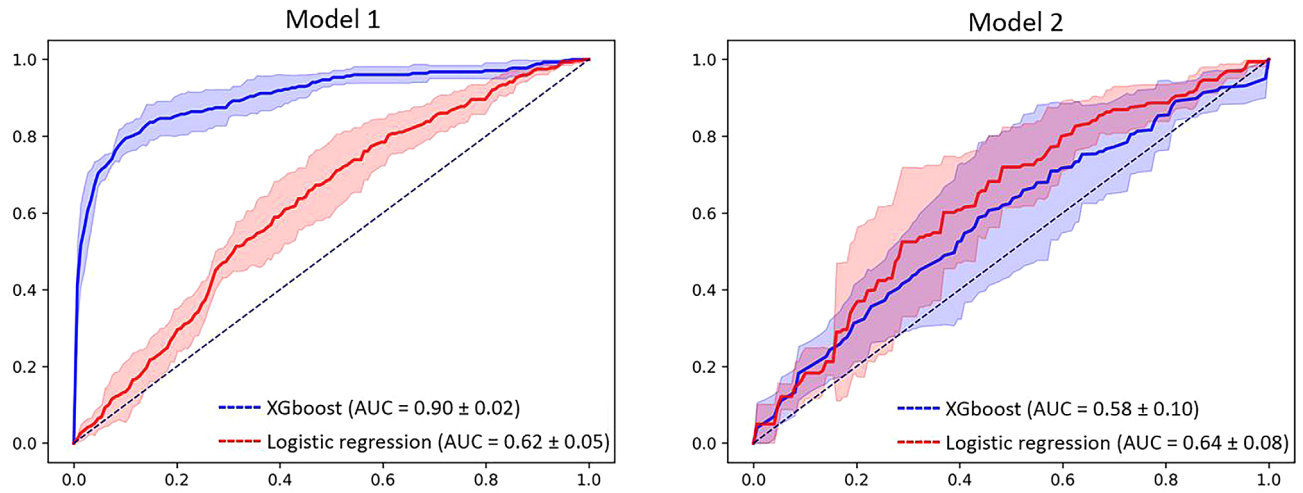



**Figure 3.** AUCs for Model 1 and Model 2 compared with logistic regression. AUC, area under the receiver operating characteristic curve. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25287/abstract.

Supplementary Table 6 and Supplementary Figures 2 and 3. As for the PR curves, the curve related to use of surgery (Model 1) demonstrated an AUC of 0.61 ± 0.10. For the prediction of surgery type (Model 2), the associated PR curve demonstrated an AUC of 0.66 ± 0.08. These models are included in Supplementary Figures 4 and 5.

## DISCUSSION

Using EHR data from an RA population at a major academic medical institution, we constructed an ML model to determine whether patients would undergo an RA-related procedure with 90% accuracy. Measures of csDMARD and oral steroid use contributed the most to developing an accurate model. We also developed a similar model to predict the type of surgery patients would undergo, which was not successful because it achieved low accuracy. Despite this result, our findings suggest that ML models have the potential to improve understanding of the general disease trajectory of patients with RA, which can inform shared decision making and facilitate coordination of surgical care.

Prior to model development, it was necessary to create a dataset and determine whether it was representative of the population of patients with RA. Like prior studies, we evaluated differences in the demographics, comorbidities, and medication use of patients who did and did not have surgery for RA.[1–3] The reported significance of these differences varies greatly across the literature and with respect to our findings.[18] For example, we found that just over 5% of patients in our cohort had surgery for RA, compared with the 15% to 25% of patients who were reported to have surgery in other studies.[19–21] The lower incidence of surgery we identified is likely associated with changes in procedure rates over time and specific practice patterns at the study institution. It is also possible that a portion of patients in our study underwent surgery after 2021, the last year from which data were collected. The outcomes of the ML models in this study also depended on the identification of variables that were not only relevant to RA, but could be acquired in a simple, standardized manner. A limitation of prior ML studies predicting RA disease progression was their reliance on genomic or patient-reported data that are difficult to collect in large amounts over long periods of time.[22,23] The variables we included were easily obtained via the EHR, facilitating construction of a data set that was adequate for model development.

Our findings indicate that certain medication variables are important for accurate model performance. Specifically, the variables for csDMARD and corticosteroid use had the greatest influence on the outcome of Model 1, which predicted the use of surgery. This result suggests that a portion of patients who undergo surgery for RA are undertreated. Rather than escalate to biologic therapy for disease that is poorly controlled with csDMARDs and steroids, they continue to take these medications. When joint damage becomes severe, patients taking csDMARDs and steroids may ultimately require surgery. The reason for this is multifactorial. Although access to bDMARDs and tsDMARDs is limited by high costs, when controlling for disease severity and insurance status, provider and regional factors are also associated with longer time to biologic initiation.[24–26] Although bDMARD and tsDMARD use had a lesser impact on Model 1 development, this is not surprising considering the relative infrequency with which these medications were prescribed in our study population. Similarly, the comorbidity and demographic factors included in our study likely had a minor impact because the cohort was relatively homogenous with respect to these factors. Overall, our findings suggest that accurate tracking of medication use is important to predict use of surgery among patients with RA.

The use of ML prediction has the potential to address several current challenges in the provision of surgery to patients with RA. Some variation in the use of surgery has been attributed to disagreement among rheumatologists and surgeons on the timing and outcomes of certain procedures.[2,27] It is also possible that patients do not seek out surgical care owing to a lack of knowledge or physician bias.[28–30] In addition, trends in the use of surgery are changing in response to the development of bDMARDs and tsDMARDs, which can be highly effective in retarding RA disease progression.[19,31–33] Despite these advancements, a portion of patients go on to develop structural change, joint pain and loss of function that may warrant surgery.[12] Ideally, an ML tool such as the one proposed in this study could be integrated into the EHR so that providers may be alerted of a patient's future need for surgical evaluation. Considering that all relevant variables are available in the EHR, predictions could be generated automatically, avoiding the need to insert numerous variables into a separate calculator. This would provide time to refer the patient to the appropriate specialist and make treatment changes based on the patient's goals to pursue more aggressive medical treatment or prepare for possible surgery.

Our findings must be interpreted in the context of several limitations. We recognize the concern that ML is a black box because the outcome of a model is not easily explained based on the inputs. However, it is important to note that a prediction tool such as the one proposed in this study is meant to support decision-making, rather than replace it. Another limitation is that the use of surgery is not only influenced by patients' objective clinical characteristics, but also by their treatment preferences and access to care.[29,34] Considering the retrospective nature of this study, we were unable to account for these factors in the prediction. Data imbalance was also a limitation for Model 1, meaning that there was imbalance in the number of surgical versus nonsurgical patients. This interfered with the model's precision, as demonstrated by the poor PR curve (Supplementary Figure 4). We also note that Model 2 did not successfully predict

whether surgical patients would receive major or minor procedures. This is likely explained by the small sample size, which limits the development of accurate regression models regardless of technique. This is evident in Figure 3 (Model 2) because both the linear regression and ML models performed poorly. The low AUC of the PR curve for Model 2 is also a consequence of the small sample size (Supplementary Figure 5). We also relied on a procedure categorization framework defined based on prior literature rather than quantitative measures. It is possible that this was not the most appropriate way to differentiate procedures. Furthermore, we were limited by the lack of laboratory data available in our institution's EHR. Finally, we did not perform external validation of the model to confirm that it performs accurately in different patient populations. Further refinement and validation will improve model accuracy and reproducibility across different health care systems with variable practice characteristics and resources.

The results of this study indicate that it is possible to train an ML model to predict the use of surgery for RA using EHR data. ML tools have the potential to improve recognition of patients who may benefit from surgery for RA. This may aid in the development of referral guidelines for RA–related procedures, enhance patient education, and improve communication between patients and their care teams.

## AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr Chung had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study conception and design.** Baxter, Lin, Wallace, Kuo, Chung.
**Acquisition of data.** Baxter, Wallace, Chen.
**Analysis and interpretation of data.** Baxter, Lin, Wallace, Chen, Kuo, Chung.

## REFERENCES


1. Gwinnutt JM, Symmons DPM, MacGregor AJ, et al. Predictors of and outcomes following orthopaedic joint surgery in patients with early rheumatoid arthritis followed for 20 years. Rheumatology (Oxford) 2017;56(9):1510–1517.
2. Nikiphorou E, Carpenter L, Norton S, et al. Can rheumatologists predict eventual need for orthopaedic intervention in patients with rheumatoid arthritis? Results of a systematic review and analysis of two UK inception cohorts. Curr Rheumatol Rep 2017;19(3):12.
3. Kapetanovic MC, Lindqvist E, Saxne T. Orthopaedic surgery in patients with rheumatoid arthritis over 20 years: prevalence and predictive factors of large joint replacement. Ann Rheum Dis 2008; 67(10):1412–1416.
4. Smolen JS, Landewé RBM, Bergstra SA, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Ann Rheum Dis 2023;82(1):3–18.
5. Norgeot B, Glicksberg BS, Trupin L, et al. Assessment of a deep learning model based on electronic health record data to forecast clinical outcomes in patients with rheumatoid arthritis. JAMA Netw Open 2019;2(3):e190606.
6. Vodencarevic A, Tascilar K, Hartmann F, et al. Advanced machine learning for predicting individual risk of flares in rheumatoid arthritis patients tapering biologic drugs. Arthritis Res Ther 2021;23(1):67.
7. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594.
8. Curtis JR, Xie F, Zhou H, et al. Use of ICD-10 diagnosis codes to identify seropositive and seronegative rheumatoid arthritis when lab results are not available. Arthritis Res Ther 2020;22(1):242.
9. Singh JA, Saag KG, Bridges SL Jr., et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol 2016;68(1):1–26.
10. Ng B, Aslam F, Petersen NJ, et al. Identification of rheumatoid arthritis patients using an administrative database: a Veterans Affairs study. Arthritis Care Res (Hoboken) 2012;64(10):1490–1496.
11. Trieb K, Hofstaetter SG. Treatment strategies in surgery for rheumatoid arthritis. Eur J Radiol 2009;71(2):204–210.
12. Dwivedi S, Testa EJ, Modest JM, et al. Surgical management of rheumatoid arthritis of the hand. R I Med J (2013) 2020;103(4):32–36.
13. Vasudev R. What is one hot encoding? Why and when do you have to use it? Medium. August 2, 2017. Accessed November 13, 2023. https://medium.com/hackernoon/what-is-one-hot-encoding-why-and-when-do-you-have-to-use-it-e3c6186d008f.
14. Tomek I. An experiment with the edited nearest-neighbor rule. IEEE Trans Syst Man Cyber 1976;SMC-6(6):448–452.
15. Chen R, Zhang S, Li J, et al. A study on predicting the length of hospital stay for Chinese patients with ischemic stroke based on the XGBoost algorithm. BMC Med Inform Decis Mak 2023;23(1):49.
16. Moore A, Bell M. XGBoost, a novel explainable AI technique, in the prediction of myocardial infarction: a UK Biobank cohort study. Clin Med Insights Cardiol 2022;16:11795468221133611.
17. Nembrini S, König IR, Wright MN. The revival of the Gini importance? Bioinformatics 2018;34(21):3711–3718.
18. Capelusnik D, Aletaha D. Baseline predictors of different types of treatment success in rheumatoid arthritis. Ann Rheum Dis 2022; 81(2):153–158.
19. Nikiphorou E, Carpenter L, Morris S, et al. Hand and foot surgery rates in rheumatoid arthritis have declined from 1986 to 2011, but large-joint replacement rates remain unchanged: results from two UK inception cohorts. Arthritis Rheumatol 2014;66(5):1081–1089.
20. Shourt CA, Crowson CS, Gabriel SE, et al. Orthopedic surgery among patients with rheumatoid arthritis 1980–2007: a population-based study focused on surgery rates, sex, and mortality. J Rheumatol 2012;39(3):481–485.
21. Pedersen AB, Mor A, Mehnert F, et al. Rheumatoid arthritis: trends in antirheumatic drug use, c-reactive protein levels, and surgical burden. J Rheumatol 2015;42(12):2247–2254.
22. Kedra J, Davergne T, Braithwaite B, et al. Machine learning approaches to improve disease management of patients with rheumatoid arthritis: review and future directions. Expert Rev Clin Immunol 2021;17(12):1311–1321.
23. Joo YB, Kim Y, Park Y, et al. Biological function integrated prediction of severe radiographic progression in rheumatoid arthritis: a nested case control study. Arthritis Res Ther 2017;19(1):244.
24. DeWitt EM, Lin L, Glick HA, et al. Pattern and predictors of the initiation of biologic agents for the treatment of rheumatoid arthritis in the United States: an analysis using a large observational data bank. Clin Ther 2009;31(8):1871–1880.
25. Kalkan A, Husberg M, Hallert E, et al. Physician preferences and variations in prescription of biologic drugs for rheumatoid arthritis: a register-based study of 4,010 patients in Sweden. Arthritis Care Res (Hoboken) 2015;67(12):1679–1685.

26. Tatangelo M, Tomlinson G, Paterson JM, et al. Association of patient, prescriber, and region with the initiation of first prescription of biologic disease-modifying antirheumatic drug among older patients with rheumatoid arthritis and identical health insurance coverage. JAMA Netw Open 2019;2(12):e1917053.

27. Alderman AK, Chung KC, Kim HM, et al. Effectiveness of rheumatoid hand surgery: contrasting perceptions of hand surgeons and rheumatologists. J Hand Surg Am 2003;28(1):3–11.

28. Motomiya M, Nishimoto N, Urita A, et al. Patient characteristics affecting knowledge of the possibility of surgical reconstruction for rheumatoid hand and wrist deformities. Mod Rheumatol 2021;31(3): 614–620.

29. Alderman AK, Arora AS, Kuhn L, et al. An analysis of women's and men's surgical priorities and willingness to have rheumatoid hand surgery. J Hand Surg Am 2006;31(9):1447–1453.

30. Nowell WB, Venkatachalam S, Stake C, et al. Identifying patient decisions and related information needs during decision making related to total knee arthroplasty. J Comp Eff Res 2020;9(16):1153–1166.

31. Park JW, Hwang JS, Gong SH. Rheumatoid hand surgery in the era of biologic therapy: a rheumatologist-oriented overview. J Rheum Dis 2021;28(4):192–201.

32. Richter MD, Crowson CS, Matteson EL, et al. Orthopedic surgery among patients with rheumatoid arthritis: a population-based study to identify risk factors, sex differences, and time trends. Arthritis Care Res (Hoboken) 2018;70(10):1546–1550.

33. Chung KC, Burns PB, Wilgis EF, et al. A multicenter clinical trial in rheumatoid arthritis comparing silicone metacarpophalangeal joint arthroplasty with medical treatment. J Hand Surg Am 2009;34(5):815–823.

34. Vina ER, Quinones C. Understanding the role and challenges of patient preferences in disparities in rheumatologic disease care. Rheum Dis Clin North Am 2021;47(1):83–96.