

RESEARCH

Open Access



# Predicting abatacept retention using machine learning

Rieke Alten<sup>1\*</sup>, Claire Behar<sup>2</sup>, Pierre Merckaert<sup>3</sup>, Ebenezer Afari<sup>4</sup>, Virginie Vannier-Moreau<sup>5</sup>, Anael Ohayon<sup>5</sup>, Sean E. Connolly<sup>6</sup>, Aurélie Najm<sup>7</sup>, Pierre-Antoine Juge<sup>8</sup>, Gengyuan Liu<sup>6</sup>, Angshu Rai<sup>6</sup>, Yedid Elbez<sup>9</sup> and Karissa Lozenski<sup>6</sup>

## Abstract

**Background** The incorporation of machine learning is becoming more prevalent in the clinical setting. By predicting clinical outcomes, machine learning can provide clinicians with a valuable tool for refining precision medicine approaches and improving treatment outcomes.

**Methods** This was a post hoc analysis of pooled patient-level data from the global, real-world ACTION and ASCORE trials in patients with rheumatoid arthritis (RA) initiating abatacept. Patient demographic and disease characteristics were input across 10 machine learning models used to predict 12-month treatment retention. Retention was defined as treatment for > 365 days or ≤ 365 days in patients who achieved remission or major clinical response (based on European Alliance of Associations for Rheumatology response criteria). The pooled dataset was split into a training/validation cohort for model development and a test cohort for an unbiased evaluation of performance. SHapley Additive exPlanation (SHAP) values determined the level of importance and directionality for key patient features predicting abatacept retention.

**Results** The pooled ACTION and ASCORE dataset included 5320 patients with RA (mean [standard deviation] age 57.7 [12.7] years; 79% female). The 12-month abatacept retention rate was 61% ( $n = 3236$ ) with a discontinuation rate of 39% ( $n = 2037$ ). In the training set ( $n = 4218$ ), the gradient-boosting classifier model demonstrated the best performance (testing accuracy: 62%). This model had an area under the receiver operating characteristic curve (95% confidence interval) of 0.620 (0.586, 0.653) and F1 score of 0.659 (0.625, 0.689) in the test set of patients ( $n = 1055$ ). Using this model, the five most important variables predicting 12-month abatacept retention were low body mass index (BMI), low American College of Rheumatology functional status class, anti-citrullinated protein antibody (ACPA) positivity, low Patient Global Assessment, and younger age.

**Conclusions** The gradient-boosting classifier model identified key patient features predictive of abatacept retention from this large, real-world study population. The SHAP values conveyed the directionality and importance of BMI, functional status, ACPA serostatus, Patient Global Assessment, and age for abatacept retention. Findings are consistent with previous observations and help validate the machine learning approach for predictive modelling in RA treatment, and may help inform clinical decision making.

**Trial registration** NCT02109666 (ACTION), NCT02090556 (ASCORE).

Anael Ohayon, Angshu Rai and Karissa Lozenski were affiliated to their institution at the time of analysis.

\*Correspondence:

Rieke Alten

Rieke.alten@schlosspark-klinik.de

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

**Keywords** Abatacept, Machine learning, Retention, Rheumatoid arthritis, Treatment response

## Background

Identifying efficacious treatment for patients with rheumatoid arthritis (RA) remains a clinical challenge. Currently, a treat-to-target approach is recommended for the management of RA [1, 2]. In addition to clinical response, other treatment outcome measures such as treatment retention are important to consider. Retention rates are influenced by factors such as lack of efficacy and adverse events, among many others [3]. Importantly, predicting response to treatment for patients with different profiles may help clinicians choose appropriate treatments and identify patients most likely to have success on any given therapy.

Abatacept, a selective co-stimulation modulator acting to inhibit T-cell activation [4], is approved for the treatment of moderate-to-severe active RA [5]. Previous observations from the AbataCepT In rOutiNe clinical practice (ACTION) [6] and Abatacept SubCutaneOus in Routine clinical practicE (ASCORE) [7] trials have identified patient features capable of predicting abatacept retention. In addition to conventional statistical analysis, machine learning may provide an alternative methodology that can predict specific outcomes by learning rules from existing data [8]. Machine learning techniques can allow the analysis of large datasets, such as those from observational studies, to identify factors that influence specific outcomes, such as treatment retention. After applying machine learning techniques to the ACTION trial data [9], a gradient-boosting classifier model identified key patient features similar to those identified from previous ACTION analyses using more traditional methods [6]. Subsequent analyses expanded upon the initial ACTION trial predictive model by describing the directionality of key patient features and how they impacted abatacept retention through SHapley Additive exPlanation (SHAP) values [10]. That analysis revealed that previous corticosteroid use was associated with lower retention and that American College of Rheumatology (ACR) functional class II was associated with higher retention. Consequently, these observations may provide clinicians with unsuspected but relevant patient features that can inform precision medicine approaches, thereby improving treatment retention and clinical outcomes.

Previous work has leveraged the similar methodologies between ACTION and ASCORE to perform concurrent analyses of clinical response [11, 12]. Those results emphasized the clinical value of baseline serostatus (ie, anti-citrullinated protein antibody [ACPA] positivity) on clinical response for patients with RA receiving abatacept

[11, 12]. Given the similarities in the methods used in the ACTION and ASCORE trials, this post hoc machine learning analysis pooled patient-level data from both the ACTION and ASCORE trials to expand upon the initial machine learning observations performed in ACTION. As such, the objectives of this analysis were to use machine learning techniques to assess the clinical importance and directionality of patient demographic and disease characteristics for predicting abatacept retention at 12 months in a large cohort of patients.

## Methods

### Study design and patients

This post hoc analysis included patient-level data from the ACTION [6] and ASCORE [7] trials. ACTION was a large, international observational trial of adult patients with moderate-to-severe RA who were enrolled at initiation (or within 3 months of initiation) of treatment with intravenous abatacept (body weight-adjusted dosing) across Europe and Canada for up to 2 years [6]. ASCORE was a 2-year, observational, prospective multicenter trial of subcutaneous abatacept (125 mg once weekly) for the treatment of patients with moderate-to-severe active RA [7]. All patients were  $\geq 18$  years old, with active moderate-to-severe RA (ACR/European Alliance of Associations for Rheumatology [EULAR] 2010 criteria) [13]. Additional trial details have been previously reported [6, 7].

### Endpoints and variables

Patient demographic and disease characteristics assessed at baseline were entered into the machine learning algorithms. Disease characteristics included: use of previous biologic treatment (yes/no), number of previous biologic treatments, number of previous tumor necrosis factor (TNF) inhibitors, duration of RA, baseline ACPA and rheumatoid factor (RF) serostatus, ACR functional status class (low/high), nonsteroidal anti-inflammatory drug use (yes/no), methotrexate (MTX) use (yes/no), MTX dose, baseline corticosteroid use (yes/no), baseline corticosteroid dose, tender and swollen joint count in 28 joints (TJC28 and SJC28, respectively), Physician Global Assessment, Patient Global Assessment, pain, radiographic erosion status, history of a past/present neoplasm (yes/no), abatacept monotherapy (yes/no), and abatacept route of administration (IV/SC). Disease activity was collected at baseline and at 3, 6, 12, 18, and 24 months of follow-up and included Disease Activity

Score in 28 joints (DAS28) using erythrocyte sedimentation rate (DAS28-ESR) and C-reactive protein (DAS28-CRP), Clinical Disease Activity Index (CDAI), Simplified Disease Activity Index (SDAI), and Health Assessment Questionnaire-disability index (HAQ-DI). A subset of variables was selected to improve model performance to avoid under- or over-fitting of machine learning models.

The primary endpoint of this analysis was 12-month abatacept retention. The retention label was created from the duration of exposure to abatacept and the reason for discontinuation defined as: retention=1, duration > 365 days and duration ≤ 365 days due to remission/major clinical response; or retention=0, duration ≤ 365 days not due to remission/major clinical response. Major clinical response was defined using EULAR response criteria based on DAS28 (ESR or CRP) [14]. Six-month retention was an additional endpoint employing the same retention label as described above but with timing of 183 days. Further models examined 3-month retention after 3-month follow-up (months 3–6), 6-month retention after 3-month follow-up (months 3–9), 3-month retention after 6-month follow-up (months 6–9), and 6-month retention after 6-month follow-up (months 6–12), with the same retention labels as described above and timings of 91 days for 3-month retention or 183 days for 6-month retention.

#### **Training and testing cohorts**

Before modelling, the combined ACTION/ASCORE population database was divided into two separate cohorts: a training/validation cohort and a test cohort. Sampling was stratified to ensure that both cohorts were representative and that relative class frequencies, including retention, were preserved. Class imbalance in the data was addressed by generating the ratio between the two classes from the training data and including those as sample weights when the loss function was calculated during cross validation. The machine learning algorithms used the training/validation cohort data to learn from and build the model (using the nested cross-validation process). The test cohort was then used for testing the resulting model against completely unseen data for an unbiased evaluation of performance.

#### **Machine learning model evaluation**

Feature engineering, designed using the training/validation cohort and then applied to the test cohort, was used to remove irrelevant or redundant variables from the dataset. The following 10 models were tested for predictive performance using Python: linear support vector classifier (SVC), XgBoost classifier, multi-layer perceptron classifier, logistic regression, decision tree classifier,

gradient-boosting classifier, random forest classifier, SVC-radial basis function kernel, Gaussian naive bayes, and XgBoost-Dart classifier. A nested cross-validation process was used to tune the hyperparameters (training/validation patient dataset). The model with parameters that performed the best (based on area under the receiver operating characteristic curve [AUROC] score) was selected. A confusion matrix was generated for the final model selected to summarize the performance of the machine learning algorithm. A false negative indicated that the model predicted discontinuation of abatacept treatment, when in fact treatment continued. As such, the negative predictive value (NPV) was reviewed as part of a negative rate analysis; no positive predictive value analysis was done. To control over- or under-fitting of models, stratified K-fold cross-validation (value of K was 5) was used to tune the model, find the hyperparameters, and keep a separate test dataset for evaluating performance.

#### **Model interpretation**

SHAP, a mathematical framework used to interpret machine learning models [15] that has been applied in other RA studies [16–18], was used to describe the effect of independent variables. The SHAP values described how much value each characteristic provided for predicting abatacept retention and indicated directionality at a patient level to indicate whether a characteristic positively or negatively impacted retention (eg, higher SHAP values indicated a higher likelihood of retention).

#### **Statistical analysis**

Baseline demographic and disease characteristics data are shown as mean (standard deviation) or proportions. Models were evaluated based on prediction quality metrics including training mean accuracy, validation mean accuracy, testing accuracy, precision, recall, F1-score, and AUROC.

Imputation was done after the dataset was split into training/validation and test cohorts; models were trained and applied using the training/validation cohort and then applied to the test cohort. Patients in the training/validation cohort who had no missing values for any variables comprised the ‘full set’; data from the full set was used in both the training/validation and test cohorts for calculation of imputed values for patient records with values missing for at least one variable. Missing baseline patient-level data were imputed by comparison of the record with missing values to similar records (based on other common variables) to fill in the missing value [19]. The median value was used to complete the missing values if the prior step failed.

Statistical analyses and model training were performed using Python programming software (version 3.8.13; packages used are listed in Supplementary Table 1).

### Ethical approval

This study was conducted in accordance with International Society for Pharmacoepidemiology Guidelines for Good Pharmacoepidemiology Practices [20] and applicable regulatory requirements. The ACTION and ASCORE study protocols and patient enrolment materials were approved according to local law in each participating country prior to initiation of each study.

## Results

### Dataset and study population

The pooled ACTION/ASCORE dataset included 5320 patients. Baseline characteristics of the study population are shown in Table 1. Of 5320 patients, duration of abatacept treatment was specified for  $n=5273$ . The 12-month abatacept retention rate was 61% ( $n=3236$ )

**Table 1** Baseline characteristics of the combined ACTION/ASCORE study population

| Variable                         | Overall study population (N = 5320) |
|----------------------------------|-------------------------------------|
| Age, years                       | 57.7 (12.7), $n=5315$               |
| Female, %                        | 79.0, $n=4202$                      |
| BMI, kg/m <sup>2</sup>           | 27.1 (5.7), $n=5021$                |
| RA duration, years               | 11.1 (9.4), $n=5283$                |
| TJC28                            | 9.3 (6.9), $n=5065$                 |
| SJC28                            | 6.5 (5.3), $n=5084$                 |
| DAS28 (CRP)                      | 4.8 (1.1), $n=4099$                 |
| DAS28 (ESR)                      | 5.3 (1.3), $n=3996$                 |
| Physician Global Assessment, VAS | 58.1 (19.3), $n=4536$               |
| Patient Global Assessment, VAS   | 61.7 (21.5), $n=4465$               |
| Patient pain, VAS                | 61.7 (22.0), $n=4458$               |
| HAQ-DI                           | 1.4 (0.7), $n=3891$                 |

All data are mean (SD) unless specified

**BMI** body mass index, **CRP** C-reactive protein, **DAS28** Disease Activity Score in 28 joints, **ESR** erythrocyte sedimentation rate, **HAQ-DI** Health Assessment Questionnaire-disability index, **RA** rheumatoid arthritis, **SD** standard deviation, **SJC28** swollen joint count in 28 joints, **TJC28** tender joint count in 28 joints, **VAS** visual analog scale

(Table 2). Initially, 114 variables were identified for the pooled dataset, and, after combining similar variables, a maximum number of 80 variables remained. For the 12-month retention model, after removing post-baseline, duplicate, or non-informative variables, and combining similar variables, 36 variables remained. For this model, in the training/validation cohort ( $n=4218$ ), samples for 2713 patients (992 and 1721 for the 2 training sets) had no missing values and did not require any imputation (full set).

### Machine learning models

Training and validation of the machine learning models were performed to optimize models and to select the top-performing models from the 10 originally assessed, for fine tuning (Supplementary Table 2). Six models were selected and fine tuned; of these, the gradient-boosting classifier model demonstrated the highest prediction testing accuracy (62%) and was then evaluated in the test cohort of patients. The baseline characteristics of the training/validation and testing cohorts for the 12- and 6-month models were similar and are shown in Table 3. In the test cohort, the gradient-boosting classifier model displayed an AUROC (95% confidence interval [CI]) of 0.620 (0.586, 0.653) for 12-month retention and 0.621 (0.580, 0.663) for 6-month retention. Further model details are shown in Supplementary Table 2. The NPV rate was higher than the negative rate for both the 12-month (0.49 vs negative rate 0.39) and 6-month (0.29 vs negative rate 0.22) retention.

Overall, the baseline characteristics of the training/validation and testing cohorts for the additional 6- and 3-month retention models were similar and are shown in Supplementary Table 3. AUROC (95% CI) values for the additional models in the test cohort were as follows: 3-month retention after 3-month follow-up, 0.708 (0.661, 0.751); 6-month retention after 3-month follow-up, 0.644 (0.606, 0.682); 3-month retention after 6-month follow-up, 0.690 (0.636, 0.740); and 6-month retention after 6-month follow-up, 0.640 (0.594, 0.683). Further model details are shown in Supplementary Table 4. NPV and negative rates for the additional analyses were as follows: 3-month retention after 3-month follow-up, 0.22 and 0.13, respectively; 6-month retention after 3-month follow-up, 0.32 and 0.23, respectively; 3-month retention

**Table 2** Treatment retention and discontinuation rates for the ACTION/ASCORE study population

| Variable                            | Overall study population (N = 5320) | Training/ validation cohort (n = 4218) | Test cohort (n = 1055) |
|-------------------------------------|-------------------------------------|--|------------------------|
| 1-year abatacept retention, % (n/N) | 61.4 (3236/5273)                    | 61.4 (2589/4218)                       | 61.3 (647/1055)        |
| Discontinuations, % (n/N)           | 38.6 (2037/5273)                    | 38.6 (1629/4218)                       | 38.7 (408/1055)        |

**Table 3** Key baseline data comparisons of the training/validation and test cohorts for 12- and 6-month<sup>a</sup> models

| Variable                         | Training/validation (n = 4218) | Test (n = 1055)       | P value |
|----------------------------------|--------------------------------|-----------------------|---------|
| Age, years                       | 57.7 (12.7), n = 4214          | 57.7 (12.4), n = 1055 | 0.829   |
| Female, %                        | 79.1, n = 3337                 | 78.7, n = 830         | 0.723   |
| BMI, kg/m <sup>2</sup>           | 27.1 (5.7), n = 3973           | 27.1 (5.7), n = 1006  | 0.989   |
| RA duration, years               | 11.1 (9.3), n = 4184           | 11.2 (9.8), n = 1053  | 0.590   |
| TJC28                            | 9.4 (6.8), n = 4039            | 9.3 (7.1), n = 998    | 0.720   |
| SJC28                            | 6.5 (5.4), n = 4052            | 6.3 (5.2), n = 1003   | 0.253   |
| DAS28 (CRP)                      | 4.8 (1.1), n = 3264            | 4.8 (1.1), n = 815    | 0.678   |
| DAS28 (ESR)                      | 5.2 (1.3), n = 3194            | 5.3 (1.3), n = 785    | 0.282   |
| Physician Global Assessment, VAS | 58.0 (19.4), n = 3600          | 58.8 (19.2), n = 908  | 0.254   |
| Patient Global Assessment, VAS   | 61.5 (21.6), n = 3548          | 62.4 (21.3), n = 895  | 0.234   |
| Patient pain, VAS                | 61.4 (22.0), n = 3544          | 62.8 (21.6), n = 891  | 0.085   |
| HAQ-DI                           | 1.4 (0.7), n = 3101            | 1.4 (0.7), n = 768    | 0.492   |

All data are mean (SD) unless specified. P values for continuous variables are from a t-test and for categorical variables from a chi-squared test

<sup>a</sup> 6-month retention from baseline

BMI body mass index, CRP C-reactive protein, DAS28 Disease Activity Score in 28 joints, ESR erythrocyte sedimentation rate; HAQ-DI Health Assessment Questionnaire-disability index, RA rheumatoid arthritis, SD standard deviation, SJC28 swollen joint count in 28 joints, TJC28 tender joint count in 28 joints, VAS visual analog scale

after 6-month follow-up, 0.26 and 0.14, respectively; and 6-month retention after 6-month follow-up, 0.55 and 0.24, respectively.

#### Prediction of abatacept retention at 12 months

The most important predictive variables identified for retention at 12 months are shown in Fig. 1A and were selected for examining directionality using SHAP values (Fig. 2A). The top five baseline predictors of abatacept retention at 12 months were: low (vs high) body mass index (BMI), low (vs high) ACR functional status class, positive (vs negative) ACPA serostatus, low (vs high) Patient Global Assessment, and younger (vs older) age.

Figure 3 shows individual SHAP plots for the top five baseline predictors of abatacept retention at 12 months. Baseline BMI displayed a non-linear association with retention, revealing that patients with either a lower BMI or higher BMI demonstrated lower retention over 12 months. ACR functional status showed a negative trend with retention, with low baseline functional status class being predictive of treatment retention and high baseline ACR functional status class associated with a lower chance of abatacept retention. A positive trend for ACPA positivity and treatment retention was observed: lower ACPA serostatus was associated with lower chance of retention and higher ACPA serostatus was associated with a greater chance for retention. A lower (vs higher) Patient Global Assessment score predicted abatacept retention at 12 months. The association between age and retention also showed a non-linear trend. Among younger patients, increasing age was associated with

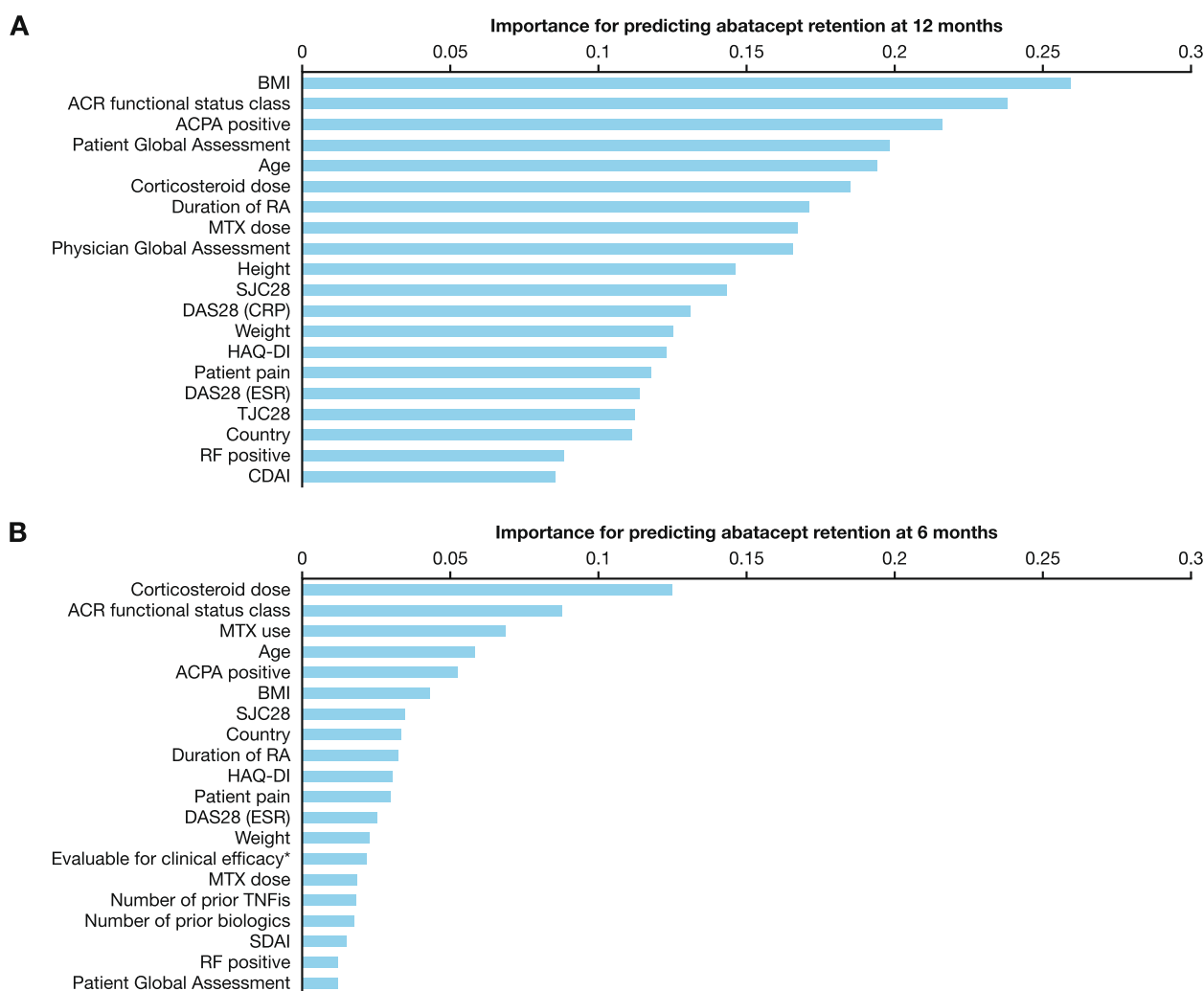
higher retention and was similar through middle-aged patients. However, there was a negative trend among older patients, revealing that as patient age increased, treatment retention decreased. Patients who were on no or low-dose corticosteroids at baseline had a higher probability of retention, and there was no association with retention as the dose increased (Supplementary Fig. 1). Additionally, shorter (vs longer) RA disease duration predicted abatacept retention at 12 months.

#### Prediction of abatacept retention at 6 months

The most important predictive variables identified for retention at 6 months are shown in Fig. 1B and were selected for examining directionality using SHAP values (Fig. 2B). The top five baseline predictors of abatacept retention at 6 months were: low (vs high) corticosteroid dose, low (vs high) ACR functional status class, receiving (vs not receiving) MTX, younger (vs older) age, and positive (vs negative) ACPA status.

Figure 4 shows individual SHAP plots for the top five baseline predictors of abatacept retention at 6 months. The lower the dose of corticosteroid at baseline, the higher the SHAP value, indicating a negative trend; therefore, patients either not receiving or receiving a smaller dose at baseline may be more likely to retain abatacept after 6 months. There was no trend as the dose increased. ACR functional status showed a negative trend with retention; thus, a lower baseline ACR functional status class was associated with a higher chance of retention and a higher functional status class was associated with a lower chance of abatacept retention over 6





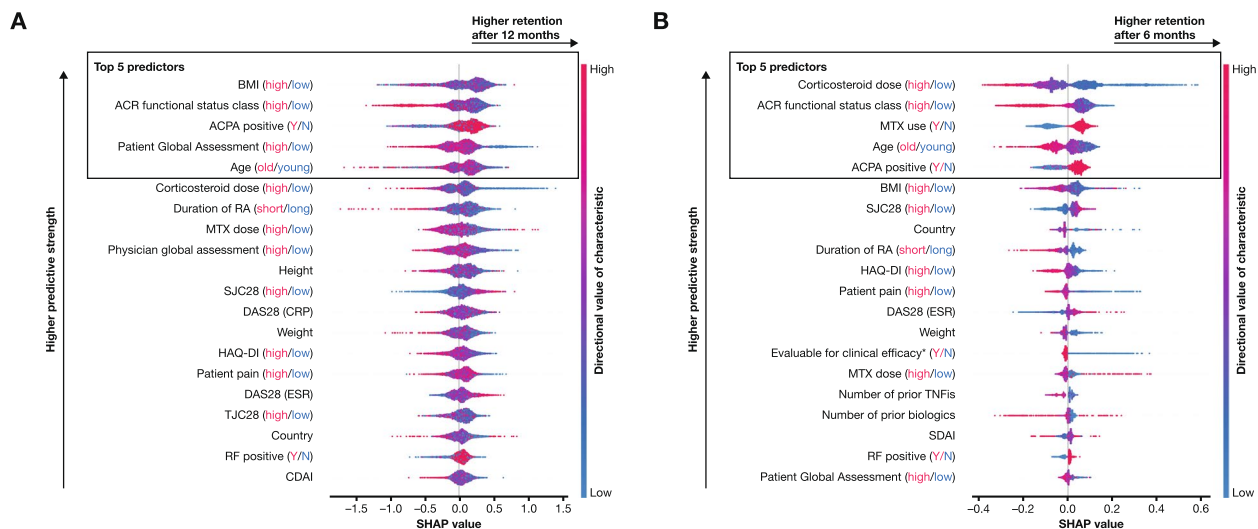
**Fig. 1** Gradient-boosting classifier variable importance for predicting abatacept retention at (a) 12 and (b) 6 months. The numerical values of variable importance in a gradient-boosting classifier provide insights into the relevance of each variable within the model's decision-making process; for example, the higher the numerical values of variable importance, the more relevant the variable (ie, corticosteroid dose) to the outcome (ie, predicting abatacept retention). \*Evaluable for secondary analysis of clinical efficacy. ACPA anti-citrullinated protein antibody, ACR American College of Rheumatology, BMI body mass index, CDAI Clinical Disease Activity Index, CRP C-reactive protein, DAS28 Disease Activity Score in 28 joints, ESR erythrocyte sedimentation rate, HAQ-DI Health Assessment Questionnaire-disability index, MTX methotrexate, RA rheumatoid arthritis, RF rheumatoid factor, SDAI Simplified Disease Activity Index, SJC28 swollen joint count in 28 joints, TJC28 tender joint count in 28 joints, TNFi tumor necrosis factor inhibitor

months. Receiving MTX at baseline was associated with a higher chance of abatacept retention at 6 months. The SHAP plot of age at baseline had two phases: firstly, there was an overall negative non-linear trend with retention with older patients more likely to retain abatacept; secondly, as baseline age increased, patients were less likely to retain abatacept. Being positive for ACPA at baseline was associated with a higher chance of retention. Additionally, BMI showed a non-linear trend with a general negative trend showing less retention with a higher BMI; there were outliers for very high and very low BMI

(Supplementary Fig. 2). A lower baseline SJC28 value was associated with a lower chance of abatacept retention, while a higher baseline SJC28 value showed a higher chance of retention. Last, RA disease duration was associated with abatacept retention at 6 months, particularly among patients with shorter disease duration.

#### Additional models for abatacept retention at 3 and 6 months

Four additional models were validated for assessing retention of abatacept either 3 months after or 6 months



**Fig. 2** Overall SHAP value plot after (a) 12 months and (b) 6 months. Figure shows the most important variables in the ACTION/ASCORE combined dataset that were predictive of abatacept retention. Colors indicate the value of the variable: red represents higher numerical values of the variable and blue represents lower numerical values. The bulges in the plot indicate more patients with that value; each dot represents a single patient. Higher SHAP values indicate a higher likelihood of retention. \*Evaluable for secondary analysis of clinical efficacy. ACPA anti-citrullinated protein antibody, ACR American College of Rheumatology, BMI body mass index, CDAI Clinical Disease Activity Index, CRP C-reactive protein, DAS28, Disease Activity Score in 28 joints, ESR erythrocyte sedimentation rate, HAQ-DI Health Assessment Questionnaire-disability index, MTX methotrexate, RA rheumatoid arthritis, RF rheumatoid factor, SDAI Simplified Disease Activity Index, SHAP SHapley Additive exPlanation, SJC28 swollen joint count in 28 joints, TJC28 tender joint count in 28 joints, TNFi tumor necrosis factor inhibitor

after either a 3-month follow-up or a 6-month follow-up period. Characteristics of the models assessed are shown in Supplementary Table 3. Baseline characteristics of the training/validation and test cohorts were similar and are shown in Supplementary Table 4.

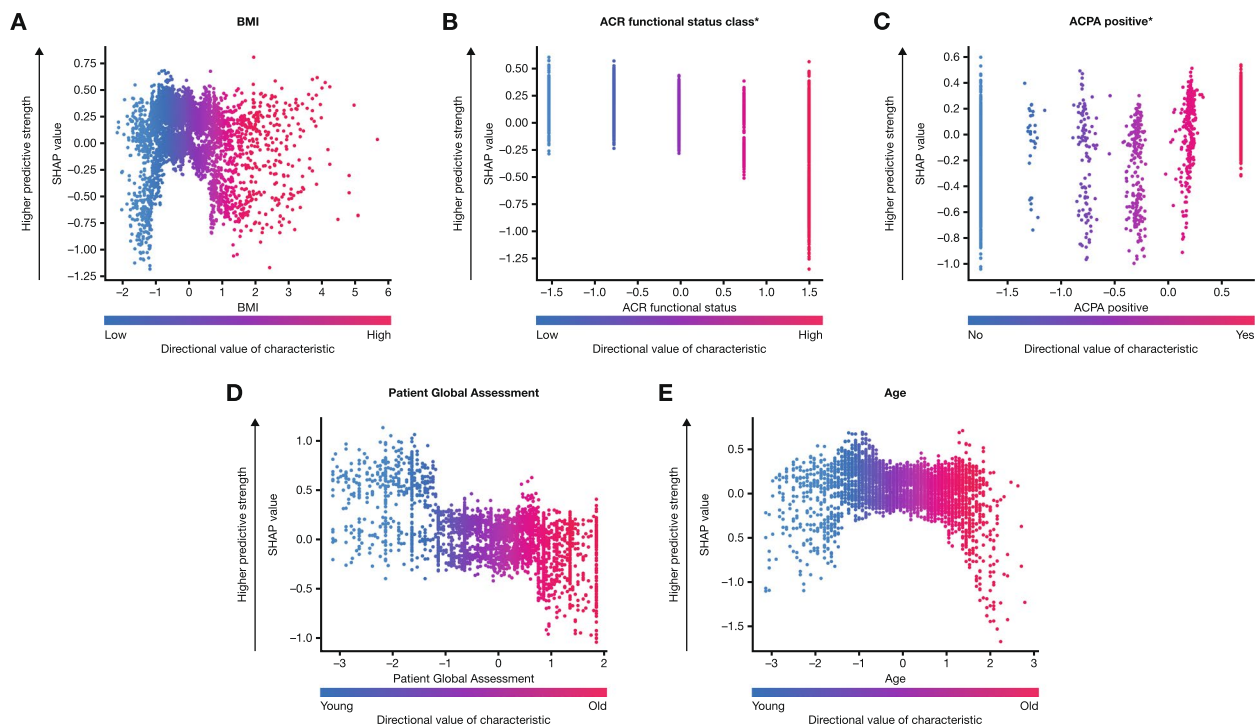
## Discussion

This post hoc, machine learning analysis of pooled data from the ACTION and ASCORE trials identified patient characteristics that were associated with abatacept retention at both 12 and 6 months: ACPA positivity, lower BMI, lower ACR functional status class, lower corticosteroid dose, and younger age. Additionally, a lower baseline Patient Global Assessment score was associated with higher chance of abatacept retention at 12 months, and patients currently receiving combination therapy (MTX) had better retention of abatacept at 6 months.

In line with the treat-to-target approach for the management of RA [1, 2], some of the patient characteristics identified in this analysis (including younger age) may illustrate the value of starting abatacept treatment early for patients with RA. Further, these results support previous observations that have demonstrated that patients with early RA and specific variables, such as ACPA positivity or low BMI, are predictive of abatacept retention and response. Results from the Assessing Very Early Rheumatoid arthritis Treatment (AVERT) trials

[21, 22] demonstrated the value of timely intervention with abatacept (vs MTX alone) for patients with early RA. Additionally, observations from the ACTION and ASCORE trials emphasize the impact that seropositivity has on clinical response among patients with RA receiving abatacept [11, 12]. Results shown here are consistent with those from the AVERT, ACTION, and ASCORE trials and lend support for the potential to develop a patient phenotype or patient stratification tool to identify those likely to respond favorably to abatacept treatment. Importantly, such variables may already be part of routine patient visits or could be implemented during routine patient visits, so they could provide a generalizable and scalable approach for implementation. For example, treating patients with RA as early as possible and targeting patients with specific clinical characteristics (eg, ACPA positivity) may not only improve treatment retention but may also augment clinical response.

Evidence demonstrating the value of machine learning among patients with RA is accumulating, with recent observations showing the ability to predict patient response to MTX [17, 23, 24], TNF inhibitors [16, 25, 26], and other biologic disease-modifying antirheumatic drugs (bDMARDs) [18, 26, 27]. Apart from our previous observations [9, 10], few data have examined abatacept treatment response among patients with RA via machine learning. Our AUROC values are similar to those



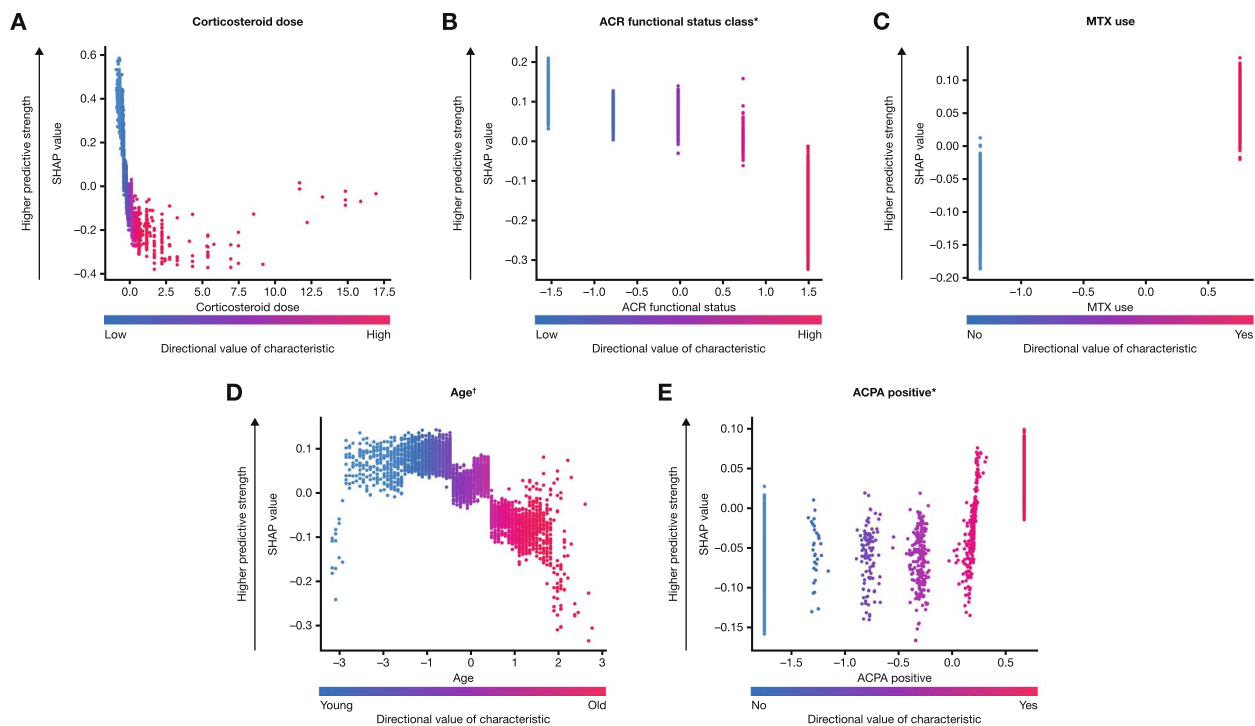
**Fig. 3** Individual SHAP value plots for top 5 characteristics predictive of abatacept 12-month retention. Colors indicate the value of the variable: red represents higher and blue represents lower. Each dot represents a single patient. Higher SHAP values indicate a higher likelihood of retention. \*For these characteristics, more columns are included here compared to the overall SHAP plot (Fig. 2) due to the step of filling missing values. Missing values are filled as predicted from available values providing a numerical output (logit); the numerical output (logit) is used to make the prediction to prevent the information loss caused by the step of transferring into binary outputs. *ACPA* anti-citrullinated protein antibody, *ACR* American College of Rheumatology, *BMI* body mass index, *SHAP* SHapley Additive exPlanations

reported by Koo et al., who predicted 12-month DAS28-ESR  $\leq 2.6$  response (AUROC, 0.598–0.679) in a Korean registry; features of importance differed between the studies. Of note, across most bDMARDs including abatacept, low ACPA levels were associated with remission, whereas our results show ACPA positivity was predictive of abatacept retention [18]; this aligns with previous ACTION and ASCORE observations [11, 12]. Potential explanations for conflicting findings may be a result of different endpoints (eg, remission vs retention) or differences in sample size. Overall, existing machine learning models have shown a good ability to predict patient response but have nuances worth mentioning. For example, final model outcomes may be affected by different disease measures, such as 20% improvement in ACR criteria response [27] versus DAS28 response [23, 26], or different follow-up intervals, such as annually [27] or less than a year [17, 24, 25], among other factors. Further, some researchers have sought to leverage data that are routinely collected in the clinical setting to build models [16, 28], whereas others have shown that the inclusion of genomic data can enhance model precision [24, 26]. Altogether, these factors are key model features that must

be considered when reviewing model performance and implementing in the clinical setting.

There was no discernible difference between the ability of the model to predict retention at 12 or at 6 months using baseline values. However, shorter retention intervals resulted in greater model performance, which was evident when predicting retention from baseline and when the additional analyses were performed within the overall follow-up timeframe (eg, 6 to 9 months). It is possible that the presence of patient features during follow-up intervals improved model performance versus missing baseline data that were imputed. Differences between AUROC and F1 score model evaluations were evident for the 12- and 6-month retention models; however, similar AUROC values were observed for each retention time point but F1 scores differed between time points. These values reflect different model qualities depending on the nature of the data, and as a result, should be selected appropriately to best describe each unique dataset. Other popular techniques such as Local Interpretable Model-agnostic Explanations and permutation importance attempt local interpretation of the model, to understand why predictions are made for individual data points or





**Fig. 4** Individual SHAP value plots for top five characteristics predictive of abatacept 6-month retention. Colors indicate the value of the variable: red represents higher and blue represents lower. Each dot represents a single patient. Higher SHAP values indicate a higher likelihood of retention. \*For these characteristics, more columns are included here compared to the overall SHAP plot (Fig. 2) due to the step of filling missing values. Missing values are filled as predicted from available values providing a numerical output (logit); the numerical output (logit) is used to make the prediction to prevent the information loss caused by the step of transferring into binary outputs. †Breaks or steps in the SHAP plots likely result from use of the gradient boosting classifier, a decision tree-based model, which uses age at certain values to split the tree decisions in the model. ACPA anti-citrullinated protein antibody, ACR American College of Rheumatology, MTX methotrexate, SHAP SHapley Additive exPlanations

instances. Such local approaches do not guarantee globally consistent interpretation of overall behavior and patterns across the entire dataset, which is important for understanding a model's general tendencies and variable importance values. However, consistent interpretation is relevant here to understand how different clinical and non-clinical factors influence retention behavior at a population level while leveraging sophisticated non-linear machine learning models.

This study has notable strengths, such as the merging of the ACTION and ASCORE databases that ultimately provided one of the largest global datasets to date of patients with RA from multiple real-world centers. Additionally, this study highlights the usefulness of machine learning for predicting abatacept retention in patients with RA and revealed key patient features that may be used to form profiles identifying patients likely to respond favorably to abatacept treatment. This study also has some limitations that should be acknowledged. First, these results reflect the current model and the data it has learned from, which may not accurately reflect the precise relationship between the variables and outcomes

in the real world. The totality of missing baseline data resulted in values needing imputation using the median value from a similar record; however, due to complexity of the data flow and calculations, it was not possible to precisely quantify this in a meaningful way. Obtained model performance was relatively limited (AUROC values in the range 0.62–0.71). Although the pooled database comprised two international trials, these data do not contain patients currently residing in all countries; thus, there is the potential for geographical differences or biases in treatment strategies that may not be generalized or practiced globally. Also, findings show that despite merging data from two studies, datasets of this relatively small magnitude ( $n = 5320$ ) do not improve model performance from our previous ACTION analyses [9, 10]. This may underscore the need for much larger datasets when implementing machine learning approaches, but it is also possible that other important covariates are missing. As we did not generate learning curves of the best performing model, we were unable to assess the effect, if any, of increasing the number of datapoints presented to the learning algorithm; lack of correlation between increased

training data and model performance may indicate that other covariates important for retention prediction were missing from the data. Route of administration was not a stratification factor for assessing imbalance between the cohorts. No specific measures of patient adherence were included as covariates. Additionally, model evaluation for false negatives may have financial implications. In such instances, the model predicts that a patient is no longer using abatacept when they actually are. These results demonstrated that the inclusion of more follow-up tests and predicting longer retention resulted in higher NPV and each of the calculated values were greater than the negative rate.

Next steps in this line of research should seek to refine existing approaches that provide improved model outcomes. For example, collecting additional data may compensate for missing data and may also reduce concerns associated with unbalanced data. Furthermore, having larger datasets would provide the models with more data-points to learn from. Alternatively, leveraging existing models as an avenue for transfer learning (where a previously developed model provides the starting point for a future model) on alternative datasets may help improve model performance. Continuing collaborative efforts with rheumatologists and clinical researchers to identify patient features that may have been overlooked in the present or previous studies should enhance development of predictive models and their corresponding prognostic ability. Ultimately, machine learning approaches have the potential to identify distinct characteristics associated with treatment response to specific bDMARDs and may help guide treatment strategies to achieve the best individual patient outcomes.

## Conclusions

The gradient-boosting classifier identified predictors of abatacept retention over 12 months from the pooled ACTION and ASCORE study populations. By including SHAP analyses, both the directionality and importance of patient features from a large, real-world study population were identified. Some of the identified patient characteristics (including younger age) are indicative of early-stage RA disease status, a time in which therapeutic intervention is essential. In sum, the factors predictive of abatacept retention found in this machine learning analysis are consistent with those previously shown and help further validate the machine learning approach for predictive modelling in RA treatment, and may help inform clinical decision making.

## Abbreviations

|        |  |
|--------|--|
| ACPA   | Anti-citrullinated protein antibody    |
| ACR    | American College of Rheumatology       |
| ACTION | AbataCepT In rOutiNe clinical practice |

|        |  |
|--------|--|
| ASCORE | Abatacept SubCutaneOus in Routine clinical practice    |
| AUROC  | Area under the receiver operating characteristic curve |
| AVERT  | Assessing Very Early Rheumatoid arthritis Treatment    |
| bDMARD | Biologic disease-modifying antirheumatic drug          |
| BMI    | Body mass index  |
| CDAI   | Clinical Disease Activity Index                        |
| CI     | Confidence interval                                    |
| CRP    | C-reactive protein                                     |
| DAS28  | Disease Activity Score in 28 joints                    |
| ESR    | Erythrocyte sedimentation rate                         |
| EULAR  | European Alliance of Associations for Rheumatology     |
| HAQ-DI | Health Assessment Questionnaire-disability index       |
| MTX    | Methotrexate   |
| NPV    | Negative predictive value                              |
| RA     | Rheumatoid arthritis                                   |
| RF     | Rheumatoid factor                                      |
| SD     | Standard deviation                                     |
| SDAI   | Simplified Disease Activity Index                      |
| SHAP   | SHapley Additive exPlanation                           |
| SJC28  | Swollen joint count in 28 joints                       |
| SVC    | Support vector classifier                              |
| TJC28  | Tender joint count in 28 joints                        |
| TNF    | Tumor necrosis factor                                  |
| VAS    | Visual analog scale                                    |

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13075-025-03484-0>.

Supplementary Material 1.

## Acknowledgements

Professional medical writing and editorial assistance was provided by Rachel Rankin, PhD, and Ryan Miller at Caudex, a division of IPG Health Medical Communications, and was funded by Bristol Myers Squibb. The results of this study were presented at ACR Convergence 2020 (5–9 November 2020; presentation number: 1745), the EULAR Annual European Congress of Rheumatology 2021 (2–5 June 2021; abstract number: AB0205), and ACR Convergence 2021 (1–10 November 2021; presentation number: 1212).

## Authors' contributions

C.B., K.L., E.A., P.M., V.V.-M., Y.E., and A.R. conceived and designed the study; S.E.C. acquired data; C.B., K.L., G.L., E.A., and R.A. analyzed data; C.B., K.L., S.E.C., G.L., E.A., R.A., P.-A.J., and A.N. interpreted the data. 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.

## Funding

This study was sponsored by Bristol Myers Squibb.

## Data availability

Bristol Myers Squibb policy on data sharing may be found at <https://www.bms.com/researchers-and-partners/independent-research/data-sharing-request-process.html>.

## Declarations

### Ethics approval and consent to participate

This study was conducted in accordance with International Society for Pharmacoepidemiology Guidelines for Good Pharmacoepidemiology Practices [20] and applicable regulatory requirements. The ACTION and ASCORE study protocols and patient enrolment materials were approved according to local law in each participating country prior to initiation of each study.

### Consent for publication

Not applicable.

## Competing interests

RA: advisory role: AbbVie, Amgen, Biogen, Bristol Myers Squibb, Celltrion, Eli Lilly, Galapagos, Gilead, Janssen, Novartis, Pfizer, Roche.  
CBe: former consultant: Bristol Myers Squibb.  
PM, EA: nothing to disclose.  
VV-M, AO, GL: employee: Bristol Myers Squibb.  
SEC, KL: employee, shareholder: Bristol Myers Squibb.  
AN: advisory role: Bristol Myers Squibb, UCB; speaker/honoraria: Bristol Myers Squibb, Galapagos, Roche, UCB; grant/research support: Novartis, AstraZeneca.  
P-AJ: consultant: Bristol Myers Squibb; speaker/honoraria: AstraZeneca, Boehringer Ingelheim.  
AR: employee, shareholder: Amgen, Bristol Myers Squibb (at the time of analysis); shareholder: Genmab. Current affiliation: Eli Lilly & Co.  
YE: consultant: Bristol Myers Squibb.

## Author details

<sup>1</sup>Schlosspark-Klinik University, Berlin, Germany. <sup>2</sup>Tulsi, Paris, France. <sup>3</sup>Data Revenue GmbH, Berlin, Germany. <sup>4</sup>Private Practice, Lyon, France. <sup>5</sup>Bristol Myers Squibb, Rueil-Malmaison, France. <sup>6</sup>Bristol Myers Squibb, Princeton, NJ, USA. <sup>7</sup>School of Infection and Immunity, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. <sup>8</sup>Université de Paris, AP-HP, Hôpital Bichat Claude-Bernard, Paris, France. <sup>9</sup>Significance, Puteaux, France.

Received: 10 July 2024 Accepted: 16 January 2025

Published online: 01 February 2025

## References

- Fraenkel L, Bathon JM, England BR, St Clair EW, Arayssi T, Carandang K, et al. 2021 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis Rheumatol*. 2021;73:1108–23.
- Smolen JS, Landewe RBM, Bijlsma JWJ, Burmester GR, Dougados M, Kerschbaumer A, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann Rheum Dis*. 2020;79:685–99.
- Ebina K, Hashimoto M, Yamamoto W, Hirano T, Hara R, Katayama M, et al. Drug tolerability and reasons for discontinuation of seven biologics in elderly patients with rheumatoid arthritis -The ANSWER cohort study. *PLoS ONE*. 2019;14:e0216624.
- Liu PC, Ssu CT, Tsao YP, Liou TL, Tsai CY, Chou CT, et al. Cytotoxic T lymphocyte-associated antigen-4-Ig (CTLA-4-Ig) suppresses *Staphylococcus aureus*-induced CD80, CD86, and pro-inflammatory cytokine expression in human B cells. *Arthritis Res Ther*. 2020;22:64.
- Bristol Myers Squibb. Ocrencia (abatacept) prescribing information. 2024.
- Alten R, Mariette X, Lorenz HM, Galeazzi M, Cantagrel A, Nusslein HG, et al. Real-world predictors of 12-month intravenous abatacept retention in patients with rheumatoid arthritis in the ACTION observational study. *RMD Open*. 2017;3:e000538.
- Alten R, Mariette X, Flipo RM, Caporali R, Buch MH, Patel Y, et al. Retention of subcutaneous abatacept for the treatment of rheumatoid arthritis: real-world results from the ASCORE study: An international 2-year observational study. *Clin Rheumatol*. 2022;41:2361–73.
- Rubinger L, Gazendam A, Ekhtiari S, Bhandari M. Machine learning and artificial intelligence in research and healthcare. *Injury*. 2023;54(Suppl 3):S69–73.
- Alten R, Behar C, Boileau C, Merckaert P, Afari E, Vannier-Moreau V, et al. A novel method for predicting 1-year retention of abatacept using machine learning techniques [abstract]. *Ann Rheum Dis*. 2021;80:AB0205.
- Alten R, Behar C, Boileau C, Merckaert P, Afari E, Vannier-Moreau V, et al. Prediction of 1-year intravenous abatacept retention in patients with RA using novel machine learning techniques: directionality and importance of predictors [abstract]. *Arthritis Rheumatol*. 2021;73:S9.
- Alten R, Rauch C, Chartier M, Nurmohamed M, Connolly S, Buch MH, et al. ACPA positivity determines remission in patients with RA treated with IV and SC abatacept: a post hoc analysis of the real-world observational ACTION and ASCORE studies. *Ann Rheum Dis*. 2022;81:POS0107.
- Alten R, Rauch C, Chartier M, Nurmohamed M, Connolly S, Buch MH, et al. Anti-citrullinated protein antibody serostatus determines 2-year retention of IV and SC abatacept in patients with RA in a real-world setting. *Ann Rheum Dis*. 2022;81:POS0512.
- Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. 2010 rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann Rheum Dis*. 2010;69:1580–8.
- Fransen J, van Riel PL. The Disease Activity Score and the EULAR response criteria. *Rheum Dis Clin North Am*. 2009;35:745–57.
- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30:4765–74.
- Bouget V, Duquesne J, Hassler S, Courneade PH, Fautrel B, Guillemin F, et al. Machine learning predicts response to TNF inhibitors in rheumatoid arthritis: results on the ESPOIR and ABIRISK cohorts. *RMD Open*. 2022;8:e002442.
- Duquesne J, Bouget V, Courneade PH, Fautrel B, Guillemin F, de Jong PHP, et al. Machine learning identifies a profile of inadequate responder to methotrexate in rheumatoid arthritis. *Rheumatology (Oxford)*. 2023;62:2402–9.
- Koo BS, Eun S, Shin K, Yoon H, Hong C, Kim DH, et al. Machine learning model for identifying important clinical features for predicting remission in patients with rheumatoid arthritis treated with biologics. *Arthritis Res Ther*. 2021;23:178.
- Jazayeri A, Liang OS, Yang CC. Imputation of missing data in electronic health records based on patients' similarities. *J Healthc Inform Res*. 2020;4:295–307.
- International Society for Pharmacoepidemiology. Guidelines for Good Pharmacoepidemiology Practices (GPP). *Pharmacoepidemiol Drug Saf*. 2016;25:2–10.
- Emery P, Burmester GR, Bykerk VP, Combe BG, Furst DE, Barre E, et al. Evaluating drug-free remission with abatacept in early rheumatoid arthritis: results from the phase 3b, multicentre, randomised, active-controlled AVERT study of 24 months, with a 12-month, double-blind treatment period. *Ann Rheum Dis*. 2015;74:19–26.
- Emery P, Burmester GR, Bykerk VP, Combe BG, Furst DE, Maldonado MA, et al. Re-treatment with abatacept plus methotrexate for disease flare after complete treatment withdrawal in patients with early rheumatoid arthritis: 2-year results from the AVERT study. *RMD Open*. 2019;5:e000840.
- Duong SQ, Crowson CS, Athreya A, Atkinson EJ, Davis JM 3rd, Warrington KJ, et al. Clinical predictors of response to methotrexate in patients with rheumatoid arthritis: a machine learning approach using clinical trial data. *Arthritis Res Ther*. 2022;24:162.
- Myasoedova E, Athreya AP, Crowson CS, Davis JM 3rd, Warrington KJ, Walchak RC, et al. Toward individualized prediction of response to methotrexate in early rheumatoid arthritis: a pharmacogenomics-driven machine learning approach. *Arthritis Care Res (Hoboken)*. 2022;74:879–88.
- Tao W, Concepcion AN, Vianen M, Marijnissen ACA, Lafefber F, Radstake T, et al. Multiomics and machine learning accurately predict clinical response to adalimumab and etanercept therapy in patients with rheumatoid arthritis. *Arthritis Rheumatol*. 2021;73:212–22.
- Guan Y, Zhang H, Quang D, Wang Z, Parker SCJ, Pappas DA, et al. Machine learning to predict anti-tumor necrosis factor drug responses of rheumatoid arthritis patients by integrating clinical and genetic markers. *Arthritis Rheumatol*. 2019;71:1987–96.
- Lee S, Kang S, Eun Y, Won HH, Kim H, Lee J, et al. Machine learning-based prediction model for responses of bDMARDs in patients with rheumatoid arthritis and ankylosing spondylitis. *Arthritis Res Ther*. 2021;23:254.
- Norgeot B, Glicksberg BS, Trupin L, Lituiev D, Gianfrancesco M, Oskotsky B, 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:e190606.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.