ORIGINAL RESEARCH

A Machine Learning Approach for Prediction of CDAI Remission with TNF Inhibitors: A Concept of Precision Medicine from the FIRST Registry

Koshiro Sonomoto · Yoshihisa Fujino · Hiroaki Tanaka ·

Atsushi Nagayasu · Shingo Nakayamada · Yoshiya Tanaka

ABSTRACT

Introduction: This study aimed to develop low-cost models using machine learning approaches predicting the achievement of Clinical Disease Activity Index (CDAI) remission 6 months after initiation of tumor necrosis factor inhibitors (TNFi) as primary biologic/targeted synthetic disease-modifying antirheumatic drugs (b/tsDMARDs) for rheumatoid arthritis (RA).

Methods: Data of patients with RA initiating TNFi as first b/tsDMARD after unsuccessful methotrexate treatment were collected from the FIRST registry (August 2003 to October 2022). Baseline characteristics and 6-month CDAI were collected. The analysis used various machine learning approaches including logistic regression with stepwise variable selection, decision tree, support vector machine, and lasso logistic regression (Lasso), with 48 factors accessible in routine clinical practice for the prediction model. Robustness was ensured by k-fold cross validation.

Results: Among the approaches tested, Lasso showed the advantages in predicting CDAI remission: with a mean area under the curve 0.704, sensitivity 61.7%, and specificity 69.9%. Predicted TNFi responders achieved CDAI remission at an average rate of 53.2%, while only 26.4% of predicted TNFi non-responders achieved remission. Encouragingly, the models generated relied solely on patient-reported outcomes and quantitative parameters, excluding subjective physician input.

Conclusions: While external cohort validation is warranted for broader applicability, this study highlights the potential for a low-cost predictive model to predict CDAI remission following TNFi treatment. The approach of the study using only baseline data and 6-month CDAI measures, suggests the feasibility of establishing regional cohorts to generate low-cost models tailored to specific regions or institutions. This may facilitate the application of regional/in-

K. Sonomoto
Department of Clinical Nursing, School of Health Sciences, University of Occupational and Environmental Health, Japan, 1-1, Iseigaoka, Yahatanishi-ku, Kitakyushu 807-8555, Japan

K. Sonomoto · H. Tanaka · A. Nagayasu ·
S. Nakayamada · Y. Tanaka (✉)
The First Department of Internal Medicine, School of Medicine, University of Occupational and Environmental Health, Japan, 1-1, Iseigaoka, Yahatanishi-ku, Kitakyushu 807-8555, Japan
e-mail: tanaka@med.uoeh-u.ac.jp

Y. Fujino
Department of Environmental Epidemiology, Institute of Industrial Ecological Sciences, University of Occupational and Environmental Health, Japan, 1-1, Iseigaoka, Yahatanishi-ku, Kitakyushu 807-8555, Japan

△ Adis

house precision medicine strategies in RA management.

PLAIN LANGUAGE SUMMARY

This study aims to enhance the management of rheumatoid arthritis by predicting the likelihood of achieving the treatment target—Clinical Disease Activity Index remission within 6 months of initiating tumor necrosis factor inhibitors. In rheumatoid arthritis, the goal is often Clinical Disease Activity Index remission, and the standard approach involves using medications like methotrexate and biologic/targeted synthetic disease-modifying antirheumatic drugs. However, not all patients respond to these treatments, leading to a trial-and-error process of changing medications. Tumor necrosis factor inhibitors are commonly used as the initial biologic/targeted synthetic disease-modifying antirheumatic drugs for patients who do not respond adequately to methotrexate; however, tumor necrosis factor inhibitor treatment may not achieve effective outcomes for all patients. The study, using a cohort of patients with rheumatoid arthritis treated with tumor necrosis factor inhibitor, has developed a model predicting Clinical Disease Activity Index remission with tumor necrosis factor inhibitors. The models use only standard clinical parameters, therefore no special examination or additional cost is required for the predictions. This approach holds the potential to improve rheumatoid arthritis management by reducing the need for trial-and-error approaches and facilitating more personalized and effective treatment strategies. While further validation is necessary, the study also suggests that creating cost-effective models tailored to specific regions or institutions is possible.

Keywords: Rheumatoid arthritis; Tumor necrosis factor inhibitor; Machine learning; Sparse modeling; Precision medicine

Key Summary Points

Why carry out this study?

Advanced technologies are certainly prompting precision medicine in rheumatoid arthritis (RA). However, their complexity and cost may limit their use to a few advanced facilities.

We propose a concept of low-cost precision medicine using readily available information.

What was learned from the study?

A machine learning approach incorporating 48 clinical parameters effectively predicted Clinical Disease Activity Index (CDAI) remission at 6 months after initiation of tumor necrosis factor inhibitors (TNFi) in addition to methotrexate (MTX).

The generalizability of the models generated in this study needs to be confirmed. However, this study suggests that regional/institutional precision medicine is feasible.

INTRODUCTION

Advances in molecularly targeted anti-rheumatic therapies have revolutionized the management of rheumatoid arthritis (RA) over the past decades. The development of biologic disease-modifying anti-rheumatic drugs (bDMARDs) and targeted synthetic (ts)DMARDs has significantly improved disease control, making clinical remission the primary treatment objective for RA [1]. In addition, the concept of "beyond remission" has gained recognition.

Current treatment guidelines recommend initiating therapy with methotrexate (MTX) in the initial phase (phase I) of RA, unless contraindicated, followed by bDMARDs or Janus

kinase inhibitors (JAKis) in phase II if initial treatment proves ineffective [1, 2]. However, based on the findings of ORAL Surveillance [3] and subsequent developments, JAKis are becoming less favored as first-line therapy in phase II. As a result, current practice is leaning toward the use of MTX and subsequent bDMARDs. According to the recommendations, interleukin-6 receptor inhibitors (IL-6Ri) are preferred for patients not receiving MTX, while both tumor necrosis factor inhibitors (TNFi) and non-TNFi bDMARDs are considered for patients receiving MTX [1]. Although TNFi are considered the standard among bDMARDs [3–7], only about 70% of patients show a favorable response [8].

With this background, numerous efforts have been reported to predict the efficacy of TNFi in advance. These efforts have included serum proteins, single-nucleotide polymorphisms (SNPs), peripheral blood gene expressions, peripheral blood epigenomic status, peripheral blood immune cell populations (analyzed by flow cytometry), synovial histology, and comprehensive synovial gene analysis [8–19]. However, access to these advanced technologies may be limited to certain countries and advanced facilities due to cost, labor requirements, and the need for process standardization. Conversely, predictive models based on routine clinical data are more accessible. Vastesaeger et al. used clinical parameters (such as sex, age, health assessment questionnaire [HAQ], tender joint count [TJC] 28, erythrocyte sedimentation rate [ESR], and presence of complications) [20]. Their study, published in 2016, introduced a tool that predicts disease activity score 28 (DAS28) remission in response to TNFi treatment, although it is not consistent with current recommendations that advocate more stringent remission criteria, such as the Clinical Disease Activity Index (CDAI) [21, 22]. Notably, most predictive models have focused on DAS improvement, with some addressing CDAI and simplified disease activity index (SDAI) improvement [18–20]. However, to our knowledge, no model has been presented that predicts CDAI remission with TNFi.

The primary objective of this study was to develop a method for predicting CDAI remission in patients considering TNFi as phase II therapy after an inadequate response to MTX. To achieve our objective, our approach involved establishing a predictive model using our cohort of patients with RA. With a focus on low cost and accessibility, we exclusively used factors that are available in routine clinical practice. However, it is important to note that these models may not be universally applicable to other cohorts. Therefore, the concept of establishing regionally optimized models within each cohort becomes valuable, although this may involve potentially smaller sample sizes, leading to concerns about overtraining or overfitting. To address this concern, we planned to use Lasso, a machine learning approach known for its sparse modeling capabilities.

METHODS

Data Source

Subjects for this study were selected from FIRST registry, a database of patients with RA treated with b/tsDMARDs at the University Hospital of University of Occupational and Environmental Health, Japan (UOEH) and 28 medical institutions in nine municipalities. Patients who met the ACR/EULAR 2010 RA classification criteria and did not respond to phase I therapy were referred to UOEH for further evaluation of eligibility for b/tsDMARDs. This comprehensive evaluation included assessment of contraindications and consideration of comorbidities, such as screening for infectious diseases and malignancies [23]. Subsequently, the administration of the b/tsDMARDs was decided by the patients and the treating physician under the guidance of recommendations/guidelines [1, 2, 24]. Patients eligible for b/tsDMARDs underwent initiation of b/tsDMARDs treatment, provided informed consent, and were subsequently enrolled in FIRST registry. In Japan, due to the healthcare system, patients typically visit medical facilities at approximately 4-week intervals. Under the FIRST registry framework, patients administered b/tsDMARDs at UOEH subsequently will then visit nearby medical facilities for ongoing care. They will visit UOEH
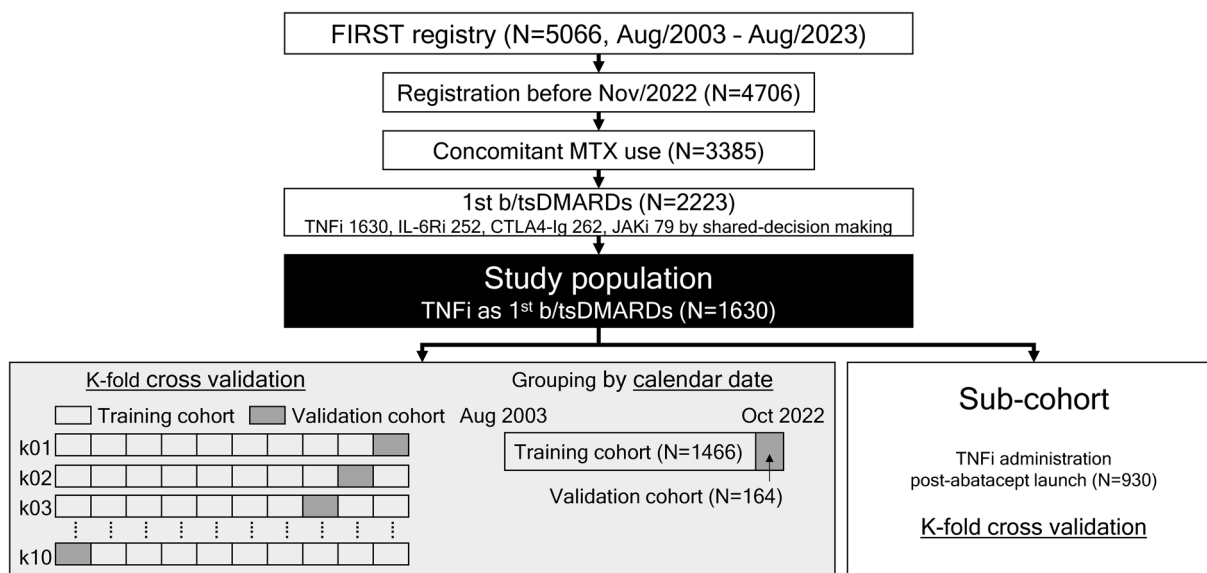
Fig. 1 Study diagram. *b/tsDMARDs* biologic/targeted synthetic disease-modifying antirheumatic drugs, *CTLA4Ig* cytotoxic T-lymphocyte-associated protein 4 immunoglobulin, *IL-6Ri* interleukin-6 receptor inhibitor, *JAKi* Janus kinase inhibitor, *MTX* methotrexate, *TNFi* tumor necrosis factor inhibitor

at 6 months, 1 year, and then annually for professional assessments. During these visits, clinical information is documented, including RA disease activity, comorbidities, and treatment status. The term b/tsDMARDs refers to TNFi, IL-6Ri, and cytotoxic T-lymphocyte-associated protein 4 immunoglobulin (CTLA4-Ig). Notably, rituximab and anakinra are not approved for the treatment of RA in Japan and are therefore not included in FIRST registry.

Selection of b/tsDMARDs for Patients on MTX in FIRST Registry

Treatment strategies for patients eligible for b/tsDMARDs after screening involve shared decision making, where options are presented considering efficacy, safety, routes of administration, and dosing intervals. Physician recommendations include appropriate vaccination and the choice of b/tsDMARDs, tailored to the patient's risk factors [24–26]. Specifically, the physician recommends pneumococcal, influenza, and herpes zoster vaccines. Since the launch of abatacept (ABT) in November 2010, it is often recommended for patients at a higher

risk of infection due to factors such as advanced age, coexisting lung disease, or concomitant glucocorticoid use.

For other patients, the options typically include TNFi (infliximab [IFX]/IFX-biosimilar [BS], etanercept [ETN]/ETN-BS, adalimumab [ADA]/ADA-BS, golimumab [GLM], certolizumab pegol [CZP]), CTLA4-Ig (ABT), IL-6Ri (tocilizumab, sarilumab) and JAKi (tofacitinib, baricitinib, peficitinib, upadacitinib, filgotinib), according to the latest recommendation[1]. Currently, JAKis are less commonly used as the first-line b/tsDMARDs. However, prior to the initial results of ORAL Surveillance trial [3], JAKis were also considered as first b/tsDMARDs, especially in highly active patients.

Handling Censoring in the Study

FIRST registry includes various medical institutions from multiple municipalities, with UOEH as the central hub. Some patients opt out of regular evaluations at UOEH due to distance or the accessibility limitations (e.g., residing in nursing facilities), resulting in censoring. Although multiple imputation was a possible

Table 1 Baseline characteristics of the b/tsDMARDs-naïve MTX users in the FIRST registry

	Total	TNFi (Study population)	IL-6Ri	CTLA4Ig	JAKi
N	2223	1630	252	262	79
Female (%)	80.3	80.3	77.4	85.5	73.4
Age	59.2 ± 14.4	57.9 ± 14.2	61.6 ± 14.6	66.0 ± 13.5	56.3 ± 13.2
Smoking (Never/Current/Past, %)	72/14/14	75/11/14	64/24/12	69/17/14	59/39/2
BMI	22.1 ± 3.7	22.1 ± 3.6	22.6 ± 4.0	22 ± 3.7	22.5 ± 4.8
eGFR (ml/min/1.73 m²)	88.7 ± 24.2	91.3 ± 24.0	82.5 ± 24.9	79.0 ± 22.9	87.2 ± 17.8
Rheumatoid arthritis					
RA duration (months)	75.7 ± 106.1	72.7 ± 100.7	74.2 ± 116.0	101.3 ± 129.6	57.0 ± 80.2
Steinbrocker stage (I/II/III/IV, %)	29/45/14/12	29/46/14/11	34/42/14/10	23/45/15/17	37/45/8/10
Morning stiffness (min)	167.0 ± 264.4	149.2 ± 216.4	262.1 ± 429.1	184.2 ± 315.4	174.4 ± 232.2
Tender joint count (28)	9.0 ± 6.6	9.3 ± 6.8	8.0 ± 6.1	7.7 ± 5.7	7.9 ± 5.7
Swollen joint count (28)	7.9 ± 5.5	8.2 ± 5.6	7.6 ± 5.5	6.5 ± 4.6	7.3 ± 4.1
Evaluator's global assessment (mm)	45.6 ± 20.8	46.5 ± 21.3	45.4 ± 20.3	40.9 ± 18.4	48.3 ± 20.5
Patient's global assessment (mm)	53.6 ± 24.7	54.6 ± 24.8	51.3 ± 25.0	50.8 ± 23.1	50.3 ± 26.8
Patient's assessment, pain (mm)	53.3 ± 25.2	54.0 ± 25.3	52.7 ± 26.3	51.0 ± 22.9	49.9 ± 27.9
HAQ	1.2 ± 0.8	1.2 ± 0.8	1.2 ± 0.9	1.2 ± 0.8	1.0 ± 0.8
CDAI	26.1 ± 13.0	26.8 ± 13.6	25.5 ± 12.2	23.4 ± 10.9	25.1 ± 11.0
RF, positive (%)	76.4	77.6	67.5	78.5	73.4
ACPA, positive (%)	71.2	72.9	63.2	69.4	72.7
ESR (mm/h)	48.4 ± 31.1	48.7 ± 31.7	53.2 ± 31.1	43.8 ± 27.3	43.8 ± 28.9
CRP (mg/dl)	2.4 ± 3.3	2.4 ± 3.4	3.2 ± 4.1	1.3 ± 1.7	1.9 ± 2.6
MMP-3 (ng/ml)	254.1 ± 308.8	260.7 ± 317.2	293.4 ± 315.7	177.5 ± 199.2	247.0 ± 365.9
Methotrexate dose (mg/week)	11.3 ± 3.7	11.2 ± 3.6	10.9 ± 4.0	11.7 ± 3.8	13.3 ± 2.9
Concomitant GC (%)	24.0	23.7	35.3	18.7	10.1

Table 1 continued

	Total	TNFi (Study population)	IL-6Ri	CTLA4Ig	JAKi
GC dose, PSL equivalent (mg/day)	1.3 ± 3.4	1.2 ± 3.5	2.2 ± 4.0	0.9 ± 2.2	0.5 ± 1.7
Concomitant csDMARDs (%)	5.0	4.8	4.8	6.1	7.6
Laboratory tests					
White blood cell count ($\times$ 1000/μl)	6.9 ± 2.2	6.9 ± 2.2	7.4 ± 2.2	6.5 ± 2.2	6.8 ± 2.4
Lymphocytes (%)	20.2	20.4	18.8	20.0	21.3
Neutrophils (%)	70.7	70.6	72.7	70.1	68.6
Albumin (g/dl)	3.8 ± 0.5	3.8 ± 0.5	3.6 ± 0.5	3.8 ± 0.5	3.9 ± 0.4
HbA1c (%)	5.8 ± 0.8	5.8 ± 0.8	6.0 ± 0.8	5.8 ± 0.5	5.8 ± 0.7
IgG (mg/dl)	1480.1 ± 439.1	1495.8 ± 432.9	1421.8 ± 443.5	1444.3 ± 465.8	1438.9 ± 456.7
IgA (mg/dl)	286.8 ± 133.7	286.5 ± 127.3	275.9 ± 130.7	298.3 ± 175.7	292.2 ± 129.5
IgM (mg/dl)	115.1 ± 63.0	117.1 ± 62.0	109.1 ± 65.8	112.2 ± 69.5	98.0 ± 51.0
CH50 (CH50/ml)	56.2 ± 9.2	55.7 ± 9.4	58.3 ± 8.5	57 ± 8.8	59.8 ± 6.5
C3 (mg/dl)	121.2 ± 23.3	122.2 ± 23.3	122.9 ± 24.9	112.7 ± 19.4	121.3 ± 23.2
C4 (mg/dl)	27.8 ± 9.2	28.1 ± 9.2	28.2 ± 9.2	25.9 ± 8.6	27.8 ± 8.9
KL-6 (U/ml)	256.8 ± 152.8	252.8 ± 154.2	255.6 ± 152.4	274.0 ± 142.9	282.7 ± 155.2
Antinuclear antibody ($\leq$ 40/80/160/320/640/1280 $\leq$ x, %)	63/15/10/5/3/4	66/14/10/4/2/4	63/15/9/3/4/6	54/19/8/7/8/4	56/25/12/3/0/4
Anti dsDNA antibody (IU/ml)	5.1 ± 9.1	5.4 ± 9.7	4.4 ± 8.1	4.3 ± 5.8	3.0 ± 2.5
Anti SS-A antibody, positive (%)	17.0	16.6	14.0	22.4	17.7
Comorbidities					
Polymyalgia rheumatica (%)	4.9	3.4	15.1	5.0	5.1
Interstitial lung disease (%)	4.3	4.2	5.2	4.6	3.8

Table 1 continued

	Total	TNFi (Study population)	IL-6Ri	CTLA4Ig	JAKi
Coexisting lung diseases (%)	20.8	18.7	18.7	37.8	15.2
NTM (%)	0.2	0.1	0.0	1.1	0.0
Past history					
Pneumonia (%)	4.6	4.0	6.0	6.9	6.3
Malignancy (%)	7.9	7.0	12.7	9.5	6.3
Fracture (%)	9.7	6.6	16.3	22.5	8.9
Pneumocystis pneumonia (%)	0.3	0.2	0.8	0.4	0.0
Interstitial lung disease (%)	0.4	0.2	0.4	1.5	0.0
Tuberculous (%)	7.7	7.3	5.6	13.0	6.3

ACPA anti-citrullinated peptide antibody, *BMI* body mass index, *b/tsDMARDs* biologic/targeted synthetic disease-modifying antirheumatic drugs, *CDAI* Clinical Disease Activity Index, *CRP* C-reactive protein, *csDMARDs* conventional synthetic DMARDs, *CTLA4Ig* cytotoxic T-lymphocyte-associated protein 4 immunoglobulin, *dsDNA* double-stranded DNA antibody, *eGFR* estimated glomerular filtration rate, *ESR* erythrocyte sedimentation rate, *GC* glucocorticoid, *HAQ* Health Assessment Questionnaire, *Ig* immunoglobulin, *IL-6Ri* interleukin-6 receptor inhibitor, *JAKi* Janus kinase inhibitor, *MMP-3* matrix metalloproteinase 3, *MTX* methotrexate, *NTM* nontuberculous mycobacteria disease, *PSL* prednisolone, *RA* rheumatoid arthritis, *RF* rheumatoid factor, *TNFi* tumor necrosis factor inhibitor

approach, we chose to exclude censored cases from this study as their occurrence appeared to be random and unrelated to the patients' medical conditions. Indeed, an analysis comparing background factors between censored and noncensored cases showed no statistically significant differences (all p values > 0.05 after Benjamini–Hochberg false discovery rate correction, data not shown).

Analysis Procedure (Fig. 1)

The primary objective of this study was to develop a predictive model for achieving CDAI remission 6 months after initiation of TNFi as the first b/tsDMARD in addition to MTX after unsuccessful MTX therapy. We included patients enrolled in the FIRST registry between August 2003 (introduction of IFX) and October 2022. Baseline characteristics at initiation of first b/tsDMARD and CDAI measurements at 6 months were collected from the FIRST registry. Patients who discontinued the first b/tsDMARD within 6 months were classified as non-remission cases (Fig. 1: Study population).

The analysis consisted of two parts. The first part focused on assessing the generalizability of a previously reported remission prediction model. We used DAS28 remission prediction model developed by Vastesaeger et al.'s (Vastesaeger's model) from a European cohort of GLM [20]. The effectiveness of Vastesaeger's model in predicting DAS28-ESR remission in study population (to assess its applicability beyond the original cohort) and CDAI remission (to assess its generalizability across different composite measures) 6 months after TNFi initiation was evaluated using the study population.

In the second part, we performed a retrospective analysis of our cohort to build a model based only on factors available in routine clinical practice.

(i) Primary analysis (Fig. 1, left): The analysis compared multivariate logistic regression with stepwise variable selection, decision tree, support vector machine (SVM) and Lasso logistic regression, a machine learning technique for sparse modeling. We assessed its robustness using k-fold cross-

Table 2 Treatment outcome of study population at 6 months

	Total	TNFi	IL6Ri	CTLA4Ig	JAKi
Subjects enrolled	2223	1630	252	262	79
Completed/ Cessation/ Censored (%)	81/ 9/ 10	82/ 10/ 8	80/5/ 15	81/8/11	84/ 3/ 13
CDAI remission (%)	32.9	33.0	33.7	27.2	48.5
DAS28-ESR remission (%)	40.9	38.3	67.3	27.9	59.1

CDAI Clinical Disease Activity Index, *CTLA4Ig* cytotoxic T-lymphocyte-associated protein 4 immunoglobulin, *DAS28* disease activity score 28, *IL-6Ri* interleukin-6 receptor inhibitor, *JAKi* Janus kinase inhibitor, *TNFi* tumor necrosis factor inhibitor

validation (k = 10). Receiver operator curve (ROC) analysis was used to calculate cutoff values to assess model performance.

(ii) Calendar cohort analysis (Fig. 1, middle): Given the increase in RA treatment options and the change in treatment strategies over time, we performed a sensitivity analysis. The cohort was divided by the cutoff date of October 2019, resulting in a training to validation cohort ratio of 9:1 (Calendar cohort).

(iii) Sub-cohort analysis (Fig. 1, right): The introduction of ABT has influenced the approach to selecting bDMARDs. Subsequently, ABT tends to be prescribed to patients at high risk of infection [25]. Therefore, we performed an analysis similar to i) on the sub cohort of patients who started TNFi as their first b/tsDMARD after the launch of ABT.

In the second part, all continuous variables are standardized through the process of z-score normalization. STATA-SE17 (for logistic
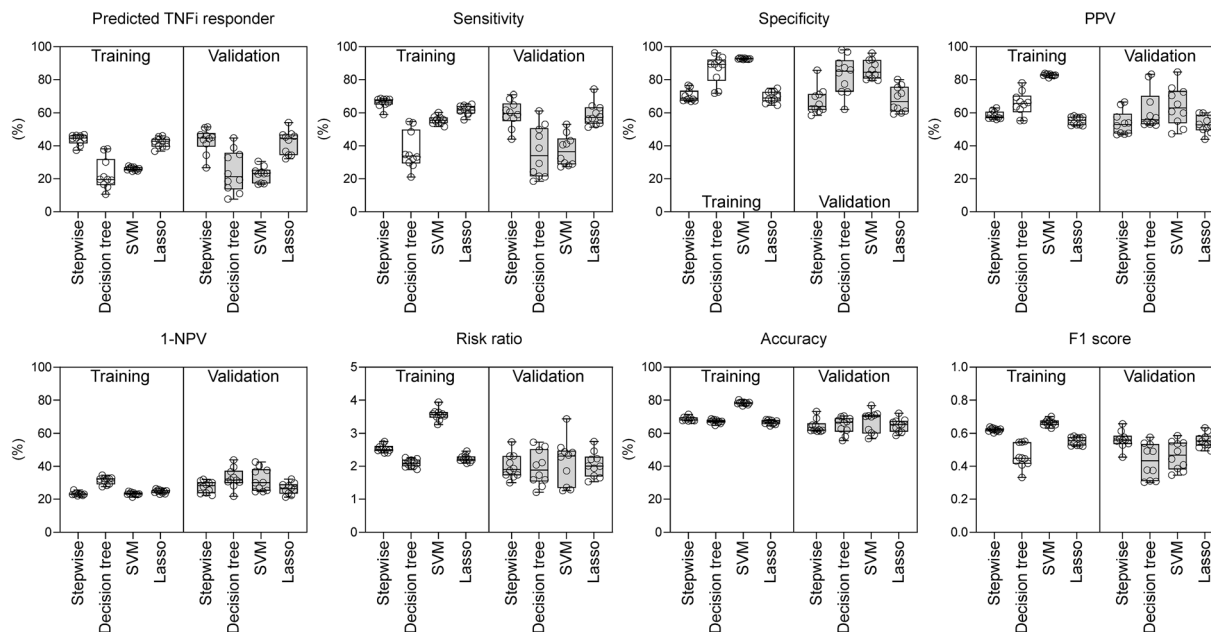
Fig. 2 Comparative analysis of different machine learning models. Multivariate logistic regression with stepwise variable selection (Stepwise), decision tree, support vector machine (SVM), and lasso logistic regression (Lasso) were used to generate Clinical Disease Activity Index (CDAI) remission prediction models. Each approach used k-fold cross validation ($k = 10$). *NPV* negative predictive value, *PPV* positive predictive value, *risk ratio* PPV/(1-NPV), *SVM* support vector machine, *TNFi* tumor necrosis factor inhibitor

regression with stepwise variable selection and Lasso logistic regression) and JMP pro 16 (for decision tree and SVM) software were used for statistical analysis.

Ethical Approval

This study followed the Declaration of Helsinki and was approved by the Ethics Committee of the University of Occupational and Environmental Health School of Medicine (#UOEHCRB21-068). Informed consent was obtained from all participants based on the Ethical Guidelines for Medical and Health Research Involving Human Subjects issued by the Japanese Ministry of Health, Labor and Welfare. Written informed consent was obtained from participants enrolled after April 2015, and written or verbal consent was obtained from others.

RESULTS

Study Diagram (Fig. 1)

Of the 5066 individuals registered in the FIRST registry, 4706 initiated their b/tsDMARDs treatment within the study period. Of these, 2223 MTX users initiated their first b/tsDMARDs, with the distribution among TNFi (N = 1630), IL-6Ri (252), CTLA4-Ig (262), and JAKi (79). The characteristics of the b/tsDMARDs naïve MTX users including the study population (MTX + TNFi) are detailed in Table 1. The entire cohort consisted of Asian individuals, with an average age of 59.2 ± 14.4 years, a body mass index (BMI) of 22.1 ± 3.7, and a mean disease duration of 75.7 ± 106.1 months. Approximately 24.0% of the patients were on glucocorticoids (GC). All patients received MTX at a mean dose of 11.3 ± 3.7 mg/week. The mean CDAI score was 26.1 ± 13.0.

Table 3 Variables selected in the Lasso-generated CDAI remission predictive models

	Training, k01	Training, k02	Training, k03	Training, k04	Training, k05	Training, k06	Training, k07	Training, k08	Training, k09	Training, k10
Female	− 0.435		− 0.385			− 0.503	− 0.389	− 0.338		
Age										
Smoking										
BMI										
eGFR	− 0.229									
RA duration	− 0.165					− 0.085		− 0.107	− 0.145	− 0.163
Steinbrocker stage										
Morning stiffness										
Tender joint count (28)										
Swollen joint count (28)										
Evaluator's global assessment										
Patient's global assessment	− 0.305	− 0.289	− 0.197	− 0.268	− 0.366	− 0.233	− 0.268	− 0.248	− 0.397	− 0.340
Patient's assessment, pain	− 0.205	− 0.222	− 0.329	− 0.210	− 0.192	− 0.266	− 0.255	− 0.300	− 0.195	− 0.220
HAQ	− 0.450	− 0.386	− 0.432	− 0.488	− 0.478	− 0.355	− 0.412	− 0.339	− 0.355	− 0.385
RF, positive										
RF ≥ 2 ULN		0.201								
RF ≥ 3 ULN	0.243						0.178			
RF_Q1	− 0.361	− 0.243	− 0.222	− 0.278	− 0.407	− 0.411	− 0.204	− 0.433		
RF_Q2										
RF_Q3										

Table 3 continued

	Training, k01	Training, k02	Training, k03	Training, k04	Training, k05	Training, k06	Training, k07	Training, k08	Training, k09	Training, k10
RF_Q4										
ACPA, positive										
ACPA ≥ 2 ULN										
ACPA ≥ 3 ULN			0.149			0.164	0.138			
ACPA_Q1										
ACPA_Q2	− 0.182			− 0.386		− 0.213	− 0.326	− 0.247	− 0.296	
ACPA_Q3										
ACPA_Q4	0.107				0.356	0.163			0.274	
ESR										
CRP						− 0.228				
MMP-3										
MTX dose	0.139									
Concomitant GC	− 0.560	− 0.056	− 0.595	− 0.425	− 0.580	− 0.476	− 0.514	− 0.620	− 0.418	− 0.467
GC dose, PSL equivalent		− 0.399								
Concomitant csDMARDs										
White blood cell count										
Lymphocytes										
Neutrophils										
Albumin						0.060				
HbA1c										
IgG	0.270	0.247	0.262	0.265	0.231	0.331	0.230	0.263	0.233	0.292

Table 3 continued

	Training, k01	Training, k02	Training, k03	Training, k04	Training, k05	Training, k06	Training, k07	Training, k08	Training, k09	Training, k10
IgA										
IgM										
CH50				0.185						
C3	0.086		0.125			0.141		0.120		
C4	0.174		0.200		0.189	0.177		0.113		
KL-6	− 0.205	− 0.178	− 0.172	− 0.168		− 0.186	− 0.199			− 0.153
Antinuclear antibody	0.213						0.178			
anti dsDNA antibody										
anti SS-A antibody, positive										
Interstitial lung disease										
Coexisting lung diseases	− 0.389					− 0.226	− 0.269			
NTM										
Pneumonia										
Malignancy	− 0.727	− 0.713	− 0.940	− 0.772	− 0.564	− 0.762	− 0.687	− 0.738	− 0.837	− 0.667
Fracture										
Pneumocystis pneumonia										
Interstitial lung disease										

Table 3 continued

	Training, k01	Training, k02	Training, k03	Training, k04	Training, k05	Training, k06	Training, k07	Training, k08	Training, k09	Training, k10
Tuberculous										

Number indicates coefficient. Continuous variables are standardized through the process of z-score normalization

ACPA anti-citrullinated peptide antibody, *BMI* body mass index, *CDAI* Clinical Disease Activity Index, *CRP* C-reactive protein, *csDMARDs* conventional synthetic disease-modifying anti-rheumatic drug, *dsDNA* double-stranded DNA antibody, *eGFR* estimated glomerular filtration rate, *ESR* erythrocyte sedimentation rate, *GC* glucocorticoid, *HAQ* Health Assessment Questionnaire, *Ig* immunoglobulin, *MMP-3* matrix metalloproteinase 3, *MTX* methotrexate, *NTM* nontuberculous mycobacteria disease, *PSL* prednisolone, *Q1–4* quartiles, *RA* rheumatoid arthritis, *RF* rheumatoid factor, *ULN* upper limit of normal

Clinically significant interstitial lung disease (ILD) was observed in only 4.3% of cases, although coexisting lung disease (such as interstitial pneumonitis, pleuritis, diffuse panbronchiolitis, bronchiectasis, old tuberculosis or inflammatory nodules detected on chest X-ray/CT) was observed in 20.8% of cases. These patients were treated with b/tsDMARDs. Of these, 90% completed the 6-month follow-up (81% continued treatment, 9% discontinued for various reasons, and 10% were censored). At 6 months, the rates of CDAI remission rates were TNFi/IL-6Ri/CTLA4-Ig/JAKi 33.0/33.7/27.2/48.5%. The rate of DAS28-ESR remission (overall 40.9%) was higher than that of CDAI remission (overall 32.9%), particularly pronounced in the case of IL-6Ri (33.7 vs. 67.3%, Table 2).

Vastesaeger's Model Predicts DAS28 Remission, Yet Shows Inferior Predictive Performance for CDAI Remission in External Cohort

Vastesaeger et al. proposed a predictive model for TNFi-induced DAS28 remission based on six baseline factors: sex, age, HAQ, TJC28, ESR, and comorbidities [20]. This model was validated in an external European cohort of 674 cases [27]. We simulated whether this model predicts DAS28-ESR remission and CDAI remission in an Asian cohort (Supplementary Materials Table S1).

In the analysis of the study population, participants with missing variables or censoring were excluded. The results showed that 40.0% achieved DAS28-ESR remission and 35.5% achieved CDAI remission at 6 months (Supplementary Materials Table S1, far left). Remission probability scores [20] were calculated for each patient using Vastesaeger's model, and its performance in predicting remission was evaluated using the specified cutoffs (20%, 30%, 40%, 50%, 67% as in Ganhao et al. [27], along with cutoffs determined via ROC analysis in the study population.

Vastesaeger's model showed significant predictive ability for DAS28-ESR remission within the study population. For example, using a 20% cutoff, 59.2% of b/tsDMARDs naïve MTX users

A
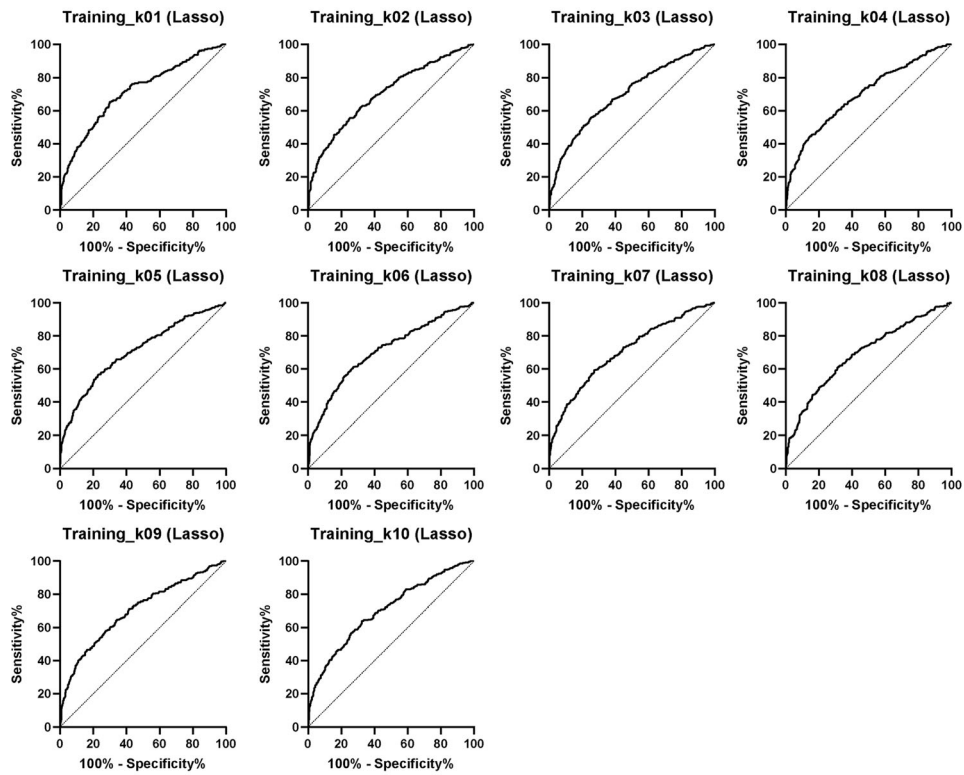
B
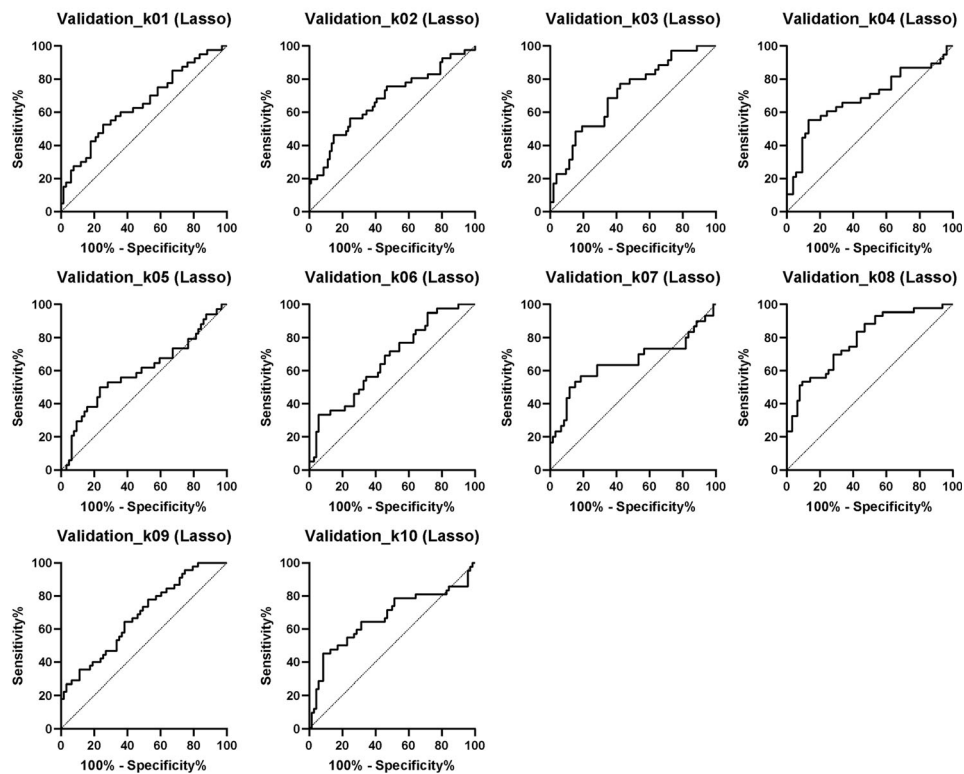
◀**Fig. 3** Areas under the curve of the Lasso-generated CDAI remission predictive models. Predictive models for Clinical Disease Activity Index (CDAI) remission were generated using training cohorts (**A**) and their performance was evaluated using validation cohorts (**B**). Figures show receiver operating characteristic (ROC) curves calculated in **A** the training cohorts and **B** the validation cohorts. The ten curves represent the k-fold cross-validation cohort

were identified as predicted TNFi responders, with 50.4% of them achieving DAS28-ESR remission at 6 months (= positive predictive value [PPV]). Conversely, only 25.0% of those classified as predicted TNFi non-responders achieved remission (= 1- negative predictive value [NPV]). The performance of the model across different cutoffs included: subjects classified as predicted TNFi responders 1.8–59.2%; PPV 50.4–69.9%; 1-NPV 25.0–39.5%. The risk ratio, which represents the ratio of remission rates in predicted TNFi responders to predicted TNFi non-responders, ranged from 1.73 to 2.03 (Supplementary Materials Table S1, top half).

Next, we evaluated the predictive ability of Vastesaeger's model for CDAI remission (Supplementary Materials Table S1, bottom half). It also showed some predictive ability: those classified as predicted TNFi responders 1.8–59.2%; PPV 40.7–54.5%; 1-NPV 27.8–35.2%. Notably, however, the risk ratios ranged from 1.38 to 1.55, indicating that Vastesaeger's model was less effective in distinguishing TNFi responders from non-responders in advance when using CDAI.

Vastesaeger's Model Fails to Predict CDAI Remission for Certain TNF Inhibitors

The Vastesaeger's model, developed using a cohort of patients treated with GLM, predicted overall CDAI remission in the study population, but may have variable predictive performance across different TNFis. Hence, the study population was stratified by each TNFi to assess the performance of Vastesaeger's model (Supplementary Materials Table S2). GLM was commonly prescribed to MTX non-users in FIRST

registry, resulting in a small sample size for this analysis (N = 29). Therefore, they were excluded.

The predictive ability of the Vastesaeger's model for CDAI remission in the study population was assessed using a cutoff value of 21% (Supplementary Materials Table S1, bottom: ROC cutoff). It successfully distinguished TNFi responders from non-responders in patients treated with ETN (PPV 36.1%, 1-NPV 18.2%, RR 1.98 [1.21–3.25]) or ADA (PPV 47.0%, 1-NPV 29.2%, RR 1.61[95% confidence interval (CI): 1.19–2.17]). However, its predictive ability was not observed for IFX (risk ratio: 1.36 [0.99–1.85]) or CZP (1.14 [0.85–1.53]) (Supplementary Materials Table S2).

Development of a CDAI Remission Predictive Model

Vastesaeger's model, based solely on clinical information, predicted DAS28-ESR remission in our Asian cohort. However, its predictive performance for CDAI remission was limited and showed instability in predictive ability across different TNFis.

Therefore, we aimed to develop a CDAI remission predictive model using FIRST registry. FIRST registry comprehensively collects clinical items related to RA. Table S3 in the Supplementary Materials shows 48 such items along with their missing rates. Before the formulation and publication of CDAI, evaluators global assessment (EGA) was not routinely documented. Anti-citrullinated protein antibodies (ACPA) was not accessible before 2007. As a result, the missing rates for EGA and ACPA are particularly high. First, we compared various machine-learning approaches; logistic regression with stepwise variable selection, decision tree, SVM and Lasso logistic regression using 48 factors and performed k-fold cross-validation, dividing the cohort into a 9:1 training: validation ratio. The models built using the training cohorts were evaluated for performance using the corresponding validation cohorts. Figure 2 summarizes the performance of the tested approaches. Stepwise and Lasso predicted response rates of 43.9% and 42.1% in

b/tsDMARDs-naïve MTX users, whereas decision tree and SVM predicted lower rates (22.8% and 26.1%). Stepwise and Lasso had higher sensitivities, whereas decision tree and SVM had higher specificities. PPV and 1-NPV reflect CDAI remission rates in predicted responders and non-responders and have implications for clinical decisions: agent selection. All approaches showed comparable PPV, 1-NPV and risk ratios (determined by the PPV/1-NPV ratio). Notably, decision tree and SVM models showed more variation. Figure S1 in the Supplementary Materials shows the risk ratios and 95% CIs of models generated using different approaches with k-fold cross-validation. Some models had a 95% CI of the risk ratio below 1, indicating potential inadequacy in separating responders from non-responders. Such models were more likely to be obtained from decision tree (40%) and SVM (30%), while Stepwise and Lasso showed their stability. Model accuracy was comparable among the four approaches, with Stepwise and Lasso showing the higher F1 scores (Fig. 2). In summary, these analyses revealed favorable aspects of model performance, particularly with Stepwise and Lasso.

The performance of the Stepwise and Lasso models, which appear to support the clinical decision to avoid agent selection on a trail-and-error basis, raises concerns about their effectiveness in external cohorts and may be specific to our cohort. In this scenario, consideration of establishing CDAI remission prediction models tailored to each institution (regional/in-house precision medicine) becomes a viable option. However, limited sample size is a potential barrier to this approach, prompting an in-depth evaluation of sparse modeling. Table 3 details the factors selected by Lasso and Fig. 3 shows the ROC curves. Among the 48 factors, patient global assessment, patient pain assessment, HAQ, concomitant GC use, history of malignancy, and serum IgG were consistently included in all models and showed notable coefficients. The performance of the Lasso-generated CDAI remission predictive model, based on the ROC analysis cutoff for predicted TNFi responders/non-responders, was as follows: AUC 0.676 [95% CI 0.642–0.711], PPV (remission in predicted TNFi responders)

53.2% [49.4–57.0%], 1-NPV (remission in predicted non-TNFi responders) 26.4% [24.0–28.9%], risk ratio 2.01 [1.75–2.26] (Table 4, bottom).

Because Vastesaeger's model showed variable stability in predicting CDAI remission across different TNFis (Supplementary Materials Table S1-2), we evaluated the performance of our Lasso-generated CDAI remission predictive model for each TNFi. Across all drugs, the Lasso-generated model consistently demonstrated an AUC of 0.667–0.736, a PPV of 48.2–56.8%, a 1-NPV of 16.3–30.8%, and a risk ratio greater than 1 (1.85–3.01) (Supplementary Materials Fig. S2). This suggests that it can predict CDAI remission regardless of the specific drug used. However, it is important to note that the GLM was not included in this analysis.

Sensitivity Analysis: Calendar Cohort

Given the expanding range of RA treatment options over time, predictive models developed using older data, which may have limited coverage of current treatments, could be less effective in the current era. To address this, we performed a sensitivity analysis on Calendar cohort (Fig. 1, middle). As shown in Table S4 in the Supplementary Materials, the Lasso-generated CDAI remission predictive model from the training-calendar cohort showed a relatively lower yet statistically significant ability to predict CDAI remission during TNFi inhibitor use (PPV 58.8%, 1-NPV 36.5%, risk ratio 1.61 [95% CI: 1.05–2.46]).

Sensitivity Analysis: Performance of Lasso-Generated CDAI Remission Predictive Model in the Post-Abatacept Setting

ABT is recommended for patients at higher risk of infection [25]. In FIRST registry, ABT is often selected for such patients, which may bias the predictive performance of the Lasso-generated CDAI remission prediction model. Therefore, we performed Lasso logistic regression using k-fold cross-validation in a cohort consisting only of patients after ABT-launch. As shown in Figure S3 in the Supplementary Materials, the

Table 4 Performance of Lasso-generated CDAI remission predictive models

Cohort	N	Remission rate (%)	Predicted TNFi responder(%)	AUC	Sensitivity (%)	Specificity (%)
Training cohort						
Training_k01	895	38.1	43.5	0.716	65.4	70.0
Training_k02	986	36.3	42.8	0.703	62.3	68.3
Training_k03	917	37.8	36.8	0.699	55.9	74.9
Training_k04	918	37.6	46.1	0.699	64.1	64.7
Training_k05	916	38.1	39.6	0.710	59.9	72.8
Training_k06	895	38.3	40.2	0.709	61.2	72.8
Training_k07	921	38.3	39.5	0.705	59.2	72.7
Training_k08	907	37.5	42.9	0.691	61.5	68.3
Training_k09	915	37.2	45.6	0.701	64.7	65.7
Training_k10	996	35.8	43.7	0.702	63.9	67.6
Mean [95% CI]				0.704 [0.699–0.709]	61.7 [59.6–63.8]	69.9 [67.5–72.3]
Validation cohort						
Validation_k01	107	37.4	46.7	0.649	60.0	61.2
Validation_k02	122	33.6	34.4	0.675	53.7	75.3
Validation_k03	87	40.2	34.5	0.708	51.4	76.9
Validation_k04	92	41.3	43.5	0.686	63.2	70.4
Validation_k05	98	34.7	36.7	0.597	52.9	71.9
Validation_k06	109	35.8	45.9	0.663	56.4	60.0
Validation_k07	90	33.3	32.2	0.651	56.7	80.0
Validation_k08	107	40.2	54.2	0.786	74.4	59.4
Validation_k09	108	41.7	45.4	0.680	57.8	63.5
Validation_k10	112	37.5	47.3	0.667	64.3	62.9
Mean [95% CI]				0.676 [0.642–0.711]	59.0 [54.2–63.8]	68.0 [62.5–73.5]

Table 4 continued

	PPV (%)	NPV (%)	1-NPV (%)	Risk ratio [95% CI]	Accuracy (%)	F1 score
Training cohort						
	57.3	76.7	23.3	2.46 [2.05–2.94]	68.3	0.573
	52.8	76.1	23.9	2.21 [1.86–2.62]	66.1	0.528
	57.6	73.6	26.4	2.18 [1.85–2.57]	67.7	0.576
	52.2	74.9	25.1	2.09 [1.75–2.49]	64.5	0.522
	57.6	74.7	25.3	2.27 [1.92–2.69]	67.9	0.576
	58.3	75.1	24.9	2.35 [1.98–2.78]	68.4	0.583
	57.4	74.1	25.9	2.22 [1.88–2.62]	67.5	0.574
	53.7	74.7	25.3	2.12 [1.78–2.53]	65.7	0.537
	52.8	75.9	24.1	2.19 [1.83–2.62]	65.4	0.528
	52.4	77.0	23.0	2.28 [1.91–2.72]	66.3	0.524
	55.2 [53.4–57.1]	75.3 [74.5–76.1]	24.7 [23.9–25.5]	2.24 [2.17–2.30]	66.8 [65.8–67.8]	0.582 [0.573–0.592]
Validation cohort						
	48.0	71.9	28.1	1.71 [1.03–2.84]	60.7	0.533
	52.4	76.3	23.7	2.21 [1.36–3.59]	68.0	0.530
	60.0	70.2	29.8	2.01 [1.23–3.30]	66.7	0.554
	60.0	73.1	26.9	2.23 [1.33–3.73]	67.4	0.615
	50.0	74.2	25.8	1.94 [1.14–3.30]	65.3	0.514
	44.0	71.2	28.8	1.53 [0.92–2.54]	58.7	0.494
	58.6	78.7	21.3	2.75 [1.55–4.87]	72.2	0.576
	55.2	77.6	22.4	2.46 [1.39–4.34]	65.4	0.634
	53.1	67.8	32.2	1.65 [1.05–2.60]	61.1	0.553
	50.9	74.6	25.4	2.00 [1.20–3.34]	63.4	0.568

Table 4 continued

PPV (%)	NPV (%)	1-NPV (%)	Risk ratio [95% CI]	Accuracy (%)	F1 score
53.2 [49.4–57.0]	73.6 [71.1–76.0]	26.4 [24.0–28.9]	2.01 [1.75–2.26]	64.9 [62.0–67.8]	0.557 [0.526–0.588]

Cut-off values were determined through receiver operating characteristic curve analysis

AUC area under the curve, *CDAI* Clinical Disease Activity Index, *CI* confidence interval, *NPV* negative predictive value, *PPV* positive predictive value, *Risk ratio* PPV/(1-NPV), *TNFi* tumor necrosis factor inhibitor

validation cohort demonstrated a robust predictive performance with a PPV of 54.5% [95% CI: 49.2–59.8%], a 1-NPV of 29.5% [25.0–34.0%], and a risk ratio of 1.92 [1.62–2.23].

This sensitivity analysis also resulted in a reduced sample size (N = 1630–930). When comparing the model performance between Stepwise and Lasso in the 930 patients, Lasso shows higher AUC, sensitivity, specificity, PPV, risk ratio, accuracy, F1 score and lower 1-NPV (remission rate in predicted non-responders). These results highlight the benefits of sparse modeling (Fig. 4).

Clinical Application of the Lasso-Generated CDAI Remission Predictive Model

For clinical implementation, we considered the cutoff values of the Lasso-generated CDAI remission predictive models. Table 5 shows the performance validation of the models which are shown in Table 4. Using a log-odds cutoff of 0.400 (close to the cutoffs calculated by ROC analysis: 0.408–0.467), 41% (37–43%) were classified as predicted TNFi responders, of which 55% (53–56%) achieved CDAI remission 6 months after TNFi initiation, while 26% (24–27%) of the predicted TNFi non-responders achieved remission. In this case, it may be a decision-making option to start TNFi in 41% of MTX users who are naïve to b/tsDMARDs and to consider non-TNFi/JAKi for the remaining 59%. However, TNFi is considered the standard of care based on its extensive history and evidence and is therefore expected to be used in a larger patient population. Using log-odds of 0.300, 61% were classified as predicted TNFi responders, with CDAI remission rates of 47% in predicted TNFi responders and 23% in predicted TNFi non-responders. Other cutoffs may be considered based on the specific circumstances of the institution (i.e., remission rates with non-TNFis and JAKis).
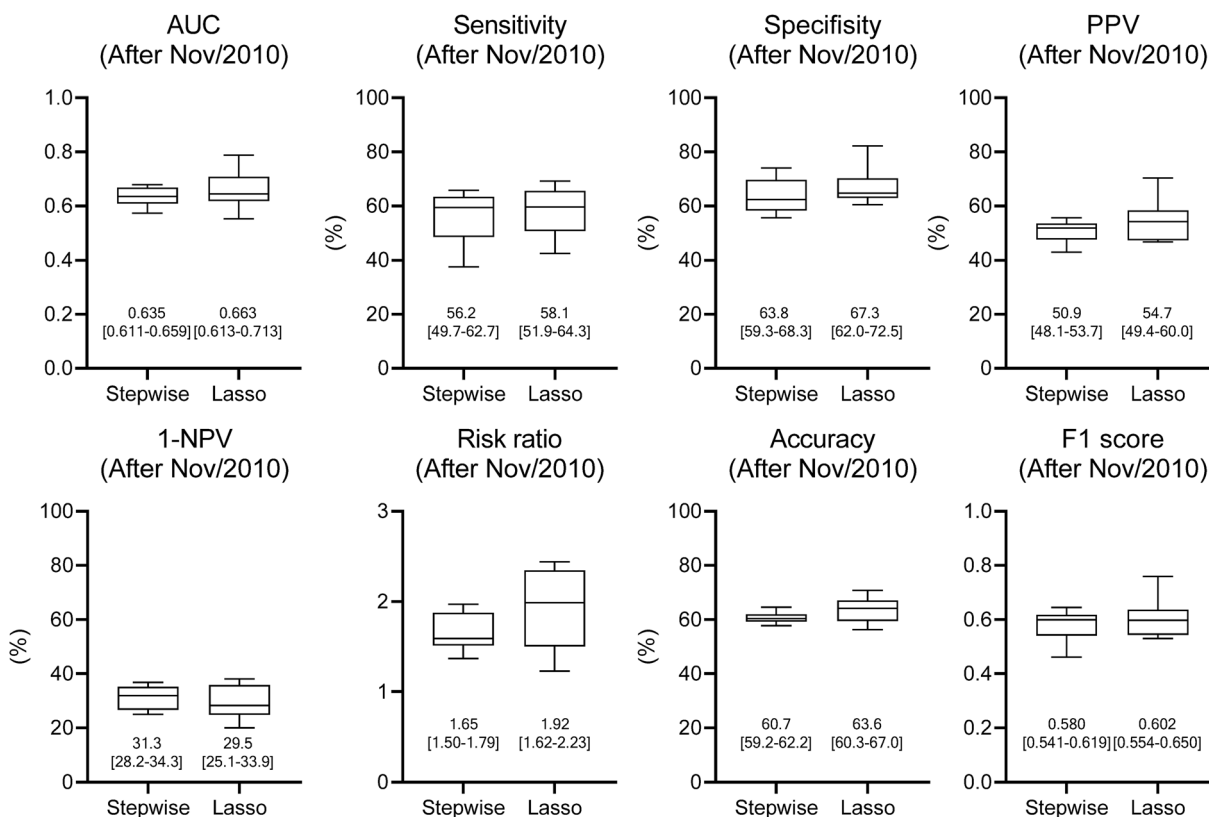
Fig. 4 Stepwise- vs. Lasso-generated CDAI remission predictive models in sparse cohort. The box plots along and corresponding values show the mean and 95% confidence intervals (CI) of performance for Clinical Disease Activity Index (CDAI) remission predictive models generated multivariate logistic regression by stepwise variable selection (Stepwise) or lasso logistic regression (Lasso). Both approaches include ten models generated using the k-fold cross-validation method. The cohort includes 930 subjects who were enrolled after November 2010. Cutoff values were determined by receiver operating characteristic curve analysis. *AUC* area under the curve, *NPV* negative predictive value, *PPV* positive predictive value, *Risk ratio* PPV/(1-NPV)

DISCUSSION

While advances in b/tsDMARD treatments have significantly improved the management of RA, no drug is universally effective for all patients. A model that predicts an individual patient's response to treatment can reduce trial-and-error treatment decisions and will certainly aid in the selection of treatment options. However, despite numerous attempts, no established method has been developed [28]. The aim of this research was to create a simple and affordable model using machine learning techniques that could predict CDAI remission 6 months after starting first b/tsDMARDs treatment with TNFi, in addition to MTX. The model was designed for practical use in standard clinical settings and relied solely on clinical information available in routine care. In our initial comparison, we assessed multivariate logistic regression analysis, decision tree, SVM, and Lasso logistic regression using K-fold cross-validation. Of the four approaches tested, Stepwise and Lasso showed advantages in terms of stability (lower deviation) and F1 score (Fig. 2). Furthermore, Lasso demonstrated robustness in the context of the sparse cohort, as shown in Fig. 4. These results suggest a potential strategy of selecting an alternative b/tsDMARD class (e.g., IL-6Ri, CTLA4-Ig, JAKi) for Lasso-predicted TNFi non-responders (Supplementary Materials Figure S4).

Predicting remission with TNFi treatment has been studied extensively. Initially,

Table 5 Optimization of the cutoff value predicting CDAI remission

Cohort	Cutoff log-odds by Lasso model	Predicted TNFi responder	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	1-NPV (%)	Risk ratio	Likelihood ratio	Accuracy (%)	F1 score
Cohort_k01	0.500	26	44	84	63	71	29	2.17	1.56	68.9	0.516
	0.400	43	64	70	56	76	24	2.34	1.91	67.5	0.600
	0.300	61	77	49	48	77	23	2.13	2.99	59.5	0.591
	0.200	82	90	23	42	79	21	1.97	6.85	48.6	0.570
	0.150	93	97	10	40	86	14	2.74	14.2	42.9	0.565
Cohort_k02	0.500	22	38	88	64	72	28	2.25	1.47	69.9	0.479
	0.400	38	56	72	53	74	26	2.08	1.99	66.3	0.545
	0.300	58	75	51	46	78	22	2.09	2.95	59.3	0.569
	0.200	80	89	25	40	79	21	1.94	6.69	47.7	0.550
	0.150	92	96	10	38	83	17	2.18	15.0	41.2	0.541
Cohort_k03	0.500	26	43	85	64	71	29	2.21	1.52	69.2	0.516
	0.400	42	59	68	54	73	27	2.00	2.14	64.9	0.563
	0.300	63	77	46	47	77	23	2.05	3.24	58.3	0.586
	0.200	83	92	22	42	81	19	2.18	6.87	48.3	0.574
	0.150	95	98	7	39	85	15	2.55	19.7	41.6	0.561
Cohort_k04	0.500	25	44	86	66	72	28	2.33	1.45	70.2	0.527
	0.400	42	60	69	54	74	26	2.07	2.07	65.5	0.569
	0.300	62	75	47	46	76	24	1.90	3.43	57.5	0.574
	0.200	83	91	22	41	79	21	1.96	7.47	47.7	0.568
	0.150	95	98	7	39	85	15	2.65	18.7	41.7	0.560
Cohort_k05	0.500	25	42	86	65	71	29	2.23	1.50	69.4	0.511
	0.400	42	61	69	55	74	26	2.13	2.02	66.1	0.574
	0.300	62	76	47	46	76	24	1.95	3.32	57.9	0.576

Table 5 continued

Cohort	Cutoff log-odds by Lasso model	Predicted TNFi responder	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	1-NPV (%)	Risk ratio	Likelihood ratio	Accuracy (%)	F1 score
	0.200	83	91	22	42	81	19	2.14	6.77	48.3	0.571
	0.150	95	97	7	39	79	21	1.87	25.7	40.8	0.554
Cohort_k06	0.500	26	44	85	64	71	29	2.22	1.53	69.3	0.523
	0.400	42	61	70	55	74	26	2.17	1.98	66.4	0.580
	0.300	61	77	49	48	77	23	2.13	3.00	59.5	0.591
	0.200	83	92	23	42	82	18	2.31	6.27	49.1	0.578
	0.150	94	98	8	40	86	14	2.87	16.5	42.2	0.563
Cohort_k07	0.500	26	43	85	63	71	29	2.15	1.56	68.7	0.508
	0.400	41	60	71	56	74	26	2.17	1.95	66.7	0.578
	0.300	62	77	48	47	77	23	2.04	3.17	58.6	0.583
	0.200	83	90	22	41	77	23	1.81	7.99	47.4	0.563
	0.150	94	97	8	39	82	18	2.13	19.7	41.6	0.558
Cohort_k08	0.500	25	43	85	63	71	29	2.18	1.54	69.0	0.509
	0.400	42	61	69	54	74	26	2.11	2.04	65.9	0.573
	0.300	61	78	48	48	78	22	2.17	3.00	59.4	0.590
	0.200	82	91	23	42	81	19	2.19	6.54	48.6	0.573
	0.150	94	98	8	39	84	16	2.52	18.0	41.7	0.559
Cohort_k09	0.500	24	42	86	65	71	29	2.23	1.49	69.5	0.506
	0.400	40	58	71	54	73	27	2.04	2.05	65.7	0.558
	0.300	61	75	48	47	76	24	1.98	3.21	58.4	0.577
	0.200	83	90	22	41	78	22	1.87	7.71	47.4	0.563
	0.150	95	98	7	39	83	17	2.24	22.2	41.0	0.555
Cohort_k10	0.500	22	39	88	65	72	28	2.31	1.44	70.4	0.488

Table 5 continued

Cohort	Cutoff log-odds by Lasso model	Predicted TNFi responder	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	1-NPV (%)	Risk ratio	Likelihood ratio	Accuracy (%)	F1 score
	0.400	37	57	74	55	75	25	2.23	1.85	67.8	0.560
	0.300	59	75	50	46	78	22	2.07	3.03	58.8	0.567
	0.200	80	89	25	40	80	20	1.98	6.42	48.0	0.551
	0.150	92	96	10	38	84	16	2.31	14.5	41.2	0.542
Summary	**0.500**	**25**	**42**	**86**	**64**	**71**	**29**	**2.23**	**1.51**	**69.5**	**0.508**
	0.400	**41**	**60**	**70**	**55**	**74**	**26**	**2.13**	**2.00**	**66.3**	**0.570**
	0.300	**61**	**76**	**48**	**47**	**77**	**23**	**2.05**	**3.13**	**58.7**	**0.580**
	0.200	**82**	**90**	**23**	**41**	**80**	**20**	**2.04**	**6.96**	**48.1**	**0.566**
	0.150	**94**	**97**	**8**	**39**	**84**	**16**	**2.41**	**18.4**	**41.6**	**0.556**

Performance metrics of the Lasso-generated Clinical Disease Activity Index (CDAI) remission predictive model using the cohorts shown in Fig. 3 and Table 4
NA not available, *NPV* negative predictive value, *PPV* positive predictive value, *risk ratio* PPV/(1-NPV), *TNFi* tumor necrosis factor inhibitor

researchers tried to use various markers such as age, gender, disease duration, serum biomarkers (rheumatoid factor: RF, ACPA, inflammatory mediators), genetic information (SNP), blood gene patterns, and immune cell types from blood samples (examined by flow cytometry) [8, 9, 11, 13, 15]. However, these methods have not been widely accepted due to issues with accuracy and reproducibility [8, 9, 11, 13]. Recent efforts have focused on comprehensive gene expression analysis in peripheral blood [10–12] and synovial tissue [14] to predict treatment response. Several studies have introduced a multi-omics approach, combining factors such as peripheral blood gene expression and epigenetic status [17]. Other attempts have used synovial biopsies, clustering synovial tissue types based on histological/genetic multi-omics to predict treatment response [16, 18]. However, while these advanced methods appear promising, they require further testing for reproducibility. In addition, these methods may not be widely accessible due to cost and complexity of analysis, limiting their use to specific healthcare settings. Thus, the potential use of predictive models that integrate readily available clinical data seems noteworthy, especially in scenarios with ethical or economic constraints on access to advanced technology.

In this context, a notable study emerged in 2012 (published in 2019) that presented a machine learning-based method that integrated clinical and SNP data to predict responsiveness to TNFi treatment (assessed by change in DAS28 at 24 months) [29]. This analysis, conducted across international cohorts, showed minimal SNP contribution due to significant interethnic variation. It highlighted the sufficiency of four clinical parameters (sex, age, MTX use, and baseline DAS28) for accurate prediction [29]. These findings offer encouraging prospects for achieving high predictive efficacy even in resource-constrained healthcare settings. However, concerns have been raised about the translation of these findings into daily practice due to its focus on DAS28 remission at 24 months and the potential increased computational burden associated with Gaussian process regression.

Vastesaeger et al. reported in 2016 a simple matrix derived from clinical information (sex, age, HAQ, TJC28, ESR, presence of comorbidities) to predict treatment response in csDMARD non-responders augmented with GLM [20]. This predictive model has demonstrated efficacy in an external European cohort including multiple TNFis [27]. However, this tool also defines DAS28 remission as the outcome, which is not consistent with current recommendations for more stringent remission criteria such as CDAI [21, 22]. In fact, while Vastesaeger's model accurately predicted DAS28-ESR remission in our Aian cohort, its predictive performance for CDAI remission was inferior (Supplementary Materials Table S1). Furthermore, the predictive performance of Vastesaeger's model for CDAI remission did not show stability across different TNFis (Supplementary Materials Table S2).

This study demonstrated the development of a predictive model using a machine learning technique, Lasso, capable of predicting CDAI remission 6 months after TNFi initiation (Table 4). Notably, this model stably predicted remission across IFX, ETN, ADA, and CZP (Supplementary Materials Figure S2).

Reports by Vastesaeger [20] and Guan[29] demonstrated the potential of predictive tools based solely on clinical information to be applicable to external cohorts. Interestingly, all models generated by lasso regression with K-fold cross-validation in our study selected only PRO and quantitative data, excluding subjective physician input. K-fold cross-validation generated 10 CDAI remission predictive models by lasso logistic regression, showing general consistency in variable selection from a pool of 48 factors relevant to daily RA practice: Sex, RA duration, PROs, RF, ACPA, concomitant GC use, and history of malignancy. These factors are known contributors to refractory disease [8, 9, 30]. This suggest the potential for extrapolation of our model; however, it should be noted that our models also selected certain factors, such as RF and antinuclear antibody (ANA) titers (Table 3), for which measurements lack international standardization. Therefore, external validation in other cohorts remains crucial to confirm the reliability of our study results.

Should our findings prove to be less generalizable, it may be worthwhile to explore the tailored prediction models for different regions or institutions (regional/in-house precision medicine). Our study relied on data at the initiation of b/tsDMARDs and subsequent CDAI at month 6, highlighting the feasibility of establishing observational cohorts and creating tailored predictive models for different regions or institutions. The conventional multivariate logistic model with stepwise variable selection was also explored for model building. This approach yielded positive results in our cohort of 1630 subjects. However, the simulation of sample size reduction showed that the Stepwise model decreased its effectiveness while Lasso model remained effective, thus we advocate for the use of Lasso logistic regression because of its ability to mitigate concerns regarding overtraining, especially in cases with limited sample sizes, a major concern with this approach.

This study has certain limitations. The accuracy of the Lasso-generated CDAI remission predictive model in the Calendar cohort decreased (Supplementary Materials Table S4). In Calendar cohort, which was split in a 9:1 ratio with a cutoff of October 2019, increased censoring was observed in the validation cohort (32 vs. 5% in training cohort). This was undoubtedly influenced by the COVID-19 situation, which have had an impact on the reduced accuracy observed.

Second, this study had fewer subjects treated with GLM. Thus, the proposed predictive model and the concept of regional/in-house precision medicine may not be feasible in cohorts with higher rates of GLM use.

Third, our selection strategy among b/tsDMARDs classes (TNFi, IL-6Ri, CTLA4-Ig, JAKi) might influence the study results. For example, prescribing ABT to individuals at high risk of infection in accordance with guideline recommendations [25]. However, this strategy is likely to be widely adopted across regions and institutions, which may not diminish the impact of our study results.

The limited predictive power of the machine learning models in this study may raise questions about the relevance of the study results. While the integration of advanced technologies such as proteomics or transcriptome analysis of blood or synovium has the potential to improve predictive power, it comes with higher costs and feasibility challenges. Likewise, unexplored statistical approaches such as deep neural networks or Gaussian process regression could improve model effectiveness, although these algorithms have high computational costs. It is important to recognize that our study, based on a single cohort, currently has limitations in terms of generalizability. Therefore, the objective of this study was not to determine the most effective model for our cohort, but rather to propose the feasibility of the concept: establishing a predictive model within each cohort. Each cohort can explore different statistical approaches and select the most appropriate among those generated.

CONCLUSIONS

This study suggested the potential of a predictive model for CDAI remission following TNFi treatment using Lasso-based machine learning. While validation in external cohorts is required to confirm its extrapolation, promising indications emerge from two previous studies.

This study also suggests the feasibility of creating cost-effective models customized for individual regions/institutes and implementing regional/in-house precision medicine approaches.

ACKNOWLEDGEMENTS

Tanaka, Atsushi Nagayasu. Shingo Nakaya-mada, Yoshiya Tanaka. Substantial contributions to analysis and interpretation of data: Koshiro Sonomoto, Yoshihisa Fujino. All authors contributed to the drafting the article or revising it critically for important intellectual content, as well as the final approval of the version of the article to be published.

Data Availability. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of Interest. Koshiro Sonomoto has received speaking fee from Chugai, Astellas, AbbVie, Taisho, Janssen, Eli Lilly Japan, Ayumi, Gilead Sciences and has received research funding from UCB Japan. Shingo Nakayamada has received consulting fees, speaking fees, lecture fees, and/or honoraria from AstraZeneca, GlaxoSmithKline, Pfizer, Bristol-Myers, Astellas, Asahi-kasei, AbbVie, Chugai, Sanofi, Eisai, Gilead Sciences, Mitsubishi-Tanabe, Janssen, Eli Lilly, and Ayumi. Yoshiya Tanaka has received speaking fees and/or honoraria from Eli Lilly, AstraZeneca, AbbVie, Gilead, Chugai, Behringer-Ingelheim, GlaxoSmithKline, Eisai, Taisho, Bristol-Myers, Pfizer, Taiho, received research grants from Mitsubishi-Tanabe, Eisai, Chugai, Taisho. Yoshiya Tanaka is an Editorial Board member of *Rheumatology and Therapy*. Yoshiya Tanaka was not involved in the selection of peer reviewers for the manuscript nor any of the subsequent editorial decisions. Yoshihisa Fujino, Hiroaki Tanaka, and Atsushi Nagayasu have nothing to disclose.

Ethical Approval. This study followed the Declaration of Helsinki and was approved by the Ethics Committee of the University of Occupational and Environmental Health School of Medicine (#UOEHCRB21-068). Informed consent was obtained from all participants based on the Ethical Guidelines for Medical and Health Research Involving Human Subjects issued by the Japanese Ministry of Health, Labor and Welfare. Written informed consent was obtained from participants enrolled after April 2015, and written or verbal consent was obtained from others.

Patient Involvement. Patients were not involved in the design of or recruitment of this study.

REFERENCES

1. Smolen JS, Landewe 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.

2. Kawahito Y, Morinobu A, Kaneko Y, et al. Drug treatment algorithm and recommendations from the 2020 update of the Japan College of Rheumatology clinical practice guidelines for the management of rheumatoid arthritis-secondary publication. Mod Rheumatol. 2023;33(1):21–35.

3. Ytterberg SR, Bhatt DL, Mikuls TR, et al. Cardiovascular and cancer risk with tofacitinib in rheumatoid arthritis. N Engl J Med. 2022;386(4): 316–26.

4. Taylor PC, Keystone EC, van der Heijde D, et al. Baricitinib versus placebo or adalimumab in rheumatoid arthritis. N Engl J Med. 2017;376(7): 652–62.

5. Fleischmann R, Mysler E, Hall S, et al. Efficacy and safety of tofacitinib monotherapy, tofacitinib with methotrexate, and adalimumab with methotrexate in patients with rheumatoid arthritis (ORAL Strategy): a phase 3b/4, double-blind, head-to-head, randomised controlled trial. Lancet. 2017;390(10093):457–68.

6. Fleischmann R, Pangan AL, Song IH, et al. Upadacitinib versus placebo or adalimumab in patients with rheumatoid arthritis and an inadequate response to methotrexate: results of a phase Iii, double-blind, randomized controlled trial. Arthritis Rheumatol. 2019;71(11):1788–800.

7. Combe B, Kivitz A, Tanaka Y, et al. Filgotinib versus placebo or adalimumab in patients with rheumatoid arthritis and inadequate response to methotrexate: a phase III randomised clinical trial. Ann Rheum Dis. 2021;80(7):848–58.

8. Wijbrandts CA, Tak PP. Prediction of response to targeted treatment in rheumatoid arthritis. Mayo Clin Proc. 2017;92(7):1129–43.

9. Callaghan CA, Boyter AC, Mullen AB, McRorie ER. Biological therapy for rheumatoid arthritis: is personalised medicine possible? Eur J Hosp Pharm. 2014;21(4):229–37.

10. Julia A, Erra A, Palacio C, et al. An eight-gene blood expression profile predicts the response to infliximab in rheumatoid arthritis. PLoS ONE. 2009;4(10): e7556.

11. Toonen EJ, Gilissen C, Franke B, et al. Validation study of existing gene expression signatures for anti-TNF treatment in patients with rheumatoid arthritis. PLoS ONE. 2012;7(3): e33199.

12. Thomson TM, Lescarbeau RM, Drubin DA, et al. Blood-based identification of non-responders to anti-TNF therapy in rheumatoid arthritis. BMC Med Genomics. 2015;8:26.

13. Bek S, Bojesen AB, Nielsen JV, et al. Systematic review and meta-analysis: pharmacogenetics of anti-TNF treatment response in rheumatoid arthritis. Pharmacogenomics J. 2017;17(5):403–11.

14. Aterido A, Canete JD, Tornero J, et al. A Combined transcriptomic and genomic analysis identifies a gene signature associated with the response to anti-TNF therapy in rheumatoid arthritis. Front Immunol. 2019;10:1459.

15. Silvagni E, Giollo A, Sakellariou G, et al. One year in review 2020: novelties in the treatment of rheumatoid arthritis. Clin Exp Rheumatol. 2020;38(2):181–94.

16. Nerviani A, Di Cicco M, Mahto A, et al. A Pauci-immune synovial pathotype predicts inadequate response to TNFalpha-blockade in rheumatoid arthritis patients. Front Immunol. 2020;11:845.

17. Tao W, Concepcion AN, Vianen M, et al. Multiomics and machine learning accurately predict clinical response to adalimumab and etanercept therapy in patients with rheumatoid arthritis. Arthritis Rheumatol. 2021;73(2):212–22.

18. Humby F, Durez P, Buch MH, et al. Rituximab versus tocilizumab in anti-TNF inadequate responder patients with rheumatoid arthritis (R4RA): 16-week outcomes of a stratified, biopsy-driven, multicentre, open-label, phase 4 randomised controlled trial. Lancet. 2021;397(10271):305–17.

19. Ho CH, Silva AA, Giles JT, et al. Validation of a bioassay for predicting response to tumor necrosis factor inhibitors in rheumatoid arthritis. Arthritis Rheumatol. 2021;73(6):1086–7.

20. Vastesaeger N, Kutzbach AG, Amital H, et al. Prediction of remission and low disease activity in disease-modifying anti-rheumatic drug-refractory patients with rheumatoid arthritis treated with golimumab. Rheumatology (Oxford). 2016;55(8): 1466–76.

21. Aletaha D, Wang X, Zhong S, et al. Differences in disease activity measures in patients with rheumatoid arthritis who achieved DAS, SDAI, or CDAI remission but not Boolean remission. Semin Arthritis Rheum. 2020;50(2):276–84.

22. Aletaha D, Wang X, Zhong S, et al. Differences in disease activity measures in patients with rheumatoid arthritis who achieved DAS, SDAI, or CDAI remission but not Boolean remission. Semin Arthritis Rheum. 2020;50(4):643–4.

23. Miyata H, Sonomoto K, Fukuyo S, et al. Computed tomography for malignancy screening in patients with rheumatoid arthritis before initiation of disease-modifying antirheumatic drug. Rheumatology (Oxford). 2023;62(10):3339–49.

24. Fraenkel L, Bathon JM, England BR, et al. 2021 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Rheumatol. 2021;73(7):1108–23.

25. Holroyd CR, Seth R, Bukhari M, et al. The British Society for Rheumatology biologic DMARD safety guidelines in inflammatory arthritis. Rheumatology (Oxford). 2019;58(2):372.

26. Sepriano A, Kerschbaumer A, Bergstra SA, et al. Safety of synthetic and biological DMARDs: a systematic literature review informing the 2022 update of the EULAR recommendations for the management of rheumatoid arthritis. Ann Rheum Dis. 2023;82(1):107–18.

27. Ganhao S, Lucas R, Fonseca JE, et al. Remission and low disease activity matrix tools: results in real-world rheumatoid arthritis patients under anti-TNF therapy. Acta Reumatol Port. 2020;45(4):245–52.

28. Poddubnyy D. Precision medicine in rheumatology: are we getting closer? Lancet. 2021;397(10271): 258–9.

29. Guan Y, Zhang H, Quang D, 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(12):1987–96.

30. Nagy G, Roodenrijs NMT, Welsing PM, et al. EULAR definition of difficult-to-treat rheumatoid arthritis. Ann Rheum Dis. 2021;80(1):31–5.