SUPPLEMENTARY METHOD:

All statistical analyses were performed using R (version 4.0.2, R Foundation for Statistical Computing, Vienna, Austria). A p-value of <0.05 was considered statistically significant.

 Latent class mixed modelling was used to investigate trajectories of monocyte count changes in the first year after initiation of anti-TNF. The monocyte count change was adjusted for age, gender, baseline disease activity score, C-reactive protein, types of anti-TNF, antibody status, concomitant disease modifying anti-rheumatic drug (DMARD), and prednisolone. The final model of 3-latent classes was chosen by lowest Bayesian information criteria (BIC) with quadratic time effect and beta spline using lcmm function from the lcmm(1) package (Entropy 0.914) for R. For latent class mixed modelling, missing values were imputed using multivariate imputation by chained equations of the Markov chain Monte Carlo method under the 'missing at random' assumption (2).

Variable selection

We applied partial least square regression for variable selection for the final model shown in Figure 1E, and this was further verified by the supervised machine learning algorithm of random forest survival analysis.

For partial least square regression 10-fold cross-validation with 50 repetitions was applied to prevent model overfitting by using plsRcox (3). Model with the best discriminatory performance (compared by overall estimation error rate) was selected and variables contributing to the model were chosen for the final multivariate model.

Random forests for survival analysis was done using maximally selected rank statistics with ranger (4) package. For random forest model, decision trees were built to construct the model using a bootstrap dataset consisting of randomly selected samples from the original dataset (number of trees, n=10,000, msplit=5). Top variables with higher permutation importance score to predict discontinuation of TNFi were reported.

Supplementary Table 1: Baseline characteristics of patients who attained remission from two independent cohorts.

Variables	Cohort 1, N = 92[1]	Cohort 2, N = 43[1]
Age, Years	59 (14)	56 (17)
Gender		
Female	68 / 92 (74%)	34 / 43 (79%)
Male	24 / 92 (26%)	9 / 43 (21%)
Race		
Non-Caucasian	15 / 92 (16%)	18 / 43 (42%)
Caucasian	77 / 92 (84%)	25 / 43 (58%)
Disease duration, months	17 (7)	19 (6)
Tender joint count	13 (8)	14 (6)
Swollen joint count	7.7 (5.0)	7.1 (3.4)
Patient global assessment (range 0-100 mm)	73 (16)	75 (17)
C-Reactive protein, mg/L	19 (25)	13 (11)
DAS28-ESR	6.25 (0.89)	5.92 (1.18)
Rheumatoid factor		
Positive	62 / 92 (67%)	36 / 43 (84%)
ACPA (anti-citrullinated peptide antibody)		
Positive	65 / 92 (71%)	31 / 43 (72%)
Concomitant Methotrexate	55 / 92 (60%)	29 / 43 (67%)
Concomitant Hydroxychloroquine	27 / 92 (29%)	9 / 43 (21%)
Concomitant Prednisolone	10 / 92 (11%)	13 / 43 (30%)
Type of TNFi (Tumour necrotic factor inhibitor)		
Etanercept	42 / 92 (46%)	20 / 43 (47%)
Adalimumab	50 / 92 (54%)	23 / 43 (53%)

[1]*Statistics presented: mean (SD); n = number (%)*

Supplementary Table 2: Baseline characteristics of patients shown for each of the 3 latent classes.

Variables	Class 1, N = 41[1]	Class 2, N = 19[1]	Class 3, N = 75[1]
Age, Years	58 (17)	65 (14)	56 (14)
Gender			
Female	28 (68%)	17 (89%)	57 (76%)
Male	13 (32%)	2 (11%)	18 (24%)
Race			
Non-Caucasian	11 (27%)	5 (26%)	15 (20%)
Caucasian	30 (73%)	14 (74%)	60 (80%)
Disease duration, months	16 (4)	15 (10)	17 (7)
Tender joint count	13 (7)	12 (6)	13 (7)
Swollen joint count	8.3 (5.5)	7.7 (3.8)	8.2 (4)
Patient global assessment (range 0-100 mm)	71 (18)	76 (13)	74 (16)
C-Reactive protein, mg/L	19 (23)	11 (13)	17 (23)
DAS28-ESR	5.88 (0.98)	5.66 (0.86)	5.88 (1.06)
Rheumatoid factor			
Positive	29 (71%)	14 (74%)	55 (73%)
ACPA (anti-citrullinated peptide			
Positive	30 (70%)	13 (68%)	53 (71%)
Concomitant Methotrexate	33 (80%)	9 (47%)	42 (56%)
Concomitant Hydroxychloroquine	12 (29%)	5 (26%)	19 (25%)
Concomitant Prednisolone	6 (15%)	5 (26%)	12 (16%)
Type of TNFi (Tumour necrotic			
Etanercept	21 (51%)	9 (47%)	33 (44%)
Adalimumab	20 (49%)	10 (53%)	42 (56%)

[1]*Statistics presented: mean (SD); n = number (%)*
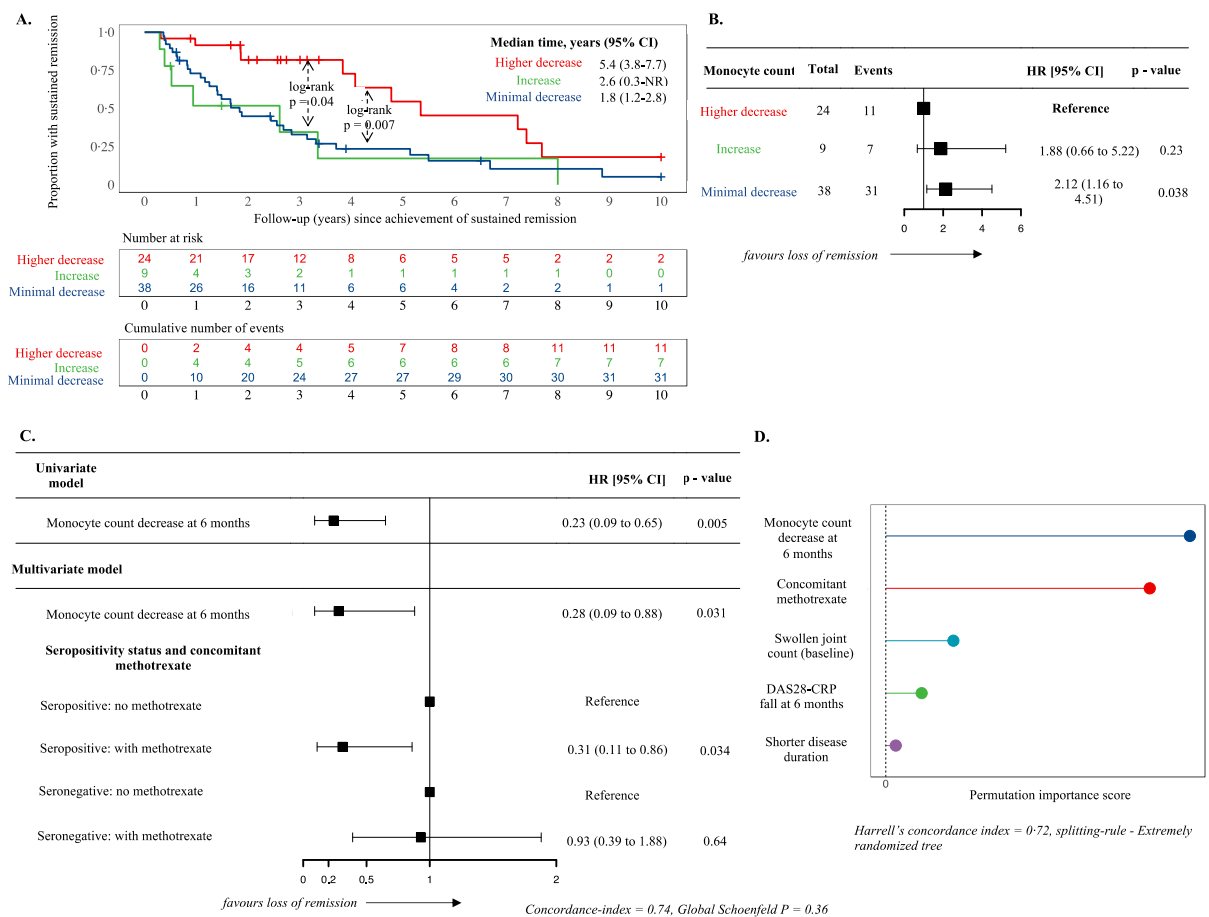
Supplementary Figure 1 (A-D): Monocyte count change and other variables predicting loss of remission in anti-TNF treated rheumatoid arthritis patients (where remission was defined by CDAI<2.8). A. Kaplan-Meier curve of the monocyte latent class to predict loss of remission of anti-TNF. **B** Cox-regression to predict the loss of remission stratified by the latent classes, adjusted by propensity score. **C.** Univariate and Multiple cox-regression model to predict loss of remission of TNFi using monocyte count change (by each 0.1-unit decrease) at 6 months and adjusted by variables selected by partial least square regression. **D.** Random Forest model - to predict loss of remission of anti-TNF. The top 5-variables (ranked by permutation importance score) contributing to the loss of remission are shown.

HR = Hazard ratio, 95% CI = 95 % confidence interval.

Reference

1. Cecile Proust-Lima VP, Benoit Liquet (2017). Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. Journal of Statistical Software,. 2017;78(2):1-56.
2. Stef van Buuren KG-O. Mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011;45(3):1-67.
3. Maumy-Bertrand FBaM. Partial Least Squares Regression for Cox Models and Related Techniques. R package. 2021;version 1.7.6.
4. Marvin N. Wright AZ. A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software. 2017;77(1):17.