




# Predicting Remission Among Patients With Rheumatoid Arthritis Starting Tocilizumab Monotherapy: Model Derivation and Remission Score Development

Jamie E. Collins,<sup>1</sup>  Fredrik D. Johansson,<sup>2</sup> Sara Gale,<sup>3</sup> Seoyoung Kim,<sup>1</sup>  Swastina Shrestha,<sup>1</sup> David Sontag,<sup>2</sup> Jacklyn Stratton,<sup>1</sup> Huong Trinh,<sup>3</sup> Chang Xu,<sup>1</sup> Elena Losina,<sup>1</sup>  and Daniel H. Solomon<sup>1</sup>

**Objective.** Most patients with rheumatoid arthritis (RA) strive to consolidate their treatment from methotrexate combinations. The objective of this analysis was to identify patients with RA most likely to achieve remission with tocilizumab (TCZ) monotherapy by developing and validating a prediction model and associated remission score.

**Methods.** We identified four TCZ monotherapy randomized controlled trials in RA and chose two for derivation and two for internal validation. Remission was defined as a Clinical Disease Activity Index score less than 2.8 at 24 weeks post randomization. We used logistic regression to assess the association between each predictor and remission. After selecting variables and assessing model performance in the derivation data set, we assessed model performance in the validation data set. The cohorts were combined to calculate a remission prediction score.

**Results.** The variables selected included younger age, male sex, lower baseline Clinical Disease Activity Index score, shorter RA disease duration, region of the world (Europe and South America [increased odds of remission] versus Asia and North America), no previous exposure to disease-modifying antirheumatic drugs and/or methotrexate, lower baseline Health Assessment Questionnaire Disability Index score, and baseline hematocrit. The area under the receiver operating characteristic curve was 0.739 in the derivation data set and 0.756 in the validation data set. Patients were categorized into three remission prediction categories based on the remission prediction score: 40% in the low (less than 10% probability of remission), 45% in the intermediate (10%-25% probability), and 15% in the moderate remission prediction category (greater than 25% probability).

**Conclusion.** We used easily accessible factors to develop a remission prediction score to predict RA remission at 24 weeks after initializing TCZ monotherapy. These results may provide guidance to clinicians tailoring treatment options based on clinical characteristics.

## INTRODUCTION

Remission in rheumatoid arthritis (RA) is the target for most patients and has increasingly become an achievable goal for many (1). However, it is difficult to determine which patients will reach remission through use of a given drug. Better tools to predict which patients are likely to reach remission with a specific drug would enable clinicians and patients to make better informed treatment decisions. Risk scores are a useful method for trans-

lating epidemiologic findings into clinical practice (2). Methods for risk score derivation and validation have been well described (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis [TRIPOD]) (3); such methods require adequate samples of patients that are well characterized with respect to treatments and outcomes.

Randomized controlled trials (RCTs) provide high-quality data that can be used for risk score derivation studies. Most recent RCTs in RA compare the agent of interest to a standard

Funding for this project was provided by Roche/Genentech to Brigham and Women's Hospital.

<sup>1</sup>Jamie E. Collins, PhD, Swastina Shrestha, MS, Elena Losina, PhD, Seoyoung Kim MD, ScD, Jacklyn Stratton, BS, Chang Xu, MS, Daniel H. Solomon, MD, MPH: Brigham and Women's Hospital, Boston, Massachusetts; <sup>2</sup>Fredrik D. Johansson, PhD, David Sontag, PhD: Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge; <sup>3</sup>Sara Gale, MPH, PHD, Huong Trinh, MS: Genentech, South San Francisco, California.

Dr. Gale is an employee of Roche/Genentech. Dr. Kim receives research support through grants to Brigham and Women's Hospital from Bristol-Myers Squibb, Pfizer, and AbbVie. Dr. Trinh is an employee of Roche/Genentech.

Dr. Losina receives research support through grants to Brigham and Women's Hospital from Pfizer, Flexion, and Samumed and is a consultant to Regeneron. Dr. Solomon receives salary support through research contracts to Brigham and Women's Hospital from AbbVie, Amgen, Corrona, Janssen, Pfizer, and Roche/Genentech. No other disclosures relevant to this article were reported.

Address correspondence to Jamie E. Collins, PhD, Brigham and Women's Hospital, Orthopaedic and Arthritis Center for Outcomes Research, 75 Francis Street, BTM 5016, Boston, MA 02115. E-mail: jcollins13@bwh.harvard.edu.

Submitted for publication October 1, 2019; accepted in revised form October 4, 2019.

treatment, such as methotrexate or a tumor necrosis factor (TNF) inhibitor. The majority of RCTs with biologic disease-modifying antirheumatic drugs (bDMARDs) have added the treatment of interest or a placebo to a background of methotrexate. This has been the chosen design because most bDMARDs are more effective when given with methotrexate. Tocilizumab (TCZ) is a biologic therapy for RA that has been shown to work well with or without concurrent methotrexate in helping patients achieve disease remission (4).

In light of this background, we sought to derive a prediction score for remission with TCZ monotherapy. We accessed patient-level data from four TCZ monotherapy RCTs (4–7): two were used to derive the prediction model, and two were used for internal validation. We used the internally validated model to estimate the remission prediction score in the total population.

## MATERIALS AND METHODS

**Study design and sample.** We followed the TRIPOD recommendations for derivation and validation of clinical risk prediction models (3). These recommendations describe the appropriate selection of the derivation and validation cohorts, variable selection strategies, model estimation, validation assessment, and risk score calculation. We identified four RCTs among patients with RA, ACT-RAY, ADACTA, AMBITION, and FUNCTION, and included the TCZ monotherapy arm from each (4–7). The TRIPOD statement recommends nonrandomly splitting the data into derivation and validation groups to allow for nonrandom variation between the two data sets; thus, we split the data based on the study, with patients from ACT-RAY and FUNCTION in the derivation cohort and patients from ADACTA and AMBITION in the validation cohort (3,8).

Data from the four RCTs were de-identified and supplied by the manufacturer after we obtained institutional review board approval from the Partners Healthcare Human Studies Committee. The data elements in each trial were largely collected and recorded in a consistent manner across trials. We examined case report forms and harmonized the variables when necessary.

**Study outcome (remission).** The primary outcome was disease remission at week 24 post randomization, defined by a Clinical Disease Activity Index (CDAI) score less than 2.8 (refs. 9,10). The CDAI remission criteria do not require a laboratory measurement and were chosen to make the outcome clinically useful and straightforward to obtain. The TRIPOD recommendations suggest the use of variables and outcomes that are clinically accessible (3).

**Potential predictors.** We considered a range of variables assessed at baseline as potential predictors of remission. Demographic/anthropometric variables included age, sex, race (white versus nonwhite), geographic region (North America, South

America, Europe, and Asia/Australia), and body mass index (BMI) (less than 25 kg/m<sup>2</sup>, normal weight; 25–30 kg/m<sup>2</sup>, overweight; and greater than 30 kg/m<sup>2</sup>, obesity). RA characteristics included baseline CDAI score, disease duration, Health Assessment Questionnaire Disability Index (HAQ-DI) score, C-reactive protein (CRP) level, erythrocyte sedimentation rate (ESR), and hematocrit (11). To evaluate possible nonlinear associations between remission and the continuous predictors, HAQ-DI score, CRP level, ESR, and hematocrit, we created quartiles based on the distribution of these parameters in the derivation data set. Quartiles were recategorized based on bivariate associations with remission. Final categories were as follows: HAQ-DI score: 0 to 1 versus 1.1 to 2 versus 2.1 to 3; CRP level: less than or equal to 4.13 µg/ml versus greater than 4.13 µg/ml; hematocrit: less than or equal to 36% versus greater than 36%; and ESR: less than or equal to 28 mm/h versus greater than 28 mm/h.

Prior RA treatments that were considered were baseline use of oral corticosteroids and past use of disease-modifying antirheumatic drugs (DMARDs) and methotrexate. Past DMARD use and past methotrexate use were combined into a three-level variable: neither treatment, both treatments, and past use of DMARDs but not methotrexate. Comorbid conditions included cancer, cardiovascular disease (CVD), hypertension, chronic obstructive pulmonary disease or asthma, hyperlipidemia, liver disease, renal disease, and diabetes. We created a categorical variable for number of comorbidities: zero, one, and two or more. We evaluated both individual comorbidities and the summary score; cancer, liver disease, and renal disease were not evaluated individually because of a small number of subjects reporting these conditions but were included in the summary variable.

In our primary analysis, we selected subjects with complete data on all covariates and the outcome. In the sensitivity analysis, we augmented missing data by imputation.

**Statistical methods.** We used logistic regression to evaluate the association between remission and potential predictors. In the primary analysis, we used the odds ratio (OR) (OR in the bivariate logistic regression less than or equal to 0.67 or greater than or equal to 1.5) to determine which predictors to advance to the multivariable model. In the secondary analysis, we considered stepwise selection based on the Akaike information criterion (AIC) and stepwise selection based on the Bayesian information criterion (BIC). The AIC and BIC differ with respect to model fitting: the AIC tends to favor more complex models that risk overfitting, whereas the BIC tends to favor less complex models that risk underfitting (12). We forced sex, age, and baseline CDAI score into each model. We present the area under the receiver operating characteristic curve (AUROC), the AIC, and the BIC for each model. The AUROCs are presented both with and without 10-fold cross-validation. We assessed calibration graphically by plotting the observed versus the predicted probability of remission. We used locally estimated

scatterplot smoothing (LOESS) to assess calibration across the range of predicted values (13). The calibration slope quantifies the relationship between observed and expected probabilities; a well-calibrated model will have a calibration slope close to 1 (ref. 14).

We performed variable selection and assessed model fit in the derivation data set and then ran the models selected in the derivation analysis in the validation data set. We compared the parameter estimates in the derivation data set with those in the validation data set and, as a sensitivity analysis, updated the prediction model to exclude variables with different directions of association in the derivation data set versus the validation data set (8). We ran the final multivariable models on the combined derivation plus validation data set to derive the final remission prediction score. To address clinical decision-making, three remission prediction groups were defined: low probability of remission, less than or equal to 10%; intermediate probability of remission, between 10% and 25%; and moderate probability of remission, greater than or equal to 25%.

In a sensitivity analysis, we imputed missing covariates and outcome data using Markov chain Monte Carlo with multiple chains (15,16). We used observed baseline covariates as well as CDAI values from weeks 2 (AMBITION and FUNCTION), 4, 8, 12, 16, and 20. For subjects with CDAI remission data missing at 24 weeks, we first checked to see if remission information was available at any time between 20 and 36 weeks. If so, we used that value. If not, we used the imputed 24-week value. To more accurately reflect the uncertainty in missing values, we created 10 imputed data sets and then used the MIANALYZE procedure in SAS to combine the results across the 10 imputations (17).

All analyses were conducted using SAS version 9.4 (SAS Institute, Inc).

## RESULTS

**Patient sample characteristics.** A total of 1019 subjects were enrolled in the TCZ monotherapy arm across the four trials: 276 in ACT-RAY, 163 in ADACTA, 288 in AMBITION, and 292 in FUNCTION. Seventy-eight (8%) subjects were missing CDAI data at week 24, and 88 (9%) subjects were missing one or more covariates. The primary analysis, which included patients with complete data, included 853 (84%) subjects. The included and excluded participants were similar with respect to baseline clinical and demographic characteristics (see Appendix Table 1). Among the subjects included in the analytic data set, the median age was 53 years, 80% were women, and 80% were white. At baseline, the median HAQ-DI score was 1.6, the median CDAI score was 40.1, and the median disease duration was 21 months.

There were differences in study characteristics and the setting across the four RCTs (see Appendix Table 2). The ACT-RAY, ADACTA, and AMBITION trials enrolled subjects with established RA and inadequate response to methotrexate, whereas the FUNCTION

trial enrolled methotrexate-naïve subjects with early RA. Whereas the median baseline CDAI score was similar across the studies, ranging from 37.9 in ACT-RAY to 42.1 in AMBITION, disease duration was varied, with a median RA disease duration of 3 months in FUNCTION, 2.8 years in AMBITION, 4.7 years in ADACTA, and 5.2 years in ACT-RAY. The percentage achieving remission at 24 weeks ranged from 9.5% (ACT-RAY) to 22.6% (FUNCTION).

The derivation data set included 473 subjects (221 from ACT-RAY and 252 from FUNCTION), and the validation data set included 380 subjects (143 from ADACTA and 237 from AMBITION). The subjects in the derivation and validation data sets were similar with respect to demographic and clinical characteristics (see Table 1). Seventy-eight (16%) subjects in the derivation data set and 49 (13%) subjects in the validation data set achieved remission at 24 weeks.

**Derivation and validation of the prediction model.** In the derivation data set, the following patient characteristics were associated with increased odds of remission: located in Europe and South America versus North America, no history of DMARD or methotrexate use or only past DMARD use versus history of both, shorter disease duration, low (0-1) HAQ-DI score versus higher (>1), higher ESR (greater than 28 mm/h), higher hematocrit (greater than 36%), no CVD, and diabetes (see Appendix Table 3).

Results of multivariable logistic regression for the three selection approaches are presented in Table 2. In the AIC and BIC approaches, the same set of variables were selected. Model fit was similar: the AUROC was 0.739 in the OR-based selection model and 0.728 in the AIC/BIC model. These values dropped to 0.656 and 0.670, respectively, under 10-fold cross-validation. Region, history of DMARD or methotrexate use, and baseline HAQ-DI score were common between the two models, whereas in the OR-based approach, ESR, hematocrit, CVD, and diabetes were additionally selected. The models were well calibrated, with calibration slopes close to 1, and LOESS graphs suggest good calibration, with the smoothed line close to the diagonal line for all estimated probabilities (see Figure 1A, Appendix Figure 1A).

After selecting predictors in the derivation data set, we ran the selected models in the validation data set. The models using the variables selected in the derivation data set performed well in the validation data set: the AUROC was 0.756 (0.632 using 10-fold cross-validation) for the OR-based selection model and 0.731 (0.639 using 10-fold cross-validation) for the AIC/BIC model (see Table 2). The models were well calibrated, with calibration slopes close to 1 and LOESS graphs suggesting adequate calibration (see Figure 1B, Appendix Figure 1B).

**Remission prediction score.** For each predictor, the strength and direction of associations were similar in the derivation and validation data sets. The exception was history of DMARD or methotrexate use: a history of DMARD, but not methotrexate, use (versus a history of both) was associated with modest increased

odds of remission in the derivation data set but decreased odds of remission in the validation data set. As a sensitivity analysis, we reran the selection procedures, omitting this variable as a potential predictor. Results of multivariable logistic regression in the combined derivation plus validation data set, including ORs and parameter estimates for the remission prediction score, are pro-

vided in Table 3. Region of the world was associated with odds of remission: subjects in Europe (OR 2.6) and South America (OR 2.8) had increased odds of remission compared with subjects in North America. Subjects with less severe disease at baseline had higher odds of remission, including those with no prior exposure to methotrexate or DMARDs (OR 2.0), and those with a baseline HAQ-DI of

**Table 1.** Cohort characteristics for derivation and validation samples<sup>a</sup>

Characteristic	Derivation (n = 473)	Validation (n = 380)
Remission CDAI (CDAI < 2.8 at wk 24)		
No	395 (83.5%)	331 (87.1%)
Yes	78 (16.5%)	49 (12.9%)
Age, y	52.0 (44.0, 60.0)	53.0 (43.5, 61.5)
BMI, kg/m <sup>2</sup>	26.5 (23.2, 30.3)	26.6 (23.9, 31.0)
BMI category		
Normal	182 (38.5%)	134 (35.3%)
Overweight	161 (34.0%)	134 (35.3%)
Obesity	130 (27.5%)	112 (29.5%)
Sex		
Female	364 (77.0%)	316 (83.2%)
Male	109 (23.0%)	64 (16.8%)
Race		
Nonwhite	81 (17.1%)	92 (24.2%)
White	392 (82.9%)	288 (75.8%)
Region		
Asia, Australia, New Zealand, Africa, Turkey	42 (8.9%)	46 (12.1%)
Europe	274 (57.9%)	138 (36.3%)
North America	91 (19.2%)	114 (30.0%)
South America	66 (14.0%)	82 (21.6%)
HAQ-DI	1.6 (1.0, 2.0)	1.6 (1.1, 2.0)
Disease duration, y	0.9 (0.2, 4.8)	3.6 (1.0, 9.6)
Baseline CDAI	39.1 (30.1, 48.6)	41.3 (32.8, 50.5)
Patient's global assessment of disease activity	69.0 (52.0, 81.0)	68.0 (50.0, 81.0)
Physician's global assessment of disease activity	64.0 (51.0, 76.0)	64.0 (52.0, 76.0)
Swollen joint count (28 joints)	10.0 (7.0, 15.0)	12.0 (8.0, 16.0)
Tender joint count (28 joints)	15.0 (10.0, 21.0)	17.0 (11.0, 23.0)
ESR result, mm/h	40.0 (29.0, 60.0)	42.0 (30.0, 61.5)
High-sensitivity CRP, mg/l	10.2 (4.0, 24.5)	15.7 (6.9, 38.5)
Hematocrit, %	0.39 (0.36, 0.41)	0.39 (0.36, 0.42)
Hemoglobin, mg/dl	12.6 (11.7, 13.8)	12.6 (11.5, 13.5)
Past DMARD and/or methotrexate use		
Both no	193 (40.8%)	98 (25.8%)
Both yes	221 (46.7%)	220 (57.9%)
DMARD, yes; methotrexate, no	59 (12.5%)	62 (16.3%)
Past non-TNF inhibitor biologic DMARD use	1 (0.2%)	1 (0.3%)
Past TNF blocker use	1 (0.2%)	19 (5.0%)
Baseline use of oral corticosteroids	221 (46.7%)	200 (52.6%)
History of cancer (not nonmelanomatous skin cancer)	7 (1.5%)	9 (2.4%)
History of CVD (HF, CAD, PVD)	38 (8.0%)	26 (6.8%)
History of hypertension	138 (29.2%)	133 (35.0%)
History of COPD/asthma	29 (6.1%)	28 (7.4%)
History of hyperlipidemia	76 (16.1%)	76 (20.0%)
History of liver disease	5 (1.1%)	7 (1.8%)
History of renal disease	2 (0.4%)	3 (0.8%)
History of diabetes	33 (7.0%)	27 (7.1%)
Number of comorbidities		
None	253 (53.5%)	192 (50.5%)
One	145 (30.7%)	105 (27.6%)
Two or more	75 (15.9%)	83 (21.8%)

Abbreviation: BMI, body mass index; CAD, coronary artery disease; CDAI, Clinical Disease Activity Index; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; CVD, cardiovascular disease; DMARD, disease-modifying antirheumatic drug; ESR, erythrocyte sedimentation rate; HAQ-DI, Health Assessment Questionnaire Disability Index; HF, heart failure; PVD, peripheral vascular disease; TNF, tumor necrosis factor.

<sup>a</sup>n (%) is presented for categorical variables. The median (25th percentile, 75th percentile) is presented for continuous variables.

**Table 2.** Multivariable logistic regression<sup>a</sup>

Model Characteristics	Derivation		Validation	
	OR > 1.5 or OR < 0.67	Stepwise, AIC/BIC <sup>b</sup>	OR > 1.5 or OR < 0.67	Stepwise, AIC/BIC <sup>c</sup>
AUROC (95% CI)	0.739 (0.679-0.800)	0.728 (0.666-0.790)	0.756 (0.689-0.824)	0.731 (0.657-0.806)
AUROC (10-fold cross-validation) (95% CI)	0.656 (0.588-0.725)	0.669 (0.600-0.737)	0.632 (0.549-0.716)	0.639 (0.552-0.727)
AIC	407.046	401.738	288.281	283.436
BIC	473.591	447.488	351.330	326.777
Calibration slope	0.995	0.996	0.971	1.023
Predictors Included	P and OR (95% CI)	P and OR (95% CI)	P and OR (95% CI)	P and OR (95% CI)
Age (OR per 1-y increase)	0.1506 0.98 (0.96-1.01)	0.1135 0.98 (0.96-1.00)	0.1688 0.98 (0.95-1.01)	0.1424 0.98 (0.96-1.01)
Baseline CDAI (OR per 1-U increase)	0.1916 0.98 (0.96-1.01)	0.2277 0.99 (0.96-1.01)	0.0490 0.97 (0.94-1.00)	0.0919 0.97 (0.95-1.00)
Disease duration (OR per 12-mo increase)	0.4009 0.97 (0.91-1.04)	... ...	0.8205 0.99 (0.94-1.05)	... ...
Sex	0.6441	0.6884	0.5047	0.5748
Female	Reference	Reference	Reference	Reference
Male	1.16 (0.62-2.17)	1.13 (0.61-2.09)	1.32 (0.58-3.01)	1.26 (0.57-2.78)
Region	0.0159	0.0137	0.0043	0.0064
Europe	2.33 (1.03-5.27)	2.29 (1.02-5.15)	3.93 (1.63-9.46)	3.71 (1.56-8.81)
North America	Reference	Reference	Reference	Reference
Asia, Australia, New Zealand, Africa, Turkey	0.99 (0.3-3.23)	0.95 (0.29-3.04)	0.83 (0.20-3.42)	0.88 (0.22-3.62)
South America	3.62 (1.43-9.17)	3.55 (1.42-8.85)	1.74 (0.61-5.01)	1.71 (0.61-4.82)
Past DMARD and methotrexate use	0.0194	0.0002	0.0690	0.0825
Both yes	Reference	Reference	Reference	Reference
Both no	2.59 (1.22-5.53)	3.56 (1.91-6.65)	1.38 (0.63-3.02)	1.26 (0.61-2.60)
DMARD, yes and methotrexate, no	1.19 (0.42-3.32)	1.62 (0.63-4.16)	0.32 (0.10-1.00)	0.32 (0.10-0.99)
ESR	0.1905	...	0.8256	...
Quartile 1 (2-28 mm/h)	Reference	...	Reference	...
Quartiles 2-4 (29-160 mm/h)	1.6 (0.79-3.21)	...	1.10 (0.47-2.55)	...
HAQ-DI	0.0647	0.0444	0.3217	0.3209
Quartile 1 (0-1)	1.77 (0.76-4.13)	1.80 (0.79-4.07)	1.91 (0.64-5.71)	1.83 (0.63-5.34)
Quartiles 2-3 (1.125-2)	0.84 (0.40-1.77)	0.83 (0.40-1.71)	1.11 (0.42-2.90)	1.06 (0.41-2.72)
Quartile 4 (2.125-3)	Reference	Reference	Reference	Reference
Hematocrit	0.3272	...	0.5292	...
Quartile 1 (29%-36%)	Reference	...	Reference	...
Quartiles 2-4 (37%-51%)	1.40 (0.72-2.72)	...	0.79 (0.38-1.65)	...
CVD	0.4435	...	0.3742	...
Yes	Reference	...	Reference	...
No	1.57 (0.5-4.96)	...	2.09 (0.41-10.67)	...
Diabetes	0.4003	...	0.0333	...
Yes	1.50 (0.58-3.89)	...	3.52 (1.10-11.23)	...
No	Reference	...	Reference	...

Abbreviation: AIC, Akaike information criterion; AUROC, area under the receiver operating characteristic curve; BIC, Bayesian information criterion; CDAI, Clinical Disease Activity Index; CI, confidence interval; CVD, cardiovascular disease; DMARD, disease-modifying antirheumatic drug; ESR, erythrocyte sedimentation rate; HAQ-DI, Health Assessment Questionnaire Disability Index; OR, odds ratio.

<sup>a</sup>Presented for predictors: P value and OR (95% CI).

<sup>b</sup>AIC and BIC selection results in the same model.

0-1 (OR 1.7), whereas each one point increase in baseline CDAI was associated with a 0.98-fold decreased odds of remission.

Using the parameter estimates from Table 3, we can compute the risk of remission associated with various combinations of baseline covariates for our primary model:

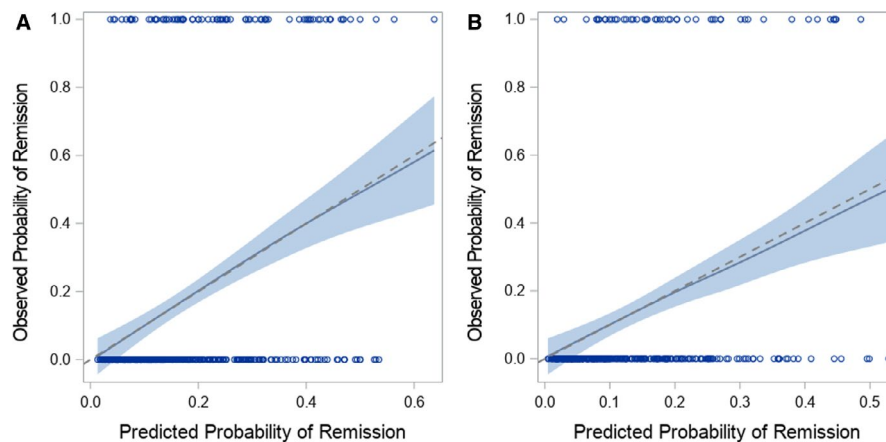
Risk of remission = (−0.0166 × age [per year]) − (0.0206 × CDAI) − (0.0229 × disease duration [per year]) + (0.1926 × sex [male = 1]) − (0.0844 × world region Asia [Asia, Australia, New Zealand, South Africa, or Turkey = 1]) + (0.9616 × world region Europe [Europe = 1]) + (1.034 × world region South America [South America = 1]) + (0.6933 × methotrexate/DMARD history [history of neither = 1]) − (0.3471 × methotrexate/DMARD history [history of DMARD but

not methotrexate = 1]) + (0.4247 × ESR [ESR >28 = 1]) + (0.5365 × HAQ-DI [HAQ-DI score ≤1 = 1]) − (0.0881 × HAQ-DI [1 < HAQ-DI score ≤2 = 1]) + (0.01618 × hematocrit [Hematocrit >0.36 = 1]) + (0.5379 × CVD [no = 1]) + (0.7239 × diabetes [yes = 1]).

The probability of remission can then be calculated as follows:

$$P = \frac{e^{(risk\ score-2.1)}}{1 + e^{(risk\ score-2.1)}}$$

For example, a male subject aged 52 years in North America with a baseline CDAI score of 41, a disease duration of 5 years, a history of DMARD (but not methotrexate) use, an ESR between



**Figure 1.** Locally estimated scatterplot smoothing (LOESS) curve to assess calibration in the odds ratio (OR)-based model. **A**, Derivation data set. **B**, Validation data set. The predicted probability of remission from the multivariable logistic regression model is along the x-axis, and the actual probability of remission is along the y-axis. The dashed diagonal line represents the line of perfect calibration. Systematic deviation from the diagonal line indicates that the model may not be well calibrated.

29 and 160, an HAQ-DI score between 1.125 and 2, hematocrit between 37% and 51%, and CVD and diabetes has an estimated probability of disease remission at 24 weeks of 5.4%. A subject with the same characteristics but less severe disease (baseline CDAI score of 30, no DMARD or methotrexate history, baseline HAQ-DI score of 0-1) has an estimated probability of disease remission of 27.6%.

By using the remission prediction score, 337 subjects were deemed to have a low probability of remission (less than 10% predicted probability of remission). Of these, 25 (7%) achieved remission. Three hundred eighty-six subjects were deemed to have an intermediate probability of remission (10%-25% predicted probability); of these, 57 (15%) achieved remission. One hundred thirty subjects were deemed to have a moderate probability of remission (greater than 25% predicted probability); of these, 45 (35%) achieved remission.

**Sensitivity analysis for missing data.** After multiple imputation, we included all subjects: 568 in the derivation sample and 451 in the validation sample. The results were similar to those in the main analysis (see Appendix Tables 4 and 5). The discrimination was slightly attenuated, with an AUROC of 0.730 in the OR-based model and 0.719 in the AIC/BIC model in the derivation data set. In the validation data set, the AUROC was 0.736 in the OR-based model and 0.721 in the AIC/BIC model. The strength and direction of association were largely the same.

## DISCUSSION

We derived a remission prediction score using data from four RCTs of TCZ monotherapy in RA. We applied three selection methods to evaluate the robustness of the results. The models were well calibrated and demonstrated moderate discrimination. In addition to demographic variables, markers of baseline disease

severity were consistently selected as important predictors. Higher baseline CDAI scores, a history of exposure to methotrexate and DMARDs, and higher baseline HAQ-DI scores were associated with lower odds of remission. The remission prediction score allowed us to categorize patients into three distinct remission prediction categories, with a clear gradient of patients reaching remission: 40% in the low (less than 10% probability of remission), 45% in the intermediate (10%-25% probability), and 15% in the moderate (greater than 25% probability) remission prediction category.

We pursued these analyses to assess whether a formal prediction score modeling process would allow for the categorization of patients with RA and their probabilities of remission. Following the TRIPOD recommendations on prediction score derivation and validation, we used easily available variables and focused on an important clinical scenario: reaching remission by 24 weeks on TCZ monotherapy. This prediction score study demonstrates that easily available clinical variables may be useful in predicting remission with TCZ monotherapy.

We focused on which patients were more likely to reach remission at 24 weeks after using monotherapy with TCZ. The available data set did not allow us to examine how well this prediction score correlated with remission for other bDMARDs. A remission prediction score that considers a more generalized remission may provide useful information for clinicians and patients when considering prognosis. However, it is also possible that other variables would be important for other agents. This type of analysis should be considered across bDMARDs. Each trial included in this analysis had exclusion criteria regarding disease severity and prior exposure to biologic agents; only two subjects had prior exposure to non-TNF inhibitor bDMARDs. Prescribing patterns may be different in a real-world setting; we plan on using registry data for such analyses.

The current set of analyses might have been more strongly predictive of remission if we included biomarkers, such as omics data. However, we focused on clinically accessible data, as

**Table 3.** Multivariable logistic regression in the combined derivation and validation data sets

Model Characteristics	OR > 1.5 or OR < 0.67	Stepwise, AIC/BIC	Stepwise, AIC (Sensitivity Analysis)	
AUROC (95% CI)	0.715 (0.666-0.764)	0.701 (0.651-0.751)	0.693 (0.644-0.743)	
AUROC (10-fold cross-validation) (95% CI)	0.667 (0.614-0.720)	0.664 (0.611-0.717)	0.656 (0.604-0.707)	
AIC	681.030	680.549	688.407	
BIC	757.010	732.785	740.643	
Calibration slope	1.03	1.03	1.013	
	P and OR (95% CI)	Parameter Estimate (SE)	P and OR (95% CI)	Parameter Estimate (SE)
Intercept	...	-2.1000 (0.9192)	...	-1.2417 (0.7096)
Age (OR per 1-y increase)	0.0511 0.98 (0.97-1.00)	...	0.0355 0.98 (0.97-1.00)	...
Baseline CDAI (OR per 1-U increase)	0.0241 0.98 (0.96-1.00)	...	0.0372 0.98 (0.96-1.00)	...
Disease duration (OR per 12-mo increase)	0.2565 0.98 (0.94-1.02)	...	...	...
Sex	0.4417 Reference	...	0.4465 Reference	...
Female	Reference	...	Reference	...
Male	1.21 (0.74-1.98)	0.1926 (0.2503)	1.20 (0.75-1.94)	0.3127 (0.2389)
Region	0.0007 Reference	...	0.0014 Reference	...
Europe	2.62 (1.46-4.69)	0.9616 (0.2980)	2.48 (1.39-4.43)	0.8738 (0.2932)
North America	Reference	...	Reference	...
Asia, Australia, New Zealand, Africa, Turkey	0.92 (0.38-2.23)	-0.0844 (0.4513)	0.90 (0.37-2.18)	0.0183 (0.4449)
South America	2.81 (1.42-5.56)	1.0340 (0.3482)	2.60 (1.33-5.09)	1.0234 (0.3418)
Past DMARD and methotrexate use	0.0021 Reference	...	<0.0001 Reference	...
Both yes	Reference	...	Reference	...
Both no	2.00 (1.22-3.27)	0.6933 (0.2505)	2.38 (1.54-3.68)	...
DMARD, yes and methotrexate, no	0.71 (0.35-1.43)	-0.3471 (0.3595)	0.83 (0.42-1.64)	...
ESR	0.1124 Reference	...	...	...
Quartile 1 (2-28 mm/h)	1.53 (0.91-2.58)	0.4247 (0.2675)	...	0.1273 Reference
Quartiles 2-4 (29-160 mm/h)	Reference	...	Reference	...
HAQ-DI	0.0297 Reference	...	0.0208 Reference	...
Quartile 1 (0-1)	1.71 (0.89-3.27)	0.5365 (0.3308)	1.64 (0.87-3.09)	0.3765 (0.3184)
Quartiles 2-3 (1.125-2)	0.92 (0.52-1.62)	-0.0881 (0.2922)	0.86 (0.49-1.50)	-0.2103 (0.2851)
Quartile 4 (2.125-3)	Reference	...	Reference	...
Hematocrit	0.5090 Reference	...	...	...
Quartile 1 (29%-36%)	1.18 (0.73-1.90)	0.1618 (0.2451)	...	...
Quartiles 2-4 (37%-51%)	Reference	...	Reference	...
CVD	0.2545 Reference	...	...	...
Yes	Reference	...	Reference	...
No	1.71 (0.68-4.32)	0.5379 (0.4720)	...	...
Diabetes	0.0505 Reference	...	...	...
Yes	2.06 (1.00-4.26)	0.7239 (0.3701)	...	...
No	Reference	...	Reference	...

Abbreviation: AIC, Akaike information criterion; AUROC, area under the receiver operating characteristic curve; BIC, Bayesian information criterion; CDAI, Clinical Disease Activity Index; CI, confidence interval; CVD, cardiovascular disease; DMARD, disease-modifying antirheumatic drug; ESR, erythrocyte sedimentation rate; HAQ-DI, Health Assessment Questionnaire Disability Index; OR, odds ratio; SE, standard error.

recommended by TRIPOD. We followed other TRIPOD suggestions in the analysis and reporting of this study (3). The use of RCT data, with well phenotyped patients and outcomes, is a strength of this initial derivation and internal validation study. In secondary analyses, we considered different variable selection techniques to evaluate the robustness of our initial variable selection. Although fewer predictors were selected in the AIC- and BIC-based selection methods compared with the OR-based approach, there was substantial overlap with sex, region, DMARD/methotrexate history, and HAQ-DI score for all models, with similar parameter estimates.

This study has several limitations. Although we followed the TRIPOD recommendations to nonrandomly split the data for internal validation, the current work requires external validation; we plan to pursue this by using real-world data from a large RA registry (18). Data were collected from four separate RCTs. There was some inconsistency in variables between trials, but we used a rigorous process to harmonize the data when possible. Some potentially important predictors, including rheumatoid factor, smoking history, functional assessment of chronic illness therapy (FACIT) score, and erosion score, were not collected in a consistent manner across all four RCTs and were thus not included in this analysis. Only one RCT included patients with early RA (7); although this may make our prediction more robust to patients with all stages of RA, it is possible that different predictors could be important for patients with early versus established disease. The four RCTs had different comparison groups; thus, we focused on the TCZ monotherapy arm only to derive the remission prediction score. Although this prediction score uses readily available clinical and demographic factors, future work could consider whether additional predictors, such as biomarkers or genetic data, can improve prediction. Machine-learning techniques could be considered to better understand complex patterns and interaction among the predictors, particularly in the setting of real-world registry data with many potential predictors. Our analysis focused on predicting disease remission rather than low disease activity. Given the relatively short disease duration (median 1.8 years) we felt that this was the more clinically meaningful outcome in these cohorts (19). In addition, remission was the outcome of interest in several of the trials, and therefore the remission prediction score model mirrors the primary outcomes of the trials. While overall rates of remission were comparable to a recent systematic review of studies using the treat-to-target strategy, the probability of remission was low, and those in the 'moderate' risk of remission category had only a >25% chance of remission (1). Future work could consider both remission and low disease activity, particularly in settings with patients with established disease.

In conclusion, easily accessible clinical variables from RCTs were used to create a remission prediction score that was derived and validated, with AUROCs showing good discrimination. The score correlated well with remission at 24 weeks and was robust to different variable selection methods. If the prediction score is found to be valid in external cohorts, it could be useful in identifying patients who could benefit from TCZ monotherapy. Until

we have promising biomarkers, clinical variables can be used to provide clinicians and patients with valuable information about the likelihood of remission.

## AUTHOR CONTRIBUTIONS

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

**Study conception and design.** Collins, Johansson, Gale, Sontag, Trinh, Losina, Solomon.

**Acquisition of data.** Gale, Trinh, Solomon.

**Analysis and interpretation of data.** Collins, Johansson, Gale, Kim, Shrestha, Sontag, Stratton, Trinh, Xu, Losina, Solomon.

## ROLE OF THE STUDY SPONSOR

This study was funded by Genentech but solely conducted at the Brigham and Women's Hospital. Genentech provided the funding to conduct the study and obtain the study databases but had no role in the analysis. The sponsor was given the opportunity to make nonbinding comments on interpretation of the data and a draft of the manuscript, but the authors retained the right of publication and to determine the final wording.

## REFERENCES

1. Yu C, Jin S, Wang Y, Jiang N, Wu C, Wang Q, et al. Remission rate and predictors of remission in patients with rheumatoid arthritis under treat-to-target strategy in real-world studies: a systematic review and meta-analysis. *Clin Rheumatol* 2019;38:727–38.
2. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;338:b375.
3. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement [published erratum appears in *Ann Intern Med* 2015;162:600]. *Ann Intern Med* 2015;162:55–63.
4. Dougados M, Kissel K, Conaghan PG, Mola EM, Schett G, Gerli R, et al. Clinical, radiographic and immunogenic effects after 1 year of tocilizumab-based treatment strategies in rheumatoid arthritis: the ACT-RAY study. *Ann Rheum Dis* 2014;73:803–9.
5. Gabay C, Emery P, van Vollenhoven R, Dikranian A, Alten R, Pavelka K, et al. Tocilizumab monotherapy versus adalimumab monotherapy for treatment of rheumatoid arthritis (ADACTA): a randomised, double-blind, controlled phase 4 trial [published errata appear in *Lancet* 2013;381:1540 and in *Lancet* 2013;382:1878]. *Lancet* 2013;381:1541–50.
6. Jones G, Sebba A, Gu J, Lowenstein MB, Calvo A, Gomez-Reino JJ, et al. Comparison of tocilizumab monotherapy versus methotrexate monotherapy in patients with moderate to severe rheumatoid arthritis: the AMBITION study. *Ann Rheum Dis* 2010;69:88–96.
7. Burmester GR, Rigby WF, van Vollenhoven RF, Kay J, Rubbert-Roth A, Kelman A, et al. Tocilizumab in early progressive rheumatoid arthritis: FUNCTION, a randomised controlled trial. *Ann Rheum Dis* 2016;75:1081–91.
8. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models. II. External validation, model updating, and impact assessment. *Heart* 2012;98:691–8.
9. Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their



- usefulness and validity in rheumatoid arthritis. *Clin Exp Rheumatol* 2005;23 Suppl 39:S100–8.
10. Aletaha D, Smolen JS. The definition and measurement of disease modification in inflammatory rheumatic diseases. *Rheum Dis Clin North Am* 2006;32:9–44.
  11. Bruce B, Fries JF. The Health Assessment Questionnaire (HAQ). *Clin Exp Rheumatol* 2005;23 Suppl 39:S14–8.
  12. Harrell FE Jr. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. 2nd ed. New York: Springer; 2015.
  13. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med* 2014;33:517–35.
  14. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–38.
  15. Rubin DB. Multiple imputation for nonresponse in surveys. Hoboken (NJ): John Wiley & Sons; 2004.
  16. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res* 2014;23:440–59.
  17. Molenberghs G, Kenward MG. Missing data in clinical studies. Chichester (UK): John Wiley & Sons; 2007.
  18. Kremer JM. The CORRONA database. *Autoimmun Rev* 2006;5:46–54.
  19. Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. *Ann Rheum Dis* 2016;75:3–15.