Universiteit
Leiden
The Netherlands

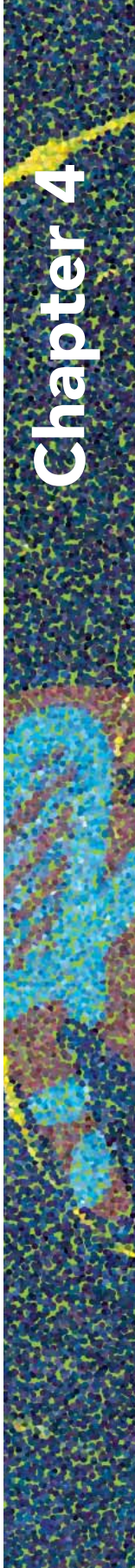
Personalized medicine in rheumatoid arthritis

Eektimmerman, F.

Citation
Eektimmerman, F. (2022, May 11). *Personalized medicine in rheumatoid arthritis*. Retrieved from https://hdl.handle.net/1887/3303689

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3303689

Note: To cite this publication please use the final published version (if applicable).



Validation of a clinical pharmacogenetic model to predict methotrexate non-response in rheumatoid arthritis patients

Frank Eektimmerman
Cornelia F. Allaart
Johanna M.W. Hazes
Moenira B. Madhar
Alfons A. den Broeder
Jaap Fransen
Jesse J. Swen
Henk-Jan Guchelaar

Pharmacogenomics 20(2), 85–93 (2019)

Abstract

Aim: To study the performance of a clinical pharmacogenetic model for the prediction of non-response to methotrexate in rheumatoid arthritis patients treated with combination therapy.

Methods: Prediction model risk scores were calculated and compared with non-response (DAS>2.4). Regression and ROC curve analyses of the prediction model were performed. Also, the sensitivity, specificity, and the positive and negative predictive values (PPV and NPV) were determined.

Results: The ROC AUC was 75% at first and 70% after second evaluation. At the second evaluation, prediction non-response had a sensitivity of 67% (CI: 54–78%), specificity of 69% (CI: 60–77%), PPV of 52% (CI: 45–60%) and NPV of 80% (CI: 73–85%).

Conclusion: The clinical pharmacogenetic model could not predict non-response in RA patients treated with methotrexate combination therapies.

INTRODUCTION

Rheumatoid arthritis (RA) is the most common form of autoimmune arthritis, affecting 0.5–1.0% of the adult population in the Western World.[1] Much of the joint damage that ultimately results in disability begins early in the course of the disease. Thus, early disease recognition, prompt diagnosis with early (intensive) treatment is critical to quickly achieve and maintain control of the inflammation and the underlying disease process. The vast majority of patients with RA start with methotrexate (MTX),[2] with the treatment goal of remission or low disease activity (Disease Activity Score [DAS] ≤2.4). MTX has been used for decades, but a considerable proportion of patients experience an inadequate response. Temporary treatment with corticosteroids has shown to increase early response rates, but after discontinuation of this, MTX can still prove insufficient response. On the other hand, some patients achieve lasting clinical remission on MTX monotherapy. To date, it remains a process of trial and error to choose the best initial treatment for newly diagnosed RA patients, although attempts have been made to identify clinical and genetic risk factors for response to MTX.

A clinical pharmacogenetic model was prior developed to predict non-response (DAS>2.4) of monotherapy MTX in early RA patients.[3] This predictive model combines clinical predictors with genetic variants related to the mechanism of action of MTX (Table 4-1). Based on the summed score in the model, patients are divided into predicted responders (summed score of ≤3.5), intermediate responders (summed score between 3.5 and 6.0) or non-responders (summed score ≥6.0). The predicted non-responders and predicted responders were used to calculate the predictive parameters for the clinical outcome low disease activity (DAS<2.4).

The originally derived prediction model, in patients treated with MTX monotherapy (n=205), showed a sensitivity of 86% (95% confidence interval (CI:) 76–93%) and specificity of 95% (CI: 82–99%) with an AUC of 85% (CI: 80–91%) for the prediction of MTX non-responders. Cross-validation in a small group of 38 early RA patients treated with MTX monotherapy supported the obtained results, although with worse sensitivity and specificity of respectively 70% (CI: 35–93%) and 72% (CI: 47–90%). A subsequent study (n=71) in MTX treated patients with preceding DMARD failure confirmed that the model performs modestly well in predicting MTX non-response, with a sensitivity of 81% (CI: 61–94%), a specificity of 47% (28–66%) and AUC of 77% (CI: not available).[4] Also, a recent replication study, that combined predicted intermediate responders with predicted responders, showed in a large number of MTX monotherapy treated RA patients (n=720) a sensitivity of 50% (CI: 45–55%), a specificity of 75% (CI: 69–80%) and an AUC of 66% (CI: not available).[5]

Since in daily clinical practice RA patients are frequently treated with MTX based (sometimes temporary) combination therapies at an early disease stage – although debate remains

whether these combinations are superior to MTX alone – the performance of the prediction model in these patients is of importance. This study aimed to evaluate the test characteristics of the pharmacogenetic model to predict MTX non-response in RA patients treated with combination therapies.

PATIENTS & METHODS

The recommendations in the TRIPOD statements[6] and the STARD guidelines[7] were used for the describing of the methods and results of the study.

Study participants

Retrospective data of 314 patients were collected from three academic hospitals in the Netherlands: Radboud University Medical Centre, Nijmegen (RUMC), Erasmus Medical Centre, Rotterdam (EMC) and Leiden University Medical Center, Leiden (LUMC). Included patients derived from the tREACH trial[8] (EMC), the IMPROVED study[9] (LUMC), and the early RA inception cohort[10] (Radboud UMC). The period for patient recruitment was between 1989 and 2009, 2007 and 2010, and 2007 and 2011 for respectively the early inception cohort, the IMPROVED study, and the tREACH trial.

Eligible patients were diagnosed with RA, based on the ACR 1987 or EULAR/ACR 2010 classification criteria for RA. Included patients had a treatment duration with MTX and follow-up for at least two study evaluation visits, were 18 years or older, and had not used any DMARD before the start of MTX. Further, DNA samples and clinical data included in the prediction model must be available (complete-case analysis). All patients provided written consent for participation in this study, and the institutional ethics committees approved the study protocol.

IMPROVED patients started their treatment with MTX and tapered prednisone in 7 weeks from 60 mg/day to 7.5 mg/day. At four months, patients with DAS<1.6 received tapered prednisolone to zero in 3 weeks. Patients not in remission (DAS>1.6) at four months were randomized in either 1) MTX + hydroxychloroquine + sulfasalazine and prednisolone, or 2) MTX and adalimumab. The given doses in IMPROVED were: MTX 25 mg/week, hydroxychloroquine 400 mg/day, sulfasalazine 2 g/day, prednisolone 7.5 mg/day, and adalimumab 40 mg/2 weeks.

tREACh patients started their treatment with either 1) MTX + sulfasalazine and hydroxychloroquine with glucocorticosteroids intramuscularly, 2) MTX + sulfasalazine and hydroxychloroquine with oral glucocorticosteroids, or 3) MTX + tapered oral glucocorticosteroids.

Glucocorticosteroids were either given as a single intramuscular dose (either methylprednisolone 120 mg or triamcinolone 80 mg) or an oral tapering scheme of prednisolone in 9 weeks from 15 mg/day to 2.5 mg/day. Patients that not achieved low disease activity (DAS>2.5) at three months switched to MTX with etanercept. The given doses in tREACH were: MTX 25 mg/week, sulfasalazine 2 g/day, hydroxychloroquine 400 mg/day, etanercept 50 mg/week.

RUMC patients were asked in the outpatient clinic to participate in follow-up research (early RA inception cohort). Most RUMC patients started with MTX monotherapy. Also, circa one-fourth of patients were treated with a combination of MTX with either leflunomide or with sulfasalazine. Typically, few patients received oral corticosteroids or biological DMARDs as first-line treatment. However, often intra-articular corticosteroids are used to offer temporary relief, and in a later stage, the combination of MTX and biological DMARDs are sometimes required for adequate disease control.

Outcome and predictors

The primary endpoint was non-response set as not achieving low disease activity (DAS>2.4) at first or second evaluation visit after 3–4 months and 6–8 months after the start of therapy, respectively. The secondary endpoint was EULAR good response criteria, defined as a DAS improvement of >1.2 from baseline and with a DAS of ≤2.4 attained during the first or second evaluation.[11]

Genotyping

Four genetic variants in four genes – *MTHFD1* rs17850560, *AMPD1* rs17602729, *ITPA* rs1127354, *ATIC* rs2372536 – were genotyped in all patients using the TaqMan technique. A TaqMan assay performed quantitative genotyping with a real-time polymerase chain reaction using the LightCycler® 480 (Roche Diagnostics, Mannheim, Germany) following the manufacturer's protocol. The program LightCycler® 480 Endpoint Genotyping analysis software (Roche Diagnostics, Mannheim, Germany) was used to call the genotype results. Each variant was tested for Hardy-Weinberg equilibrium, and a $p<0.05$ was considered as deviance.

Statistical analysis

On baseline, first and second evaluation, the variables between the three cohorts were evaluated. To test differences between the observed responders (attained low disease activity; DAS≤2.4) and non-responders (DAS>2.4), variables at the second evaluation were compared. The variables in the prediction model (Table 4-1) at the first and second evaluation were entered into a logistic regression model and checked if those variables

showed the same effect as in the discovery study. The included variables with the same weighted scores were associated with actual response (low disease activity, DAS>2.4). The associations were reported as betas and OR with the corresponding p-values.

Receiver Operating Characteristic (ROC) curves of the prediction model with the four pharmacogenetic variants (pharmacogenetic model) and without (clinical model) were plotted and the area under the curve (AUC) was calculated.

Based on the summed score in the model, patients are divided into predicted responders (summed score of ≤3.5), intermediate responders (summed score between 3.5 and 6.0) or non-responders (summed score ≥6.0). To assess the performance of the prediction model the sensitivity, specificity, PPV and NPV were calculated. The intermediate responders were ignored in the calculation of the predictive parameters, but were used in the calculation of the AUC of the ROC curve.

Table 4-1. The pharmacogenetic model to predict non-response to methotrexate

Score	Variable
0	Male gender DAS at baseline ≤3.8 RF-negative non-smoker
1	Female gender RF-negative smoker RF-positive non-smoker *MTHFD1 1958* AA genotype *AMPD1 34* CC genotype *ATIC* 347 G-allele carrier
2	RF-positive smoker *ITPA* 94 A-allele carrier
3	DAS at baseline >3.8 and ≤5.1
3.5	DAS at baseline >5.1

A higher summed scores indicate a higher probability of non-response to methotrexate.
Abbreviations: DAS: Disease Activity Score, RF: Rheumatoid Factor.

Time evaluation of the different cohorts

The evaluation time differed intra-and interstudy. For instance, the tREACH study evaluation points were planned quarterly, while the IMPROVED study assessment was planned every four months. To check if those evaluation points influence the DAS and the prediction model, association between visiting times and the DAS were tested using Chi-square test and additionally, visually inspected for a pattern by a scatterplot.

Intended sample size

Based on the development study, the amount of minimal included patients was guided to an expected 40% prevalence of non-responders and a point estimate of 85% sensitivity. As a result, at least 264 patients required to be included to achieve a confidence limit of >75 with 0.95 probability.[12] We planned to include 320 patients to achieve some margins of error and misjudgment of the frequency of non-responders. This sample would allow 80% power to detect differences in sensitivity between responders and non-responders.

All statistical analyses were performed using RStudio version 1.0.136 (RStudio, Boston, MA) and IBM®. SPSS® Statistics 24.0 version (SPSS INC, Chicago, Illinois, USA). P-values lower than 0.05 were considered significant.

RESULTS

Cohort differences

Patient baseline characteristics were similar between the three study cohorts, except for age, smoking, ESR, CRP, VAS, and drug treatment (Table 4-2). The mean DAS at baseline was 3.49 (SD±0.98, range 0.67–6.77), 34 patients (11%) had a DAS below 2.4, patients median age was 54 years (range 18–87 years), the majority was female (69%), rheumatoid factor and anti-citrullinated protein antibodies were positive in 70 and 67% of patients, respectively. At first and second evaluation respectively, the mean MTX dosage of all included patients was 23.3±4.1 and 22.1±5.2 mg/week, and the given weekly MTX dosage was approximately the same between the cohorts. On the contrary, concomitant drug treatment differed between the groups, for example, RUMC patients started their treatment with fewer oral corticosteroids and less concomitant DMARDs than EMC and LUMC, on both evaluation points.

Study outcomes

After the first and second evaluation respectively, 215 (68%) and 223 patients (71%) achieved low disease activity (DAS≤2.4). EULAR good response (DAS<2.4 and DAS improvement >1.2 from baseline) was attained at the first and second visit, in respectively 165 (53%) and 169 (64%) patients. Genotype distribution of all four genetic variants were in Hardy-Weinberg equilibrium (p-value >0.05).

The patient baseline characteristics of the actual responders (DAS<2.4) and non-responders at the second evaluation are shown in Table 4-3. At baseline (start of therapy), significant differences were observed for the use of concomitant DMARDs and corticosteroids, gender,

Table 4-2. Patients characteristics at baseline, first and second evaluation.

	EMC (n=142)	LUMC (n=135)	RUMC (n=37)	Combined (n=314)
At first visit (baseline)				
Age, mean ±SD years	55.2±14.4	52.3±13.6	58.8±14.2	54.4±14.1
Female, n (%)	92 (64.8)	95 (70.4)	29 (78.4)	216 (68.8)
Smoker, n (%)	52 (36.6)	34 (25.2)	6 (16.2)	92 (29.3)
RF-positive, n (%)	102 (71.8)	90 (66.7)	27 (73.0)	219 (69.8)
ACPA positive, n (%)	101 (71.6)[#]	86 (63.7)	16 (61.5)[#]	203 (67.2)[#]
DAS, mean ±SD	3.5±0.9	3.4±1.0	3.6±1.2	3.5±1.0
ESR, mean ±SD	32.0±22.4	33.6±25.9	22.8±19.2	31.6±23.8
CRP, mean ±SD	20.4±27.9	23.1±32.0	19.0±30.4	21.5±30.0
VAS, mean ±SD	51.2±24.2	43.2±24.1	56.1±24.7	47.8±24.5
MTX doses, mean ±SD	25.0±0.0	25.0±0.0	14.24±6.88	23.72±3.89
Concomitant DMARDs, n (%)	138 (97.2)	135 (100.0)	8 (21.6)	281 (89.5)
Concomitant NSAIDs, n (%)	4 (2.8)	93 (68.9)	27 (73.0)	124 (39.5)
Concomitant corticosteroid, n(%)	136 (95.8)	135 (100.0)	3 (8.3)	247 (87.3)
Concomitant biologicals, n (%)	0 (0.0)	23 (17.0)	1 (2.7)	24 (7.6)
At first evaluation (t=3–4 months)				
DAS, mean ±SD	2.0±1.0	1.6±0.9	3.0±1.2	1.9±1.0
ΔDAS from baseline, mean ±SD	1.5±1.1	1.9±1.1	0.6±1.2	1.6±1.2
ESR, mean ±SD	17.3±14.2	12.4±10.1	21.9±22.7	15.7±14.3
CRP, mean ±SD	8.5±12.9	7.7±11.1	13.8±27.3	8.7±14.4
VAS, mean ±SD	31.1±23.0	21.7±20.5	40.6±27.0	28.2±23.3
MTX doses, mean ±SD	24.0±3.2	24.5±2.1	22.2±5.3	23.3±4.1
Concomitant DMARDs, n (%)	138 (97.2)	135 (100.0)	9 (24.3)	282 (89.9)
Concomitant NSAIDs, n (%)	18 (12.7)	65 (48.1)	24 (64.9)	107 (34.1)
Concomitant corticosteroid, n (%)	9 (6.3)	34 (25.2)	4 (10.8)	47 (15.0)
Concomitant biologicals, n (%)	0 (0.0)	23 (17.0)	1 (2.7)	24 (7.6)
At second evaluation (t=6–8 months)				
DAS, mean ±SD	1.9±0.9	1.6±0.8	2.8±0.9	1.9±1.0
ΔDAS from baseline, mean ±SD	1.7±1.1	1.8±1.1	0.8±1.0	1.6±1.1
ESR, mean ±SD	14.9±13.2	13.2±14.7	15.5±12.5	14.2±13.7
CRP, mean ±SD	7.4±11.6	7.7±16.1	6.4±10.0	7.4±13.5
VAS, mean ±SD	28.6±20.7	24.7±20.8	24.8±13.8	21.7±8.7
MTX doses, mean ±SD	22.7±4.4	22.5±5.4	21.6±5.6	22.2±5.2
Concomitant DMARDs, n (%)	141 (99.3)	134 (99.3)	10 (27.0)	285 (90.8)
Concomitant NSAIDs, n (%)	19 (13.4)	26 (19.3)	24 (64.9)	69 (22.0)
Concomitant corticosteroid, n (%)	9 (6.3)	34 (25.2)	3 (8.1)	46 (14.6)
Concomitant biologicals, n (%)	13 (9.2)	35 (25.9)	1 (2.7)	49 (15.6)

[#] Missing data.

Abbreviations: EMC: Erasmus Medical Center, LUMC: Leiden University Medical Center, RUMC: Radboud University Medical Center. RF: Rheumatoid Factor, ACPA: Anti-citrullinated protein antibodies, DAS: Disease activity score, ESR: Erythrocyte sedimentation rate, CRP: C-reactive protein, VAS: visual analogue score, DMARDs: Disease-modifying antirheumatic drugs, NSAIDs: nonsteroidal anti-inflammatory drugs.

Table 4-3. Variables at baseline of the responders and non-responders (according the second evaluation)

Baseline variables	Responders (n=223)	Non-responders (n=91)	p-value
Age, mean ±SD	53.6±14.3	56.3±13.5	0.187
Female gender, n (%)	142 (63.7)	74 (81.3)	$3.42*10^{-3}$ **
RF-positive, n (%)	159 (71.3)	60 (65.9)	0.422
Current smoker, n (%)	64 (28.7)	28 (30.8)	0.819
DAS, mean ±SD	3.3±0.9	3.9±1.0	0.056
ESR, mean ±SD	30.9±23.5	33.5±24.6	0.795
VAS, mean ±SD	44.1±24.0	56.9±23.2	$2.30*10^{-5}$ ***
CRP, mean ±SD	20.4±26.8	24.2±36.8	0.315
MTX dose, mean ±SD	24.4±2.4	21.9±5.8	0.324
Concomitant NSAIDs, n (%)	89 (39.9)	35 (38.5)	0.912
Concomitant DMARDs, n (%)	211 (94.6)	70 (76.9)	$9.17*10^{-6}$ ***
Concomitant corticosteroids, n (%)	208 (93.3)	66 (72.5)	$1.47*10^{-6}$ ***
ITPA 94 A-allele carrier, n (%)	22 (9.8)	16 (17.6)	0.087
ATIC 347 G-allele carrier, n (%)	120 (53.8)	50 (54.9)	0.954
AMPD1 34 CC genotype, n (%)	187 (83.9)	70 (76.9)	0.199
MTHFD1 1985 AA genotype, n (%)	42 (18.8)	21 (22.1)	0.486

Responders were defined as DAS≤2.4 at 6 months.
Abbreviations: RF: Rheumatoid Factor, DAS: Disease Activity Score 28, ESR: Erythrocyte Sedimentation Rate, VAS: Visual Analogue Score, CRP: C-reactive protein, MTX: methotrexate, NSAIDs: non-steroidal anti-inflammatory drugs,
DMARDs: Disease-modifying antirheumatic drugs.
* p<0.05, ** p<0.01, *** p<0.001. # Including missing data.

and the VAS. RUMC patients less often started on combination therapy with corticosteroids, less often had an EULAR response than patients in the other cohorts.

Performance of the pharmacogenetic model

Table 4-4 shows the distribution of the patients into non-responders, intermediate and responders according to the cut-off values of the pharmacogenetic model, divided into patients that achieved response (DAS<2.4) or non-response (DAS≥2.4). At first evaluation, the model for prediction non-response had a sensitivity of 67% (CI: 54–78%), specificity of 70% (CI: 61–78%), PPV of 55% (CI: 47–63%) and NPV of 79% (73–85%). At the second evaluation, the model for prediction non-response had a sensitivity of 67% (CI: 54–78%), specificity of 69% (CI: 60–77%), PPV of 52% (CI: 45–60%) and NPV of 80% (73–85%).

Table 4-4. Pharmacogenetic model at first and second evaluation with observed and predicted MTX response (n=314)

	Predicted response according to the prediction model			
	Non-responders	Intermediate	Responder	
	Score ≥6	Score 3.5–6	Score ≤3.5	Total
Observed response at first evaluation				
Non-responder	46	30	23	99
Responder	38	89	88	215
Total	84	119	111	314
Observed response at second evaluation				
Non-responder	44	25	22	91
Responder	40	94	89	223
Total	84	119	111	314

Non-responders were classified as DAS>2.4 and responders as DAS≤2.4

Regression analysis of the prediction model

Regression analyses of the variables in the prediction are shown in Supplementary Table S4-1. At both time points (first and second evaluation) only the variables female gender and DAS at baseline were significantly associated ($p<0.05$) with MTX response (DAS≤2.4). At first evaluation, RF-positive smoker, *MTHFD1*, and *ATIC* were associated with non-response, while at second evaluation this was only seen for *AMPD1* (OR<1.0). Also, the confidence intervals of most included variables cross 1.0, and this implies that those variables show no difference between the responders and non-responders.

Figure 4-1 plots the ROC curves of the pharmacogenetic and clinical model (without the four genetic variants). The AUC of the ROC curves were 74.6% and 71.5% for the pharmacogenetic model and the clinical model respectively at the first evaluation. The AUC of the second evaluation was lower than that of the first evaluation, with 69.1% and 67.1%, for the pharmacogenetic and clinical model respectively. Taken LUMC and EMC together (without the 32 RUMC patients); the AUC of the ROC were similar to the group consisting of the three cohorts.

Using the EULAR response criteria as an endpoint instead of low disease activity (DAS<2.4) leads to worse performance of the prediction model. The AUC of the ROC curves with EULAR response were AUC of 62.9 and 63.4 (pharmacogenetic), and AUC of 57.7 and 62.3 (clinical model), respectively for the first and second visit.

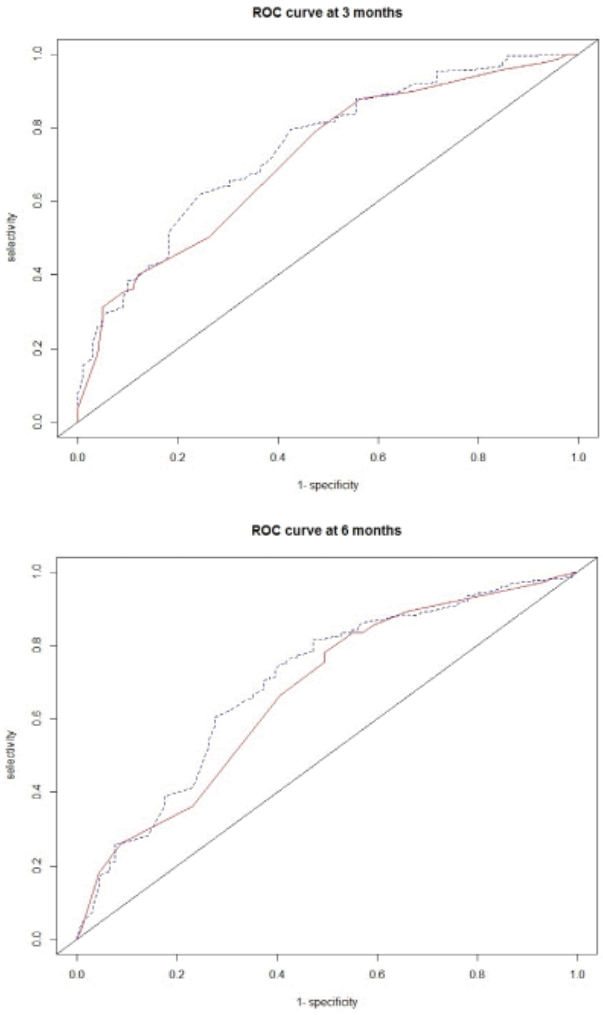


Figure 4-1. The receiver operating characteristic (ROC) curves of the pharmacogenetic model and clinical model.
The ROC curve was expressed as 1-specificity with sensitivity at first evaluation (t = 3 mo) and second evaluation (t = 6 mo). The pharmacogenetic model (blue line) contained the variables: gender, DAS28 at baseline, RF, smoking status and the genetic variants *ATIC* 347G, *IPTA* 94A, *MTHFD1* 1985AA and *AMPD1* 34CC. The clinical model (red line) contained the variables: gender, DAS28 at baseline, RF and smoking status.

Evaluation time differences

Analysis of the time after the start of MTX (evaluation time) to the DAS showed no pattern in the scatterplots in both evaluation points (Supplementary Figure S4-1). Also, no statistical difference was found between the DAS and the time visits: p-values were 0.08 and 0.56 at the first and second evaluation, respectively.

DISCUSSION

This study shows that the pharmacogenetic model, originally derived in early RA patients treated with MTX monotherapy, could not predict non-response of RA patients treated with MTX based combination therapies. Although the AUCs of the ROC curves were weak to modest (approximately 70%), the PPV, specificity and sensitivity were inadequate to predict non-response. For instance, the PPV, the complement of the false discovery rate, showed that approximately 50% of the actual non-responders (47 out of 91) were predicted as responders. Interestingly, while in MTX monotherapy the model had a sensitivity of 86% in predicting non-response, in patients treated with MTX combination therapy the PPV had a decrease to 70%. Therefore, the prediction model is not clinically applicable to predict non-response in patients treated with MTX combination therapy.

There are several possible explanations for the underperformance of the prediction model. One reason is that the included pharmacogenetic variants showed a minimal additive value in the prediction model with an AUC increase of 2.0 and 3.1% of the ROC curves, for the first and second evaluation visit respectively. The reason may be that the pharmacogenetic variants are related to the mechanism of action of MTX and adding other DMARDs as is the case in our replication cohort compared to the discovery cohort could dilute the predictive effect. While the weak predictive value of the pharmacogenetic variants was confirmed in the replication studies with MTX monotherapy treated RA patients, the variants showed a better prediction and therefore makes it a necessary component in the pharmacogenetic model.

Another potential explanation for the underperformance of the prediction model may be the baseline DAS in the prediction model. Patients in the development cohort had a high mean baseline DAS of 4.4, and as a consequence, a DAS of 3.8 was a modifier for response in the prediction model. In our cohort, however, the baseline DAS was 3.5 and showed a small contribution in the prediction model. Because the low baseline DAS, and because the use of combination therapies, the majority of the patients achieved low disease activity on both evaluation visits (circa 70%). Yet, using the EULAR response criteria, that takes the baseline DAS into account and showed ~50% responders, still results in poor prediction and is not applicable in the clinical setting. This study showed that the predictive value of the model exists mainly on the clinical values: gender, rheumatoid factor positivity, and smoking status. The use of different RA classification criteria (1987 or 2010 criteria) could also play a role in the underperformance of the prediction model. For instance, the 2010 criteria were broader, and patients could be indicated with RA in an earlier disease stage. However, no difference was found between the classification criteria of RA in baseline DAS scores.

The frequency of the predicted intermediate responders is an important indicator of the feasibility of the prediction model and could limit the clinical usefulness, as it increases the number needed to diagnose. In our study, a large group of patients (approximately 40%) were predicted intermediate responders, and for this group, no drug advice (MTX or alternative drug treatment) could be given. Therefore, it may be better to use a single cutoff value in the prediction model to get a clear distinguishment between two groups: predicted responders and predicted non-responders. For instance, this was performed in the large replication study, where responders and intermediate responders were combined into one group.

Our study has a few strengths. First, with 314 patients the study is one of the largest MTX pharmacogenetics studies published so far. Also, the estimations of the diagnostic parameters were precise, with small CIs around them. Second, patients were treated with mainly combination therapies of MTX with either another DMARD or tapered corticosteroids and thus represents treatment according to daily clinical practice. Third, the use of the TRIPOD and STARD reporting criteria ensures a full and transparent way of reporting.

The prediction of efficacy in RA seems challenging with still today no clear indicators for routine daily practice. Multiple studies tried to find predictors for the response to MTX or the discontinuation of MTX in RA patients,[13–16] but those studies lack or failed replication. Subsequently, a review on biological DMARDs showed 65 potential (bio)markers, but as well no validation studies were performed.[17] Probably, even a reasonably accurate prediction of response will not have a substantial impact on the treatment outcome. One explanation was that hospitals increasingly used the treat-to-target approach (with the DAS steered therapy) and the use of temporary corticosteroids treatment. This results in the finding of current trials that >80% of the patients are in a state of remission after one year of drug treatment. Also, the prediction models include variables that also predict to some extent non-response for alternatives for MTX. For example, sex, RF as acute phase reactants have weak predictive effects also for other (b)DMARDs. Therefore, overall, there seems little room to improve the treat-to-target and trial and error RA care vastly.

In summary, a prediction model developed to predict response to MTX monotherapy was tested in three other cohorts starting with MTX combination therapy and performed poorly. Based on patients with the treat-to-target approach, prediction models offer no added value for daily clinical practice.

Acknowledgement

The authors would like to thank the following persons for their assistance with the collection of the data or receiving the DNA: Thea van Gaalen, Els van den Enden, Ron Buijs, and Marieke Coenen. We would also like to extend our thanks to Renée Baak-Pablo, Daniëlle Klootwijk, and Rowena Schaap, for their help in genotyping.

REFERENCES

1. Scott DL, Wolfe F, Huizinga TWJ. Rheumatoid arthritis. Lancet 376, 1094–1108 (2010).
2. Singh JA, Saag KG, Bridges SL *et al*. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Care Res. (Hoboken). 68, 1–25 (2016).
3. Wessels JAM, van der Kooij SM, le Cessie S *et al*. A clinical pharmacogenetic model to predict the efficacy of methotrexate monotherapy in recent-onset rheumatoid arthritis. Arthritis Rheum. 56, 1765–1775 (2007).
4. Fransen J, Kooloos WM, Wessels JAM *et al*. Clinical pharmacogenetic model to predict response of MTX monotherapy in patients with established rheumatoid arthritis after DMARD failure. Pharmacogenomics 13, 1087–1094 (2012).
5. López-Rodríguez R, Ferreiro-Iglesias A, Lima A *et al*. Evaluation of a clinical pharmacogenetics model to predict methotrexate response in patients with rheumatoid arthritis. Pharmacogenomics J. 18, 539–545 (2018).
6. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 350, g7594 (2015).
7. Cohen JF, Korevaar DA, Altman DG *et al*. STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. BMJ Open 6, 1–17 (2016).
8. de Jong PH, Hazes JM, Kan HK *et al*. Randomised comparison of initial triple DMARD therapy with methotrexate monotherapy in combination with low-dose glucocorticoid bridging therapy; 1-year data of the tREACH trial. Ann. Rheum. Dis. 73, 1331–9 (2014).
9. Heimans L, Wevers-de Boer KVC, Visser K *et al*. A two-step treatment strategy trial in patients with early arthritis aimed at achieving remission: the IMPROVED study. Ann. Rheum. Dis. 1356–1361 (2013). doi:10.1136/annrheumdis-2013-203243
10. Kievit W, Fransen J, de Waal Malefijt MC, den Broeder AA, van Riel PLCM. Treatment changes and improved outcomes in RA: An overview of a large inception cohort from 1989 to 2009. Rheumatol. (United Kingdom) 52, 1500–1508 (2013).
11. van Gestel AM, Prevoo ML, van 't Hof MA *et al*. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis: Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. Arthritis Rheum. 39, 34–40 (1996).
12. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. J. Clin. Epidemiol. 58, 859–862 (2005).
13. Jenko B, Tomsic M, Jekic B *et al*. Clinical Pharmacogenetic Models of Treatment Response to Methotrexate Monotherapy in Slovenian and Serbian Rheumatoid Arthritis Patients: Differences in Patient's Management May Preclude Generalization of the Models. Front. Pharmacol. 9, 20 (2018).
14. Jenko B, Lusa L, Tomsic M, Praprotnik S, Dolzan V. Clinical-pharmacogenetic predictive models for MTX discontinuation due to adverse events in rheumatoid arthritis. Pharmacogenomics J. 1–7 (2016). doi:10.1038/tpj.2016.36

15. Hider SL, Silman AJ, Thomson W *et al*. Can clinical factors at presentation be used to predict outcome of treatment with methotrexate in patients with early inflammatory polyarthritis? Ann. Rheum. Dis. 68, 57–62 (2009).

16. de Rotte MCFJ, de Jong PHP, Pluijm SMF *et al*. Association of Low Baseline Levels of Erythrocyte Folate With Treatment Nonresponse at Three Months in Rheumatoid Arthritis Patients Receiving Methotrexate. Arthritis Rheum. 65, 2803–2813 (2013).

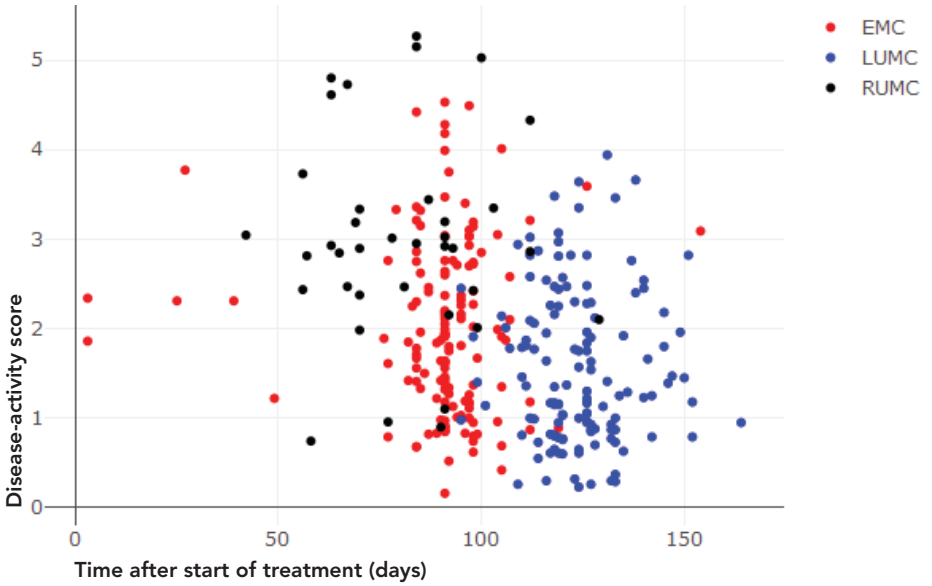
17. Cuppen BVJ, Welsing PMJ, Sprengers JJ *et al*. Personalized biological treatment for rheumatoid arthritis: a systematic review with a focus on clinical applicability. Rheumatology 55, 826–839 (2016).

Supplementary Table S4-1. Regression coefficients and odds ratios of the logistic regression models to predict MTX response

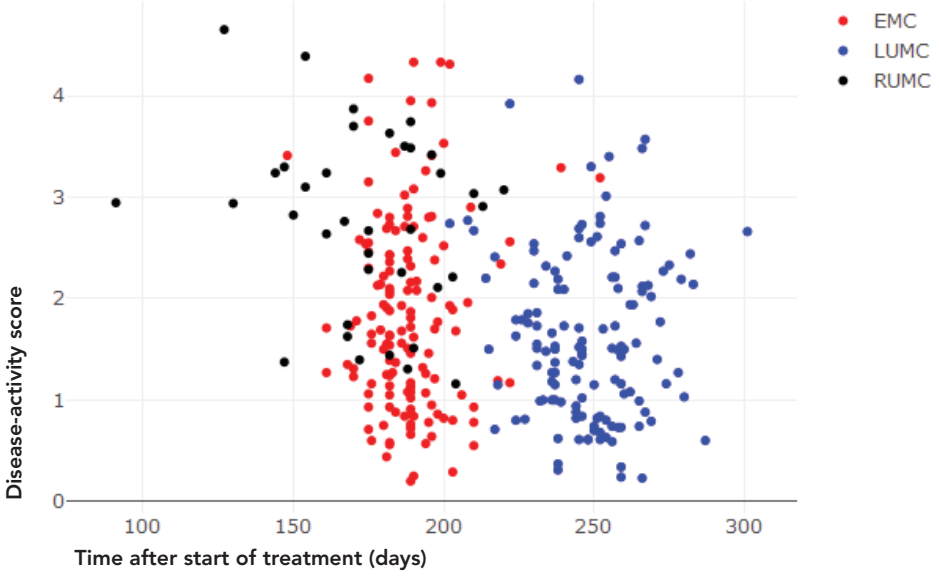
	At first evaluation visit (3–4 months)		
Variable	β	OR [95% CI]	p-value
(intercept)	2.53	12.54 [4.59–37.08]	$1.92 * 10^{-06}$ ***
Female gender	-1.72	0.18 [0.09–0.35]	$1.55 * 10^{-06}$ ***
DAS at baseline	-0.43	0.65 [0.55–0.78]	$2.01 * 10^{-06}$ ***
RF positive smoker	0.01	1.01 [0.68–1.50]	0.960
MTHFD1 1958 AA genotype	0.09	1.09 [0.57–2.14]	0.795
AMPD1 34 CC genotype	-0.03	0.97 [0.48–1.90]	0.929
ITPA 954 A-allele carrier	-0.33	0.72 [0.49–1.05]	$8.39 * 10^{-2}$ *
ATIC 347 G-allele carrier	0.21	1.23 [0.73–2.08]	0.434
	At second evaluation visit (6–8 months)		
Variable	β	OR [95% CI]	p-value
(intercept)	1.94	6.97 [2.73–18.89]	$7.92 * 10^{-05}$ ***
Female gender	-0.94	0.39 [0.20–0.71]	$2.93 * 10^{-3}$ **
DAS at baseline	-0.40	0.67 [0.57–0.80]	$6.02 * 10^{-6}$ ***
RF positive smoker	-0.04	0.96 [0.65–1.43]	0.857
MTHFD1 1958 AA genotype	-0.16	0.86 [0.45–1.65]	0.637
AMPD1 34 CC genotype	0.38	1.46 [0.75–2.79]	0.256
ITPA 954 A-allele carrier	-0.34	0.71 [0.49–1.04]	$7.56 * 10^{-2}$ *
ATIC 347 G-allele carrier	-0.09	0.92 [0.54–1.54]	0.742

Abbreviations: DAS: disease activity score, RF: rheumatoid factor, β: regression coefficient, OR: Odds ratio, CI: confidence interval.
* p<0.10, ** p<0.01, *** p<0.001.

A. First evaluation



B. Second evaluation



Supplementary Figure S4-1. Scatterplot of the time visits of the first and second evaluation (days) versus the disease-activity score.