



Validation of a Prognostic Multivariable Prediction Model for Insufficient Clinical Response to Methotrexate in Early Rheumatoid Arthritis and Its Clinical Application in Evidencio

Helen R. Gosselt · Maxime M. A. Verhoeven · Maurits C. F. J. de Rotte · Saskia M. F. Pluijm · Ittai B. Muller · Gerrit Jansen · Janneke Tekstra · Maja Bulatović-Ćalasan · Sandra G. Heil · Floris P. J. G. Lafeber · Johanna M. W. Hazes · Robert de Jonge

Received: July 14, 2020 / Published online: September 14, 2020
© The Author(s) 2020

ABSTRACT

Introduction: Methotrexate (MTX) constitutes the first-line therapy in rheumatoid arthritis (RA), yet approximately 30% of the patients do not benefit from MTX. Recently, we reported a prognostic multivariable prediction model for insufficient clinical response to MTX at 3

Johanna M. W. Hazes and Robert de Jonge contributed equally to this manuscript as senior authors.

Digital Features To view digital features for this article go to: <https://doi.org/10.6084/m9.figshare.12800714>.

Electronic Supplementary Material The online version of this article (<https://doi.org/10.1007/s40744-020-00230-7>) contains supplementary material, which is available to authorized users.

H. R. Gosselt (✉) · M. C. F. J. de Rotte · I. B. Muller · R. de Jonge
Amsterdam Gastroenterology and Metabolism, Department of Clinical Chemistry, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
e-mail: h.gosselt@amsterdamumc.nl

H. R. Gosselt · S. G. Heil
Department of Clinical Chemistry, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

M. M. A. Verhoeven · J. Tekstra · F. P. J. G. Lafeber
Department of Rheumatology and Clinical Immunology, UMC Utrecht, Utrecht, The Netherlands

months of treatment in the treatment in the Rotterdam Early Arthritis Cohort (tREACH), including baseline predictors: Disease activity score 28 (DAS28), Health Assessment Questionnaire (HAQ), erythrocyte folate, single-nucleotide polymorphisms (SNPs; *ABCB1*, *ABCC3*), smoking, and BMI. The purpose of the current study was (1) to externally validate the model and (2) to enhance the model's clinical applicability.

Methods: Erythrocyte folate and SNPs were assessed in 91 early disease-modifying anti-rheumatic drug (DMARD)-naïve RA patients starting MTX in the external validation cohort (U-Act-Early). Insufficient response (DAS28 > 3.2) was determined after 3 months and non-response after 6 months of therapy. The previously developed prediction model was

M. C. F. J. de Rotte
Amsterdam Gastroenterology and Metabolism, Department of Clinical Chemistry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

S. M. F. Pluijm
Department of Paediatric Oncology, Prinses Máxima Centre for Paediatric Oncology, Utrecht, The Netherlands

G. Jansen
Amsterdam Rheumatology and Immunology Center, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

considered successfully validated in the U-Act-Early (validation cohort) if the area under the curve (AUC) of the receiver operating characteristic (ROC) was not significantly lower than in the tREACH (derivation cohort).

Results: The AUCs in U-Act-Early at three and 6 months were 0.75 (95% CI 0.64–0.85) and 0.71 (95% CI 0.60–0.82) respectively, similar to the tREACH. Baseline DAS28 > 5.1 and HAQ > 0.6 were the strongest predictors. The model was simplified by excluding the SNPs, while still classifying 73% correctly. Furthermore, interaction terms between BMI and HAQ and BMI and erythrocyte folate significantly improved the model increasing correct classification to 75%. Results were successfully implemented in Evidencio online platform assisting clinicians in shared decision-making to intensify treatment when appropriate.

Conclusions: We successfully externally validated our recently reported prediction model for MTX non-response and enhanced its clinical application thus enabling its evaluation in a clinical trial.

Trial Registration: The U-Act-Early is registered at ClinicalTrials.gov. number: NCT01034137. tREACH is registered retrospectively at ISRCTN registry, number: ISRCTN26791028 at 23 August 2007.

Keywords: Arthritis; Health care; Methotrexate; Outcome assessment; Rheumatoid; Therapeutics

M. Bulatović-Ćalasan
Department of Internal Medicine, UMC Utrecht,
Utrecht, The Netherlands

J. M. W. Hazes
Department of Rheumatology, Erasmus MC,
University Medical Center Rotterdam, Rotterdam,
The Netherlands

Key Summary Points

Why carry out this study?

Previous multiple prediction models for non-response to methotrexate (MTX) have been proposed, which all resulted in an area under the curve (AUC) between 0.65 and 0.80, but not all models have been validated.

The purpose of the current study was to externally validate a previously developed prediction model for insufficient response to MTX and to enhance the model's applicability in clinical practice.

What has been learned from the study?

The prediction model was externally validated with an AUC of 0.75 (95% CI 0.64–0.85), enhanced for clinical applicability (AUC = 0.75, 95% CI 0.69–0.81) and successfully integrated in an online tool “Evidencio”, which can assist clinicians and patients in shared decision-making.

Patients with high risk scores for insufficient response to MTX according to our model integrated in Evidencio can immediately intensify MTX treatment with biologic disease-modifying anti-rheumatoid drugs/conventional synthetic disease-modifying anti-rheumatoid drugs (bDMARDs/csDMARDs) as proposed in the America College of Rheumatology/ European League Against Rheumatism (ACR/EULAR) 2019 updated guidelines for RA treatment [1], while sufficient responders are spared expensive step-up treatment.

Disease activity in these patients can be tightly controlled during the window of opportunity, resulting in better long-term responses and avoiding unnecessary adverse events of MTX.

INTRODUCTION

Methotrexate (MTX) is the first-line therapy in rheumatoid arthritis (RA) [1]. Although efficacious in a large proportion of patients, MTX is poorly effective in approximately 30% of early RA patients [2, 3]. Patients on MTX who do not show improvement at 3 months (insufficient responders) or do not reach the treatment target of low disease activity/remission at 6 months (non-responders) are switched to biologic disease-modifying anti-rheumatoid drug (bDMARD) therapies or novel targeted synthetic DMARD (tsDMARD) therapies, including, e.g., Janus kinase (JAK) inhibitors [1, 4, 5], with or without concomitant MTX treatment. To ensure that only patients unresponsive to MTX receive early (additional) treatment with b/tsDMARDs and those responsive to MTX are spared costly biologics or synthetic drugs, we and others have constructed models to predict MTX (non)-response [6–9]. Our prognostic multivariable prediction model for the prediction of insufficient response, defined as: disease activity score 28 (DAS28 > 3.2) at 3 months of MTX therapy, was constructed in the treatment in the Rotterdam Early Arthritis Cohort (tREACH) and included clinical predictors (DAS28 and Health Assessment Questionnaire [HAQ]), life-style predictors (smoking and BMI) and laboratory parameters involved in MTX metabolism (erythrocyte folate and single-nucleotide polymorphisms: SNPs) [9]. This model classified 80% of patients correctly (area under the curve [AUC] of the receiver operating characteristic [ROC]: 0.80 [95% CI 0.73–0.86]) and was externally validated in the MTX-Rotterdam cohort showing a similar prognostic performance (AUC 0.80 [95% CI 0.69–0.91]) even though BMI and smoking predictors were absent from this validation cohort [9]. The aim of the current study was to validate the complete prediction model, including BMI and smoking status predictors, in an external early RA cohort (U-Act-Early) from a different geographic region and to enhance the model's applicability in clinical practice [10].

METHODS

The methodology of this study followed transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines [11].

Patients

The external validation cohort consisted of 91 patients from the U-Act-Early cohort, a multicenter, double-blind, placebo-controlled strategy trial, registered at ClinicalTrials.gov (number: NCT01034137) [10]. DMARD- and glucocorticoid (GC)-naïve early RA patients were eligible for inclusion once classified as RA patients according to the 1987 America College of Rheumatology (ACR) [12] ($n = 7$, 8%) or the 2010 ACR/EULAR classification criteria ($n = 84$, 92%) [13], and had a disease duration < 12 months and active disease at baseline (disease activity score 28; DAS28 ≥ 2.6). Patients were randomly assigned to a treatment strategy with tocilizumab (TCZ) + placebo, MTX + placebo or their combination (TCZ + MTX) and treated to the target of sustained remission (i.e., a DAS28 < 2.6 and swollen joint count of ≤ 4 joints of the 28 joints assessed, during ≥ 24 weeks). All 91 patients included were derived from the initial MTX + placebo strategy arm. The starting MTX dose was 10 mg/week orally and increased stepwise 5 mg every 4 weeks up to 30 mg/week until remission or the maximum tolerable dose. During the trial, GC use was not permitted. The tREACH ($n = 285$) cohort was described earlier [9]. Importantly, in the tREACH, the optimal MTX dose of 25 mg/week was reached within 3 weeks (combined with other conventional synthetic (cs) DMARDs and/or GCs) and therapy was targeted to low disease activity (DAS28 ≤ 3.2) at 3 months. If this failed, step-up treatment with additional csDMARDs (sulfasalazine and/or hydroxychloroquine) or bDMARDs (i.e., TNF- α inhibitor) was initiated. In both cohorts, folic acid (10 mg/week) was prescribed during MTX treatment. This study was approved by the medical ethics committee of the University Medical Center Utrecht (ML22497) and the

medical ethics committee of Erasmus Medical Center (MEC-2006-252). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Written informed consent was obtained for all patients.

Outcome and Clinical Predictors

The primary outcome was insufficient MTX response after 3 months of treatment start, defined as DAS28 > 3.2, where DAS28 was based on the erythrocyte sedimentation rate (DAS28-ESR). All predictors were dichotomized prior to analyses. Cut-off values were defined previously in the tREACH cohort as: DAS28 > 5.1, HAQ > 0.6, erythrocyte folate < 750 nmol/l, current smoking, BMI > 25 kg/m², *ABCB1* rs1045642 (GG/GA vs. AA) genotype, and *ABCC3* rs4793665 (TC/CC vs. TT) genotype [9]. As erythrocyte-folate levels were slightly higher in U-Act-Early, new cut-off points were examined for erythrocyte-folate (deciles) and BMI (> 30 kg/m²) in U-Act-Early and tested for improvement of the model. The secondary outcome measure was non-response to MTX after 6 months of treatment, defined as DAS28-ESR > 3.2.

Erythrocyte Folate and Genetic Variants

In U-Act-Early, erythrocyte folate and genetic variants included in the original prediction model [9] were determined from EDTA whole blood samples stored at – 80 °C, as described elsewhere [14, 15]. DNA was obtained from whole blood using a MagNAPure Compact (Roche Life Science, Almere, The Netherlands) and genotypes were determined for *ABCB1* rs1045642 and *ABCC3* rs4793665 using real-time PCR with Taqman, as described previously [15]. Samples were tested for deviation from the Hardy–Weinberg equilibrium (HWE). SNPs were determined in the same lab and according to the same protocols as the tREACH study [9]. Importantly, there were methodological

differences in the measurement of folate between the tREACH and U-Act-Early cohorts. The Elecsys® Folate III assay (Ref 7027290190; Roche Diagnostics) has been re-standardized since 2017 in accordance with the WHO International Standard NIBSC Code 03/178. This resulted in 10% lower erythrocyte folate levels (U-Act-Early) compared to those quantified using the previous assay (tREACH). Furthermore, serum folate levels, required for folate correction in whole blood, were not available in U-Act-Early. As serum folate levels take up only a small part of the total folate concentrations, whole blood folate levels were corrected for the average serum folate concentration in the tREACH (25 nmol/l).

Statistical Analysis

Clinical and laboratory parameters for the tREACH (derivation) and U-Act-Early (validation) cohorts were compared. The difference in DAS28 at 3 months (compared to baseline) was expressed as a mean with standard deviation (\pm SD) and assessed using a paired-sample *t* test. Differences between cohort variables were tested using an independent two-group *t* test, if the assumptions of normal distribution (visual inspection) and equal variances (tested using the Levene's test) were met. If these assumptions were not met, non-parametric Mann–Whitney *U* test was performed. Differences in proportions were tested using a two-sample proportion test. Due to missing informed consent at the start of this validation study, eight subjects from the tREACH (derivation) cohort were excluded from analyses in this study. This is why the model as described previously was first re-analyzed on the tREACH data excluding these eight subjects, resulting in negligible differences in effect sizes compared to previous study [9]. Next, to validate the prediction model in the external validation cohort (U-Act-Early), the predictors, DAS28 > 5.1, HAQ > 0.6, erythrocyte folate < 750 nmol/l, current smoking, BMI > 25 kg/m², *ABCB1* rs1045642 genotype and *ABCC3* rs4793665 genotype were entered into a multivariable logistic regression and the probability for

insufficient response was calculated using the pROC package in R according to the following formula:-

, where β_0 represents the constant, β , β_2 , and β_n represent the regression coefficients for each of the predictors x_1 , x_2 , x_n . Subsequently, an ROC curve with AUC was constructed using the predicted probabilities and compared with results in the tREACH. The previously developed prediction model in tREACH (derivation) was considered successfully validated in the U-Act-Early (validation) if the area under the curve (AUC) of the receiver operating characteristic (ROC) was not significantly lower than in the tREACH. Goodness of fit between the predicted probabilities and observed values was tested using the Hosmer–Lemeshow test, where $p > 0.05$ indicated that a model fit the data well. All analyses were performed in R studio (Version: 3.5.3, “2019-03-11”). P values < 0.05 were considered statistically significant. Subjects with missing data ($N = 104$) were excluded from analyses (complete-case analysis).

To enhance the model’s clinical applicability and thus facilitate its clinical implementation, we applied the prediction model on the combined dataset of tREACH and U-Act-Early ($N = 264$) using the step-up approach. Therefore, the model could be simplified using fewer predictors and possible two-way interactions could be examined in a combined cohort with more power. Statistically significant interactions ($p < 0.05$) were added to the model. To simplify the model, we assessed individual contribution of variables to the predictive power of the model by sequential addition of predictors. Model fits were compared using the likelihood ratio test. Probability for insufficient response was calculated for each patient as well as the corresponding specificity, positive predictive value (PPV), sensitivity, and negative predictive value (NPV) were determined, using the “pROC” package in R. The final prediction model was uploaded onto the online platform ‘Evidencio’ providing a tool for clinicians to decide whether to start MTX combination therapy.

RESULTS

Cohort Comparisons

In U-Act-Early (validation cohort), mean DAS28 decreased from $5.0 (\pm 1.1)$ to $3.6 (\pm 1.6)$ during the first 3 months ($p < 0.001$). Mean DAS28 in tREACH (derivation cohort) was $5.0 (\pm 1.1)$, which decreased to a mean DAS28 of $3.1 (\pm 1.2)$, $p < 0.001$. Baseline DAS28 in U-Act-Early was comparable to that of tREACH ($p = 0.613$; Table 1). In U-Act-Early, 58 patients (64%) were categorized as insufficient MTX responders (i.e., $\text{DAS28} > 3.2$ at 3 months) compared to 114 (43%) in tREACH ($p = 0.006$). In U-Act-Early, 39 (44%) patients were classified as MTX non-responders (i.e., $\text{DAS28} > 3.2$ at 6 months), which was not significantly different from 38% in tREACH (Table 1). Additionally, U-Act-Early consisted of significantly more rheumatoid factor positive patients (81%) compared to tREACH (65%) ($p = 0.007$; Table 1), whereas no significant differences were found for anti-citrullinated protein antibody (ACPA) positivity ($p = 0.214$; Table 1). Despite the lower folate levels due to (international) re-standardization of the method, the erythrocyte-folate levels were still significantly higher in U-Act-Early compared to tREACH ($p = 0.006$; Table 1) and genotype GG/GA for *ABCB1* was significantly more frequent in tREACH ($p = 0.016$; Table 1), while genotypes for *ABCC3* were similar between cohorts. Importantly, besides MTX, co-medication was prescribed in tREACH (derivation) but not in U-Act-Early (validation; Table 1).

Validation of Prediction Model at 3 Months

The model combining all predictors accomplished an AUC of 0.75 (95% CI 0.64–0.85) (Fig. 1) in U-Act-Early (validation), which means that 75% of the patients could be classified correctly. The strongest predictor for insufficient response to MTX in U-Act-Early was baseline $\text{DAS28} > 5.1$ ($p = 0.008$; Table 2). Odds ratios (ORs) for the predictors DAS28, HAQ, erythrocyte folate, BMI and smoking in U-Act-Early (validation) were in the same direction as

Table 1 Descriptives of the derivation (tREACH) and external validation (U-Act-Early) cohorts

| Clinical parameters | tREACH Derivation cohort N = 277 | U-Act-Early Validation cohort N = 91 | P value |
|---|--|--|------------|
| DAS28 > 3.2 at 3 months | 43% | 64% | 0.006** |
| DAS28 > 3.2 at 6 months | 38% | 44% | 0.417 |
| Gender, male | 30% | 36% | 0.300 |
| Age, mean ± SD | 54 ± 14 | 53 ± 13 | 0.498 |
| Baseline DAS28, mean ± SD | 5.0 ± 1.1 | 5.0 ± 1.1 | 0.613 |
| HAQ > 0.6 | 76% | 70% | 0.330 |
| Laboratory parameters | | | |
| Erythrocyte folate, median (IQR) [#] | 862 (665–1163) | 1020 (795–1221) | 0.006** |
| <i>ABCB1</i> GG/GA | 73% | 58% | 0.016* |
| <i>ABCC3</i> TC/CC | 66% | 67% | 0.909 |
| Rheumatoid factor positive | 65% | 81% | 0.007** |
| ACPA positive | 71% | 79% | 0.214 |
| Life-style parameters | | | |
| BMI, median (IQR) | 25 (23–29) | 25 (23–29) | 0.950 |
| Current smokers, N (%) | 84 (33) | 28 (31) | 0.820 |
| Co-medication | | | |
| Other DMARDs | 56% | 0% | < 0.001*** |
| Oral corticosteroids | 58% | 0% | < 0.001*** |
| Parental corticosteroids | 28% | 0% | < 0.001*** |
| Subcutaneous | 0% | 0% | |

Missing values tREACH: DAS28 at 3 months $N = 13$, DAS28 at 6 months, $N = 28$, erythrocyte folate $N = 78$, rheumatoid factor $N = 35$, BMI $N = 3$, smoking status $N = 21$, HAQ $N = 18$, *ABCB1* $N = 21$, *ABCC3* $N = 20$. Missing values U-Act-Early: DAS28 at 6 months, $N = 2$, rheumatoid factor $N = 1$, ACPA status $N = 1$. Percentages shown are of valid data points. * $P < 0.05$ was considered significant, ** $P < 0.01$, *** $P < 0.001$

[#] Expressed in nmol/l

in the tREACH (derivation; Table 2). The OR of the *ABCC3* SNP was however in the opposite direction in U-Act-Early (OR = 0.6, 95% CI 0.23–1.79) compared to tREACH (OR = 3.1, 95% CI 1.39–6.94).

Similar results were found for the prediction of non-response at 6 months. An ROC curve was constructed with an AUC of 0.71 (95% CI 0.60–0.82; Supplementary Figure S1) in U-Act-Early (validation), which is comparable to the

predictive value of the tREACH model (derivation) at 6 months (AUC 0.75, 95% CI 0.67–0.83; Supplementary Figure S1).

Enhancement of the Model's Clinical Applicability

Next, the U-Act-Early and tREACH cohort were combined to increase power and enhance the

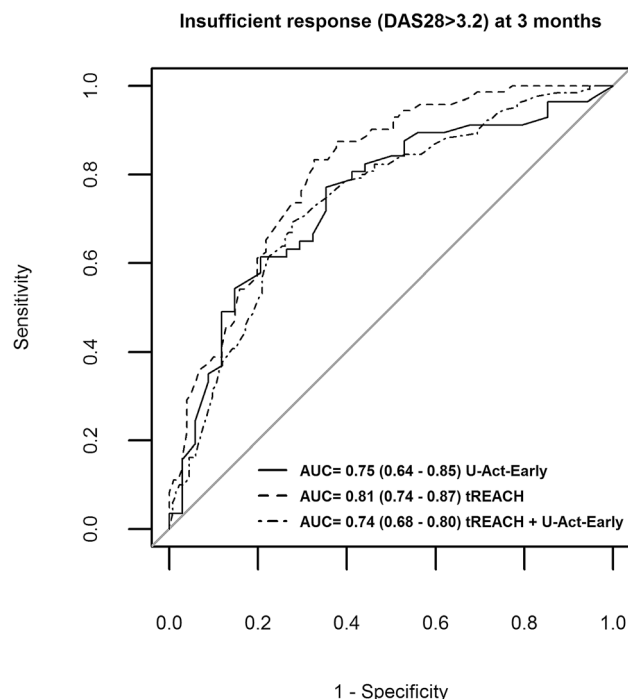


Fig. 1 ROC curve for the prediction of insufficient response (DAS28 > 3.2) to MTX after 3 months of treatment. Area under the curve (AUC) is reported as follows: AUC (95% confidence interval). Predictors were:

baseline DAS28 > 5.1, baseline HAQ > 0.6, *ABCB1* genotype, *ABCC3* genotype, baseline erythrocyte folate, BMI > 25 kg/m² and current smoking

Table 2 Validation of multivariable logistic regression models for insufficient response to MTX (DAS28 > 3.2) at 3 months of treatment in an external validation cohort (U-Act-Early)

| | tREACH derivation cohort N = 173 | U-Act-Early validation cohort N = 91 |
|----------------------------|-------------------------------------|---|
| Predictors | OR (95% CI) | OR (95% CI) |
| DAS28 > 5.1 | 3.7 (1.62–8.38)** | 4.1 (1.44–11.82)** |
| HAQ > 0.6 | 2.8 (1.15–7.00)* | 2.1 (0.67–6.35) |
| <i>ABCB1</i> GG/GA | 2.4 (1.06–5.23)* | 1.0 (0.37–2.75) |
| <i>ABCC3</i> TC/CC | 3.1 (1.39–6.94)** | 0.6 (0.23–1.79) |
| Folate < 750 nmol/l | 2.1 (0.97–4.40) | 3.4 (0.88–12.79) |
| Smoker | 4.2 (1.91–9.42)** | 1.3 (0.44–4.00) |
| BMI > 25 kg/m ² | 3.3 (1.52–7.21)** | 1.6 (0.62–4.23) |
| AUC (95% CI) | 0.81 (0.74–0.87) | 0.75 (0.64–0.85) |

The left column presents data from the derivation cohort (tREACH) and the right of the external validation cohort (U-Act-Early). OR odds ratio, CI confidence interval. Predictors that contributed significantly to the model are indicated with an asterisk, where *P < 0.05 and **P < 0.01

Table 3 Logistic model building in combined datasets: U-Act-Early + tREACH

| Model | Predictors | Log likelihood | Chi-square | <i>P</i> | AUC (95% CI) |
|-------|--|----------------|------------|----------|---------------------|
| 1 | DAS28 + HAQ | − 168.68 | | | 0.67 (0.61–0.74) |
| 2 | DAS28 + HAQ + smoking | − 165.52 | 6.32 | 0.01* | 0.70 (0.64–0.76) |
| 3 | DAS28 + HAQ + smoking + BMI | − 162.98 | 5.08 | 0.02* | 0.72 (0.66–0.78) |
| 4 | DAS28 + HAQ + smoking + BMI + erythrocyte folate | − 160.43 | 5.11 | 0.02* | 0.73 (0.67–0.79) |
| 5 | DAS28 + HAQ + smoking + BMI + erythrocyte folate + <i>ABCC3</i> | − 158.57 | 3.71 | 0.05 | 0.74 (0.68–0.80) |
| 6 | DAS28 + HAQ + smoking + BMI + erythrocyte folate + <i>ABCB1</i> | − 160.15 | 0.57 | 0.45 | 0.74 (0.68–0.80) |
| 7 | DAS28 + HAQ + smoking + BMI + erythrocyte folate + <i>ABCC3</i> + <i>ABCB1</i> | − 158.29 | 4.28 | 0.12 | 0.74 (0.68–0.80) |

Each model was compared to the previous model. Models 6 and 7 were compared to model 4. **P* value < 0.05 was considered significant. DAS28 = DAS28 > 5.1, HAQ = HAQ > 0.6, smoking = current smoking, BMI = BMI > 25 kg/m², erythrocyte folate = erythrocyte folate < 750 nmol/l, *ABCC3* = genotype TC or CC, *ABCB1* = genotype GG or GA

model's clinical applicability. In this combined cohort, the ORs for all predictors were greater than 1 and all predictors, except for the SNPs were significant (Supplementary Table S1). The combined model reached an AUC of 0.74 (95% CI 0.68–0.80; Supplementary Table S1 and Fig. 1) at 3 months.

Additionally, in this combined set, we investigated whether all predictors were required to reach 74% predictive power or whether the model could be further simplified. To do so, we analyzed changes in AUC upon sequential addition of predictors to the model. We started with the most readily available clinical predictors DAS28 > 5.1 and HAQ > 0.6, which generated an ROC with an AUC of 0.67 (95% CI 0.61–0.74; Table 3).

Upon addition of smoking to the model, the AUC significantly increased (*p* = 0.01) to 0.70 (95% CI 0.64–0.76), followed by BMI, upon which the AUC further improved to 0.72 (95% CI 0.66–0.78, *p* = 0.02). Upon addition of

erythrocyte folate to the model the AUC reached 0.73 (95% CI 0.67–0.79, *p* = 0.02). Addition of *ABCB1* and *ABCC3* genotypes did not significantly improve the model (AUC = 0.74, 95% CI 0.68–0.80, *p* = 0.12; Table 3). Hence, the model could be simplified to a model where SNP genotypes were excluded resulting in a model with predictive power of 73%.

To fine-tune the model, all two-way interaction terms between predictors were tested. An interaction term between HAQ and BMI (OR = 3.68 95% CI 1.07–13.14) significantly contributed to the model. This means that a BMI > 25 kg/m² was associated with worse disease activity when HAQ values were > 0.6. Furthermore, an interaction term between HAQ and erythrocyte folate (OR = 0.23, 95% CI 0.06–0.86) also significantly contributed to the model, indicating that low erythrocyte folate concentrations (< 750 nmol/l) significantly predicted insufficient response when HAQ

Table 4 Final prediction model enhanced for clinical implementation

| | β | OR (95% CI) | <i>p</i> |
|---|---------|-------------------------|------------|
| Intercept | − 1.67 | 0.19 (0.07–0.44) | < 0.001*** |
| Baseline DAS28 > 5.1 | 1.34 | 3.81 (2.12–6.99) | < 0.001*** |
| HAQ > 0.6 | 0.44 | 1.56 (0.58–4.33) | 0.383 |
| BMI > 25 kg/m ² | − 0.34 | 0.71 (0.24–2.04) | 0.528 |
| Erythrocyte folate < 750 nmol/l | 1.79 | 5.98 (2.00–19.09) | 0.002** |
| Smoking (current smoker) | 0.81 | 2.26 (1.25–4.16) | 0.008** |
| HAQ > 0.6 × BMI > 25 kg/m ² | 1.30 | 3.68 (1.07–13.14) | 0.040* |
| HAQ > 0.6 × Erythrocyte folate < 750 nmol/l | − 1.46 | 0.23 (0.06–0.86) | 0.031* |
| AUC | | 0.75 (95% CI 0.69–0.81) | |
| Hosmer–Lemeshow test | | | 0.634 |

β = beta coefficient of the final logistic regression model. OR (95% CI) = odds ratio with 95% confidence interval. The model was constructed in the combined dataset (tREACH + U-Act-Early, *N* = 264)

* *P* values < 0.05, ***P* values < 0.01, ****P* values < 0.001. *AUC* area under the curve. The multiplication sign indicates that there is an interaction between two predictors

values were < 0.6. Hence, interaction terms for HAQ and BMI and HAQ and erythrocyte folate were added to the model. Upon addition of these interaction terms to the model, the AUC of the final model, shown in Table 4, increased to 0.75 (95% CI 0.69–0.81). As mentioned in the Methods section, we generated new cut-off values for erythrocyte folate and the BMI in the U-Act-Early cohort which, when included, did not result in higher AUCs.

Model Translation to the Clinic

In order to apply the model in clinical practice, the prediction model was integrated into an online platform “Evidencio” [16]. Using this tool, clinicians can easily enter DAS28, HAQ, erythrocyte folate, smoking, and BMI for individual patients using sliding scales and buttons. The model then automatically calculates and presents a probability of insufficient response to MTX for this specific patient. Specificity, positive predictive value (PPV), sensitivity, and negative predictive values (NPV) were calculated for different cut-off values of these

probabilities for insufficient response and are presented in Supplementary File S1.

Our online model can be found in Evidencio using the following link: <https://www.evidencio.com/models/show/2191>. A patient-specific report presenting the results and model interpretation can be downloaded from Evidencio (Supplementary File S1). An example is shown in Fig. 2, where a patient with DAS28 score = 4.9, HAQ = 0.4, BMI = 22 kg/m², erythrocyte-folate = 720 nmol/l, and current smoking status = yes, has a probability of insufficient response of 71.7%, with corresponding specificity of 86% and PPV of 75%.

DISCUSSION

We externally validated our previously developed prediction model for insufficient response to MTX therapy at 3 and 6 months after treatment initiation in early RA patients including all predictors DAS28 > 5.1, HAQ > 0.6, *ABCB1* rs1045642 genotype, *ABCC3* rs4793665 genotype, erythrocyte folate < 750 nmol/l, current smoking and BMI > 25 kg/m². To enhance

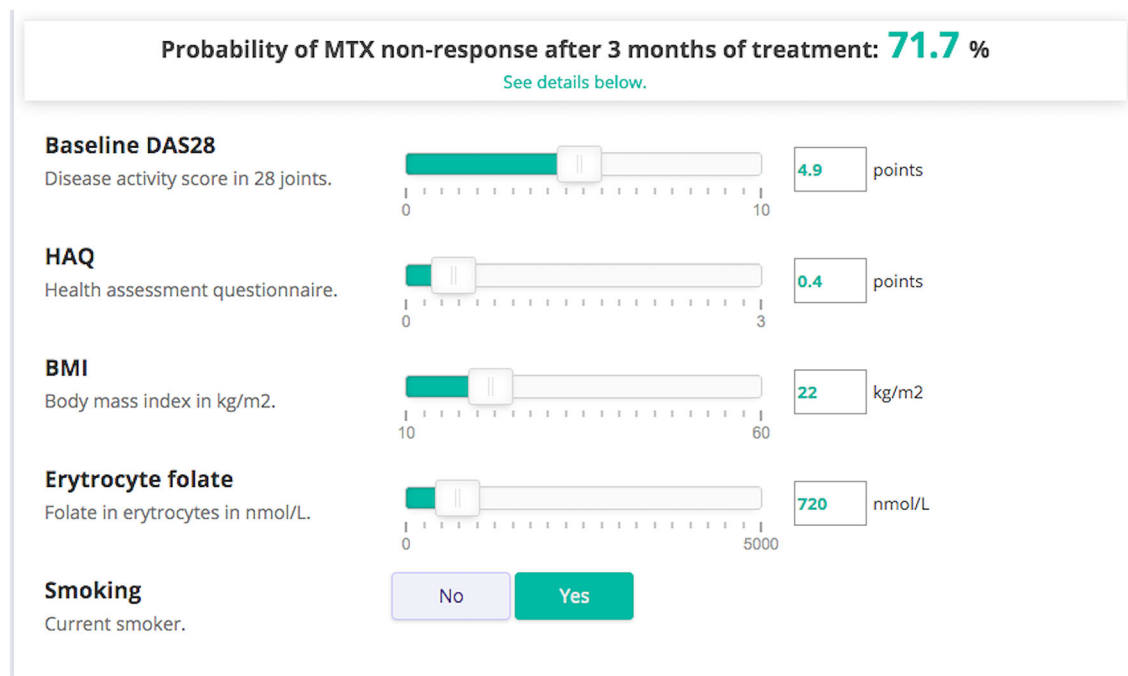


Fig. 2 Example of online platform Evidencio for the implementation of the prediction model. Values for each individual patient can be filled out using the buttons and

slides. Corresponding probability for insufficient response is automatically calculated using the prediction model

clinical applicability and facilitate implementation, the validated model was applied in the combined derivation and validation cohort. This model, without *ABCB1* rs1045642 genotype and *ABCC3* rs4793665 genotype, had an AUC of 0.75, meaning that it classified 75% of the insufficient responders correctly. Currently, according to the EULAR treatment guidelines, MTX is the first-line therapy in RA. Treatment is only up-scaled after 3 to 6 months of insufficient response to MTX, despite evidence supporting a ‘window of opportunity’ for targeted treatment [1, 17]. This window of opportunity is a limited period between diagnosis and RA progression in which the disease could still be modified, radiographic damage/functional disability could be limited, and progression could be slowed down upon early control of disease activity, for which sufficient treatment is required [17]. Our prediction model could assist in identification of insufficient responders at diagnosis: for those with high probability of insufficient response to MTX, additional biologics or JAK inhibitors could be prescribed,

while for patients with low probabilities of insufficient response these expensive treatments could be spared. This distinction at diagnosis could save precious time for insufficient responders, allowing earlier control of disease activity resulting in better long-term outcomes.

We externally validated the model for the first time in its entirety as, besides the clinical and laboratory predictors, the lifestyle predictors (BMI and smoking) were also examined in the U-Act-Early cohort (as opposed to the initial validation in the MTX-Rotterdam cohort, which lacked the life-style predictors) [9]. The strongest predictor was high disease activity at baseline (DAS28 > 5.1) confirming previous findings [7, 18]. Due to differences in treatment intensities (i.e., MTX dose and co-medication) between the derivation and the validation cohort, we investigated whether the model was applicable at 6 months despite step-up treatments after the 3-month mark. Indeed, 71% of the non-responders to MTX were classified correctly, which was similar to the 75% in the

tREACH derivation cohort at 6 months. In the combined cohort, all predictors except for *ABCB1* and *ABCC3* genotypes significantly contributed to the predictive power of the model. Addition of *ABCB1* and *ABCC3* genotypes to the model showed only minimal improvement, resulting in an absolute change in AUC of 0.01, which was not statistically significant. A meta-analysis on the relationship between *ABCB1* genotype and response to MTX in 2014 RA patients showed an association between this genotype and response to MTX, yet our patient group was too small to validate this result [9]. Another recent GWAS study did not show a relationship between *ABCB1* or *ABCC3* and treatment response [19]. Since the differences in predictive power were minor and the effect of *ABCC3* genotype pointed in opposite directions in the two cohorts, possibly indicating a spurious finding, we excluded both genotypes from the model.

In agreement with our study, increased BMI (obese > 30 kg/m²) was previously found to be associated with insufficient response to MTX in RA patients [20, 21]. It is postulated that the effect of BMI on non-response to MTX could be due to the release of proinflammatory adipokines [e.g., leptin, interleukin-6, and or tumor-necrosis factor alpha (TNF- α)] from the adipose tissue [20]. Also, the predictive power of smoking was in accordance with previous studies [6, 7, 20, 22].

So far, several prediction models for MTX non-response have been proposed, which resulted in an AUC between 0.65 and 0.85 [6–8, 22, 23]. Different outcome measures at different time points and combination therapies complicate comparison between models and their validation. However, the best-performing models all included clinical parameters and laboratory parameters, which is in line with our findings [6, 8]. We also showed that clinical predictors (DAS28 > 5.1 and HAQ > 0.6) alone classified fewer insufficient responders correctly (67%) compared to the model combining clinical, life-style (BMI and smoking) and laboratory predictors (erythrocyte folate), which classified 73% of patients correctly. Most clinical predictors and life-style predictors are easy to assess. Erythrocyte folate may not be available in every

laboratory, however the assay is relatively easy to assess [24].

Strengths of this study are that both derivation and validation studies were prospectively designed and that patients in the external validation cohort were included from different districts in the Netherlands. Limitations are that the size of the external validation cohort was limited, however the number of cases in both the internal (tREACH) and external (U-Act-Early) cohort were similar. In addition, the model was validated in an MTX monotherapy group, while it was designed in a combination (GC and csDMARD) therapy group. Commonly, however MTX is co-prescribed with a short course of glucocorticoids (prednisone) as MTX's optimal effect ensues after 8–12 weeks [1]. Despite differences in co-medication between the cohorts, the prediction model had similar predictive value and OR for predictors were in the same direction in both cohorts, indicating that co-medication did not affect the prediction of response to MTX. Another limitation is that smoking status was assessed using questionnaires; possibly biasing the results as smoking behavior could be underreported or underestimated. In future studies, cotinine, the degradation product of nicotine, could be quantified as an objective measure for smoking status, which can easily be determined in serum [25].

Furthermore, we showed that the online platform Evidencio provides an easy tool for implementation of the prediction model in clinical practice. Evidencio is freely available so that the data can be uploaded to automatically validate the model in specific cohorts. In addition, using the Evidencio platform, clinicians can directly use the model in their practice. When a new patient is diagnosed with RA, patient's information on DAS28, HAQ, erythrocyte-folate, BMI, and smoking status can be provided to Evidencio. Subsequently, a probability of insufficient response to MTX with corresponding specificity, positive predictive value (PPV), sensitivity, and negative predictive value (NPV) are provided by the tool and may help clinicians and patients in shared decision-making on step-up treatment with bDMARDs or tsDMARDs. The choice of a cut-off depends on the clinical goal. Taking into consideration the

“window of opportunity” [17] for optimal treatment, we consider it crucial to adequately treat insufficient MTX responders with additional bDMARDs/tsDMARDs. Therefore, our goal for this prediction model was to identify as many insufficient responders as possible, while at the same time attempting to restrict the use of bDMARDs/tsDMARDs to those patients who really need them, hence to avoid misclassification of sufficient responders. Considering this, a cut-off probability of 70% (of insufficient response) could be chosen. At this cut-off, 75% of patients classified as insufficient responder match actual insufficient responders (PPV) and could be treated with additional bDMARDs/tsDMARDs. Additionally, at this cut-off, 86% of all sufficient responders would be correctly classified as such (specificity) and could be spared additional treatment.

The importance of erythrocyte-folate for the predictive power of the model implies that this model is specific in predicting insufficient response to MTX, as MTX is structurally similar to folate. Hence, low erythrocyte folate levels are possibly a surrogate measure for poor MTX absorption, transportation, and MTX accumulation in the cell, as described previously [14]. However, it is possible that a certain proportion of insufficient responders to MTX are difficult-to-treat RA patients who are poorly responsive to various b/ts DMARDs [26, 27]. So far, we cannot identify difficult-to-treat RA patients in advance. Furthermore, as recently argued, treatment strategies could be more important than specific drugs, implying that these patients could still benefit from quicker and more aggressive treatment to reach a certain treatment target when earlier identified as insufficient responders [28, 29].

CONCLUSIONS

We successfully externally validated our previously published prognostic prediction model of insufficient response to MTX, which correctly classified 75% of insufficient responders at 3 months and 71% of non-responders at 6 months of treatment. The model can be used in clinical practice to identify insufficient

responders to MTX with the goal of treating them with additional biologic or JAK inhibitors as early as possible to reduce disease activity and limit joint damage. Application of the tool by means of a clinical trial is warranted.

ACKNOWLEDGEMENTS

We would like to thank all participating institutions and their staff who made this study possible and all subjects enrolled in the U-Act-Early and tREACH study for providing clinical data and material. Furthermore, we would like to thank Pieter Griffioen for his assistance in the SNP and erythrocyte folate experiments and Roche Nederland BV for financing this study and providing the study medication. Finally, we would like to thank Rick Pleijhuis, one of the developers of Evidencio for his assistance in uploading the model to facilitate clinical use.

Funding. This study was funded by Roche Nederland BV. The Rapid Service Fee was funded the Clinical Chemistry Department of the VUmc. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Prior Presentation. This manuscript is based on work that has been previously presented in a poster tour at the 2019 American College of Rheumatology (ACR) annual meeting, November 8–13 Atlanta, Georgia.

Disclosures. Helen R. Gosselt, Maxime M. A. Verhoeven, Maurits C. F. J. de Rotte, Saskia M. F. Pluijm, Ittai B. Muller, Gerrit Jansen, Janneke Tekstra, Maja Bulatović-Ćalasan, Sandra G. Heil, Floris P. J. G. Lafeber, Johanna M. W. Hazes, and Robert de Jonge have nothing to disclose.

Compliance with Ethics Guidelines. This study was approved by the medical ethics committee of the University Medical Center Utrecht (ML22497) and the medical ethics committee of Erasmus Medical Center (MEC-2006-252). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Written informed consent was obtained for all patients.

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

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

REFERENCES

- Smolen JS, Landewé RBMM, Bijlsma JWJ, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann Rheum Dis*. 2020. <https://doi.org/10.1136/annrheumdis-2019-216655>.
- Brown PM, Pratt AG, Isaacs JD. Mechanism of action of methotrexate in rheumatoid arthritis, and the search for biomarkers. *Nat Rev Rheumatol*. 2016;12:731–42. <https://doi.org/10.1038/nrrheum.2016.175>.
- Bakker MF, Jacobs JWG, Welsing PMJ, et al. Low-dose prednisone inclusion in a methotrexate-based, tight control strategy for early rheumatoid arthritis: a randomized trial. *Ann Intern Med*. 2012;156:329–39. <https://doi.org/10.7326/0003-4819-156-5-201203060-00004>.
- Taylor PC. Clinical efficacy of launched JAK inhibitors in rheumatoid arthritis. *Rheumatol (United Kingdom)*. 2019;58:i17–26. <https://doi.org/10.1093/rheumatology/key225>.
- Burmester GR, Pope JE. Novel treatment strategies in rheumatoid arthritis. *Lancet*. 2017;389:2338–48. [https://doi.org/10.1016/S0140-6736\(17\)31491-5](https://doi.org/10.1016/S0140-6736(17)31491-5).
- 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*. 2007;56:1765–75. <https://doi.org/10.1002/art.22640>.
- Teitsma XM, Jacobs JWG, Welsing PMJ, et al. Inadequate response to treat-to-target methotrexate therapy in patients with new-onset rheumatoid arthritis: development and validation of clinical predictors. *Ann Rheum Dis*. 2018;77:1261–7. <https://doi.org/10.1136/annrheumdis-2018-213035>.
- Sergeant JC, Hyrich KL, Anderson J, et al. Prediction of primary non-response to methotrexate therapy using demographic, clinical and psychosocial variables: results from the UK Rheumatoid Arthritis Medication Study (RAMS). *Arthritis Res Ther*. 2018;20:147. <https://doi.org/10.1186/s13075-018-1645-5>.
- De Rotte MC, Pluijm SM, De Jong PH, et al. Development and validation of a prognostic multivariable model to predict insufficient clinical response to methotrexate in rheumatoid arthritis. *PLoS ONE*. 2018;13:e0208534. <https://doi.org/10.1371/journal.pone.0208534>.
- Bijlsma J, Welsing P, Woodworth T, et al. Early rheumatoid arthritis treated with tocilizumab, methotrexate, or their combination (U-Act-Early): a multicentre, randomised, double-blind, double-dummy, strategy trial. *Lancet*. 2016;388:343–55. [https://doi.org/10.1016/S0140-6736\(16\)30363-4](https://doi.org/10.1016/S0140-6736(16)30363-4).
- Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis

- (TRIPOD): the TRIPOD statement. *BMC Med.* 2015;13:1. <https://doi.org/10.1186/s12916-014-0241-z>.
12. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* 1988;31:315–24. <https://doi.org/10.1002/art.1780310302>.
 13. Aletaha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League against rheumatism collaborative initiative. *Arthritis Rheum.* 2010;62:2569–81. <https://doi.org/10.1002/art.27584>.
 14. De Rotte M, De Jong P, Pluijm S, et al. Association of low baseline levels of erythrocyte folate with treatment nonresponse at 3 months in rheumatoid arthritis patients receiving methotrexate. *Arthritis Rheum.* 2013;65:2803–13. <https://doi.org/10.1002/art.38113>.
 15. De Rotte MCFJ, Bulatović-Ćalasan M, Heijstek MW, et al. *ABCB1* and *ABCC3* gene polymorphisms are associated with first-year response to methotrexate in juvenile idiopathic arthritis. *J Rheumatol.* 2012;39:2032–40. <https://doi.org/10.3899/jrheum.111593>.
 16. Van Steenbeek CD, Van Maaren MC, Siesling S, et al. Facilitating validation of prediction models: a comparison of manual and semi-automated validation using registry-based data of breast cancer patients in the Netherlands. *BMC Med Res Methodol.* 2019;19:117. <https://doi.org/10.1186/s12874-019-0761-5>.
 17. Burgers LE, Raza K, Van Der Helm-Van Mil AH. Window of opportunity in rheumatoid arthritis definitions and supporting evidence: from old to new perspectives. *RMD Open.* 2019;5:e000870. <https://doi.org/10.1136/rmdopen-2018-000870>.
 18. Smolen JS, Van Vollenhoven RF, Florentinus S, et al. Predictors of disease activity and structural progression after treatment with adalimumab plus methotrexate or continued methotrexate monotherapy in patients with early rheumatoid arthritis and suboptimal response to methotrexate. *Ann Rheum Dis.* 2018;77:1566–72. <https://doi.org/10.1136/annrheumdis-2018-213502>.
 19. Taylor JC, Bongartz T, Massey J, et al. Genome-wide association study of response to methotrexate in early rheumatoid arthritis patients. *Pharmacogenomics J.* 2018;18:528–38. <https://doi.org/10.1038/s41397-018-0025-5>.
 20. Levitsky A, Brismar K, Hafström I, et al. Obesity is a strong predictor of worse clinical outcomes and treatment responses in early rheumatoid arthritis: results from the SWEFOT trial. *RMD Open.* 2017;3:e000458. <https://doi.org/10.1136/rmdopen-2017-000458>.
 21. Jawaheer D, Olsen J, Lahiff M, et al. Gender, body mass index and rheumatoid arthritis disease activity: results from the QUEST-RA study. *Clin Exp Rheumatol.* 2010;28:454–61.
 22. Eektimmerman F, Allaart CF, Hazes JM, et al. Validation of a clinical pharmacogenetic model to predict methotrexate nonresponse in rheumatoid arthritis patients. *Pharmacogenomics.* 2019;20:85–93. <https://doi.org/10.2217/pgs-2018-0144>.
 23. 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. *Pharmacogenom J.* 2018;18:539–45. <https://doi.org/10.1038/s41397-018-0017-5>.
 24. Pfeiffer CM, Zhang M, Jabbar S. Framework for laboratory harmonization of folate measurements in low- and middle-income countries and regions. *Ann N Y Acad Sci.* 2018;1414:96–108. <https://doi.org/10.1111/nyas.13532>.
 25. Vartiainen E, Seppälä T, Lillsunde P, et al. Validation of self-reported smoking by serum cotinine measurement in a community-based study. *J Epidemiol Commun Health.* 2002;56:167–70. <https://doi.org/10.1136/jech.56.3.167>.
 26. Roodenrijs NMT, De Hair MJH, Van Der Goes MC, et al. Characteristics of difficult-to-treat rheumatoid arthritis: results of an international survey. *Ann Rheum Dis.* 2018;77:1705–9. <https://doi.org/10.1136/annrheumdis-2018-213687>.
 27. De Hair MJH, Jacobs JWG, Schoneveld JLM, et al. Difficult-to-treat rheumatoid arthritis: an area of unmet clinical need. *Rheumatology.* 2018;57:1135–44. <https://doi.org/10.1093/rheumatology/kex349>.
 28. Chatzidionysiou K, Sfikakis PP. Low rates of remission with methotrexate monotherapy in rheumatoid arthritis: review of randomised controlled trials could point towards a paradigm shift. *RMD Open.* 2019;5:993. <https://doi.org/10.1136/rmdopen-2019-000993>.
 29. Drosos AA, Pelechas E, Voulgari PV. Treatment strategies are more important than drugs in the management of rheumatoid arthritis. *Clin Rheumatol.* 2020;39:1363–8. <https://doi.org/10.1007/s10067-020-05001-x>.