Chapter 6

External Validation of the Clinical Pharmacogenetic Model for predicting MTX monotherapy efficacy using a Swedish Cohort of patients with recent-onset rheumatoid arthritis

W.M. Kooloos¹, T.W.J. Huizinga², H.-J. Guchelaar¹, L. Klareskog³, L. Padyukov³, J.A.M. Wessels¹ and R.F. van Vollenhoven³

¹ Clinical Pharmacy & Toxicology, Leiden University Medical Center, Leiden, The Netherlands.
² Rheumatology, Leiden University Medical Center, Leiden, The Netherlands.
³ Rheumatology, Karolinska University Hospital, Stockholm, Sweden

Manuscript in preparation
Abstract

Objective.
Previously, comparison of genetic, demographic and other clinical factors between methotrexate (MTX) responders and nonresponders has led to a clinical pharmacogenetic model including 4 genetic polymorphisms and 4 nongenetic factors to predict MTX therapy outcome. The aim of this study is to validate this predictive model in Swedish patients with recent-onset RA.

Methods
Genetic and nongenetic factors were collected in 355 patients participating in the Swefot clinical trial. All patients received up to 20 mg MTX monotherapy weekly for 3 to 4 months. After this treatment period, efficacy was evaluated by Disease Activity Score (DAS28). Based on individual scores calculated by factors compromising the prediction model, patients were classified as predicted responders, nonresponders and intermediate responders. Hereby, predicted response and observed response on MTX were compared by calculating accuracy, true negative and true positive predictive values (TNPV and TPPV) and by constructing receiver operating curves (ROC). Furthermore, predictive values of the original BeSt cohort, in which the predictive model was derived, and the validation cohort were compared.

Results
At baseline, no significant differences in frequencies of genetic and nongenetic factors compromising the predictive model were observed between this validation cohort and the BeSt cohort. With application of the model, the TNPV and TPPV observed in patients of the validation cohort, were significantly lower compared to the values observed in the original cohort (for TPPV 70% vs. 95%; for TNPV 68% vs. 86% respectively; all p<0.05). Regarding the number of classifiable patients, the level of accuracy of the model and the AUC, no significant differences were found between the two cohorts.

Conclusion
In this study, a pharmacogenetic model for predicting efficacy of MTX in patients with RA was validated. Our data imply that response prediction with this model is feasible in a substantial part of the patients. In order to implement this model in rheumatology clinical practice, additional replication and (ideally) performance of prospectively designed studies is eligible.
Introduction

Over the last two decades, several genetic risk factors for RA like HLA-DRB1 (shared epitope) and PTPN22 have been introduced (1;2). Besides the diagnostic ability of genetics, it has been demonstrated, although to a lesser extent, that genetics could also influence treatment outcome in RA patients (3-5). Intelligibly, a pharmacogenetic effect could be an explanation for the high variability of effective drug responses, up to 30-40%, between RA patients in large clinical trials (6;7).

Most genetic factors in relation to RA treatment outcome have been studied with MTX as the drug under study. Specifically, increasing knowledge about the hypothetical mechanism of MTX action and its role in the inflammatory pathway in RA has led to a substantial number of candidate genetic variants, mostly single nucleotide polymorphisms (SNPs), potentially influencing its efficacy in RA patients (8). However, genetics can not solely account for variability of effective drug response. Clinical and demographic factors, like disease activity- and rheumatoid factor (RF) status, have been shown to be related with treatment outcome (9-11). Still, only a part of the reported pharmacogenetic studies have also included nongenetic factors for the association with treatment outcome. However, in order to detect an individual or a synergistic effect of biomarkers on MTX response, genetic and nongenetic factors should be analyzed collectively.

In a previous study of our group, comparison of genetic, demographic and other clinical factors between MTX responders and nonresponders led to a clinical pharmacogenetic model to predict MTX therapy outcome in a Dutch cohort of patients with early RA (12). This model demonstrated a true positive (predictive) rate of 95% and a true negative (predictive) rate of 86% and was able to categorize 60% of the patients into responders or non-responders defined as achieving DAS >2.4 and DAS ≤2.4 at 6 months, respectively. Despite the potential ability of the model to predict MTX efficacy, validation in larger cohorts is required before this pharmacogenetic model can assist rheumatologists in treatment-decision making. Therefore, the aim of this study is to validate the performance of this previously designed predictive model in an independent Swedish cohort of patients with recent-onset rheumatoid arthritis treated with MTX monotherapy (Swefot).

Patients and methods

RA patients

The 355 patients enrolled in this study are originated from a cohort participating in the “Swedish Pharmacotherapy” (Swefot) randomized clinical trial (13). In this trial, the addition of conventional DMARDs was compared with the addition of anti-TNFα drugs in patients with early rheumatoid arthritis (RA), who had failed initial treatment with MTX at 3 to 4 months.

Patients were recruited by rheumatologists in fifteen rheumatology units in Sweden, which collaborated in this study. Inclusion criteria contained a diagnosis of RA according to the revised ACR (formerly ARA) criteria; RA symptom duration of less than 1 year; no prior disease modifying anti-rheumatic drug (DMARD) therapy; no oral glucocorticoid therapy or stable glucocorticoid therapy for at least 4 weeks of at most 10 mg daily prednisolone (or equivalent); Disease Activity Score (DAS) based on 28-joint count of more than 3.2. Exclusion criteria were contraindications to any of the trial medications, and prior treatment with any DMARD. All patients gave written informed consent prior to inclusion.
Study design
In the Swefot trial, all patients (n=487) started with MTX monotherapy at an initial dose of 10 mg weekly. This dosage was increased biweekly by 5 mg increasing to a maximum of 20 mg weekly. Next to MTX, folic acid suppletion was prescribed in tablets of 5 mg to be taken 1-6 times weekly. In this study, monitoring of liver enzymes and blood counts was performed and abnormalities could lead to dose adjustments based on well-established clinical routines.
At a follow-up visit at least 3 and at most 4 months after the baseline visit, disease activity scores (DAS28) (14) were estimated. If patients were responders on MTX therapy (defined in the protocol as DAS28 <3.2), MTX was continued and patients were followed in usual care. If patients did not obtain response (defined as DAS28 >3.2), patients were randomized to either arm A (the addition of sulphasalazine plus hydroxychloroquine) or arm B (the addition of infliximab) according to the study protocol (REF Swefot study).
Patients who received MTX monotherapy and were evaluated at the follow-up visit at least 3 and at most 4 months after the baseline visit and for whom DNA samples were available (N=355) were included in the current analysis.

Clinical evaluation
The clinical pharmacogenetic predictive model was based on obtaining good clinical response as defined by DAS44 (12). Specifically, responders were defined as patients who were receiving MTX and had a DAS of ≤2.4 (good clinical response). Nonresponders were defined as patients who were receiving MTX and had a DAS of >2.4.
Notably, the original model is based on the DAS44 which includes a 44-joint count. In order to validate the predictive model in the SWEFOT data, main clinical endpoint for each patient in the Swefot cohort was converted: patient’s DAS28 scores (based on 28 joints) were recalculated to DAS scores based on a 44-joint count (DAS44) using the transformation formula: DAS28 = (1,072 X DAS44) + 0,938 (15).
Of the 355 patients included for the efficacy analysis, patients were unavailable for efficacy analysis due to lack of information about RF status (N=3), missing DAS at baseline and/or after 3-4 months (N=4) or due to incomplete genotype data (N=15). Consequently, remaining patients (N=333) were included for validation of the predictive model.

Genotyping
The standard method used for DNA extraction is modified salting-out method. SNPs for analysis in this study were selected according to the genetic polymorphisms previously included in the pharmacogenetic model (12). The selected SNPs were in genes coding for adenosine monophosphate deaminase (AMPD1), aminoimidazole, carboxamide ribonucleotide transformylase (ATIC), inosine triphosphate pyrophosphatase (ITPA) and methylenetetrahydrofolate dehydrogenase (MTHFD1).
Specifically, these SNPs were analyzed: AMPD1 34C>T (rs17602729), ITPA 94A>C (rs1127354), ATIC 347C>G (rs2372536) and MTHFD1 1958G>A (rs17850560).
The method used for genotyping was TaqMan allelic Discrimination (Applied Biosystems, Foster City, U.S.A.). 5’ Nuclease assay was performed according to a standard protocol in a 384-well format with 10 ng of DNA per sample. Detection of the final fluorescent signals from probes, which targeted alleles for each SNP, was performed at 7900 Sequence Detector (Applied Biosystems, Foster City, U.S.A.).
During the genotyping the following frequency distributions of these SNPs were observed: for AMPD1 34C>T, 74% CC, 24% CT, 2% TT; for ITPA 94A>C, 1% AA, 14% AC, 85% CC; for ATIC 347C>G, 47% CC, 42% CG, 11% GG and for MTHFD1 1958G>A, 27% GG, 49% GA, 24% AA. Genotype frequencies for all 4 SNPs were in Hardy-Weinberg equilibrium in this cohort (p>0.05). The
success rates for each SNP were as follows: for AMPD1 34C>T 99%; for ITPA 94A>C 98%, for ATIC 347C>G 99% and for MTHFD1 1958G>A 98%.

**Statistical analysis**

Swefot data were analysed for validation of the previously designed predictive model for MTX efficacy in the BeSt cohort (12). Briefly, reciprocal comparison in multivariate regression analyses of 17 polymorphisms and 24 nongenetic factors in the BeSt cohort led to a predictive model for MTX efficacy. This model consisted of gender, RF and smoking status, the DAS at baseline, and 4 SNPs in the AMPD1, ATIC, ITPA, and MTHFD1 genes. Each patient was scored based on the regression coefficients of the independent variables and categorized in three groups: patient with scores of ≤3.5 (predicted response), patients with scores between 3.5 and 6 (predicted intermediate response) and patients with scores of ≥6 (predicted non-response). Additionally, a nongenetic model was developed based on gender, RF and smoking status, the DAS at baseline, which led to a subdivision of patients based on achieving a score of ≤2 (predicted responder), a score of >2 but <5.5 (intermediate predicted responder) and a score of ≥5.5 (predicted nonresponder).

For initial analysis in this study, baseline factors (included the prediction model) between patients in the BeSt cohort and the validation cohort were compared using chi-square test. Next, Swefot patients were individually scored based on the sum of points obtained by each baseline factor included in the original predictive model. Based on these calculated scores, patients were classified in one of the three groups of predicted response according to the original pharmacogenetic- and nongenetic model. Predicted response and observed response on were compared by calculating true negative and true positive predictive values (TNPV and TPPV) and levels of accuracy. Specifically, accuracy was calculated by the proportion of true results (the number of true positives en negatives) in the patient population.

Additionally, receiver operating characteristic (ROC) curves were constructed to evaluate the discriminative performance of the pharmacogenetic model in comparison with the nongenetic model in the SWEFOT cohort. Finally, levels of accuracy, ROC curves and TNPV and TPPV between the BeSt cohort and the validation cohort were compared using a chi-square test.

Notably, due to the absence of information on smoking status in the Swefot cohort, smoking status per individual could not be applied in the pharmacogenetic model. In this way, the maximum number of points which could be obtained was 10.5 (Table 1). Also, due to this absence the nongenetic model could not be optimally tested in the Swedish validation cohort and was therefore left out of the analysis.

All statistical analyses were performed using SPSS 16.0 software (SPSS Inc., Chicago, IL, USA).

**Results**

**Baseline characteristics**

The response of the included RA patients (N=333) at 3-4 months after start with treatment of MTX monotherapy was 41% according to DAS44 ≤2.4. This was not significantly different in comparison with obtained good clinical response of patients in the original BeSt cohort at 6 months (41% vs. 47%; p>0.05) (12). In addition, no significant differences in baseline factors included the prediction model between patients in the validation cohort and the BeSt population were observed (Table 1).
Table 1. Comparison of baseline factors between patients in the original (BeSt) cohort and patients in the validation (Swefot) cohort

<table>
<thead>
<tr>
<th>Variable</th>
<th>Score</th>
<th>Original cohort in % (N=186)</th>
<th>Validation cohort in % (N=333)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0</td>
<td>30</td>
<td>28</td>
</tr>
<tr>
<td>Female</td>
<td>1</td>
<td>70</td>
<td>72</td>
</tr>
<tr>
<td>DAS at baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤3.8</td>
<td>0</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>&gt;3.8 and ≤5.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd quartile</td>
<td>3</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>3</td>
<td>26</td>
<td>30</td>
</tr>
<tr>
<td>&gt;5.1</td>
<td>3.5</td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td>Rheumatoid factor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negativity</td>
<td>0</td>
<td>32</td>
<td>30</td>
</tr>
<tr>
<td>Positivity</td>
<td>1</td>
<td>68</td>
<td>70</td>
</tr>
<tr>
<td>MTHFD1 1958 AA</td>
<td>1</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>AMPD1 34 CC genotype</td>
<td>1</td>
<td>74</td>
<td>75</td>
</tr>
<tr>
<td>ITPA 94 A-allele carrier</td>
<td>2</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>ATIC 347 G-allele carrier</td>
<td>1</td>
<td>53</td>
<td>53</td>
</tr>
</tbody>
</table>

a. Abbreviations: AMPD1 = adenosine monophosphate deaminase; ATIC = aminoimidazole, carboxamide ribonucleotide transformylase; ITPA = inosine triphosphate pyrophosphatase; MTHFD1 = methylenetetrahydrofolate dehydrogenase; DAS = Disease activity score
b. No significant differences between the validation cohort and the population were observed for gender, DAS at baseline (based on quartiles), RF status and MTHFD1, ATIC, AMPD1, and ITPA genotype frequencies (p>0.05)
Validation of the pharmacogenetic model in the Swedish cohort

Assigned scores as defined by the pharmacogenetic model for the Swedish cohort ranged from 1 to 10.5. Application of the pharmacogenetic model cut off values for predicted (non)response to MTX monotherapy resulted in a TPPV rate of 70% (38 of 54 patients) and a TNPV rate of 68% (122 of 179 patients). In total, 233 patients (70%) in the Swefot cohort were classified as either predicted responder or nonresponder, whereas 100 patients (30%) were classified as predicted intermediate responders (Table 2). Hereby, the accuracy of the model in this cohort of patients was 48%, which represents the proportion of true results (TP+TN/All patients= 33+122/333).

In figure 1, a ROC of the pharmacogenetic model is demonstrated. Hereby, the discriminative ability of the model (AUC-area under the curve) was 75% (95%C.I. 70%-81%).

<table>
<thead>
<tr>
<th>Predicted response</th>
<th>Predicted response</th>
<th>Predicted intermediate response</th>
<th>Predicted non-response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed response</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pharmacogenetic model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responder (DAS ≤2.4)</td>
<td>38</td>
<td>42</td>
<td>57</td>
</tr>
<tr>
<td>Nonresponder (DAS &gt;2.4)</td>
<td>16</td>
<td>58</td>
<td>122</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>54</td>
<td>100</td>
<td>179</td>
</tr>
</tbody>
</table>

Table 2. Comparison of the number of observed and predicted response to MTX therapy at 3-4 months for patients in the validation (Swefot) cohort.

b. Abbreviations: DAS= Disease activity score.

Comparison performance original cohort and validation cohort

TPPV and TNPV demonstrated in this analysis, were significantly lower in the validation cohort compared to the predictive values observed in the original cohort (for TPPV 70% vs. 95%; p<0.0001 and for TNPV 68% vs. 86%; p=0.004, respectively) (12). Regarding the number of patients classified, level of accuracy of the model and AUC no significant differences were found between the Swefot cohort and BeSt cohort (for number of patients classified 70% vs 60%; p=0.182; for accuracy 48% vs. 53%; p=0.572 and for AUC 75% vs. 85%; p=0.111, respectively) (Table 3).
Figure 1. Receiver operating characteristic curve for predicting response to methotrexate in the validation (Swefot) cohort

Table 3. Comparison of predictive values for response to MTX therapy at 3-4 months for patients in the validation (Swefot) cohort* and patients in the original (BeSt) cohort*^b^

<table>
<thead>
<tr>
<th>Predictive values</th>
<th>Original Cohort in % (N=186)</th>
<th>Validation cohort in % (N=333)^b^</th>
<th>P-value^b^</th>
</tr>
</thead>
<tbody>
<tr>
<td>N of patients classified (%)</td>
<td>60</td>
<td>70</td>
<td>0.182</td>
</tr>
<tr>
<td>TPPV (%)</td>
<td>95</td>
<td>70</td>
<td>0.0001</td>
</tr>
<tr>
<td>TNPV (%)</td>
<td>86</td>
<td>68</td>
<td>0.0040</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>53</td>
<td>48</td>
<td>0.572</td>
</tr>
<tr>
<td>AUC (%) (95% C.I.)</td>
<td>86 (80-91)</td>
<td>75 (70-81)</td>
<td>0.111</td>
</tr>
</tbody>
</table>

a. Abbreviations: AUC= area under the curve; TPPV= true positive predictive values; TNPV= true negative predictive values
b. Tested by chi-square test
**Discussion**

In this study, a model for predicting the efficacy of MTX monotherapy was validated in a cohort of Swedish RA patients (Swefot). Importantly, no significant differences in frequencies of genetic and nongenetic factors compromising the predictive model were observed between this validation cohort and the BeSt cohort at baseline. Application of this pharmacogenetic model resulted in a TPPV rate of 70% and a TNPV rate of 68%. In total, 233 patients (70%) were classified as either predictive responder or nonresponder. With application of the model, the TPPV and TNPV observed in patients of the validation cohort were significantly lower compared to the values found in the original cohort. However, for the predictive values accuracy, number of patients classified and AUC the results were comparable. Overall, these validation data imply that efficacy of a substantial part of early RA patients treated with MTX could be predicted by this clinical pharmacogenetic model.

Interestingly, the performance of the pharmacogenetic model was already found to be comparable at initial validation in a separate, but much smaller, group of Dutch RA patients (N=38) in the original manuscript (12). In this Dutch group TPPV rate was 70 % and TNPV rate was 68%, whereas 68% of the patients could be categorized as either predictive responder or nonresponder. However, in comparison with the original BeSt cohort significant differences were observed concerning TPPV and TNPV. Several explanations are possible to declare these differences.

Firstly, differences in the performance of the model may partially be due to the lower response rate achieved (DAS ≤2.4) in the validation cohort with a smaller period of time on MTX therapy before evaluation of response compared with the original cohort (41% at 3 months vs. 47% at 6 months, respectively).

Next, no significant associations of the four individual SNPs included in this model and treatment response were found in the validation cohort (p>0.05). In the analyses of the original BeSt cohort, the four genetic variants increased the AUC of the pharmacogenetic model with 9% compared with the nongenetic model (85% vs. 76%, respectively) (12). However, in the validation cohort, a clear difference between AUCs of the pharmacogenetic- and nongenetic models could not be analyzed due to absence of smoking status. Conclusively, it remains unclear, if the four variants are true markers for MTX response, if other variants in the four genes are responsible for the effect on treatment outcome, or whether the extent of genes or the effect of other genes involved in the mechanism of action of MTX is more important than the current genes. Since the MTX responsive phenotype may be considered polygenic, selecting SNPs according to a candidate gene (mono- or oligogenetic) approach will repeatedly lead to only a limited explanation of variance in MTX response. Hereby, other genes could be more involved in MTX’s mechanism of action than *MTHFD1*, *ATIC*, *AMPD1*, and *ITPA*. Future research on the mechanism of action of MTX is therefore required.

Finally, in this study the DAS28 of each patient was converted to the original DAS based on a 44 joint count. However, since cut-off criteria for response are applied to these scores, different response rates could be observed. Specifically, in the Swefot trial, patients were defined as responder if a DAS28 of less than or equal to 3.2 was achieved. Responders in the BeSt study were defined as achieving an original DAS of less than or equal to 2.4. Comparison of these response rates revealed that 18 patients were defined responder according to DAS, but nonresponder according to DAS28. These patients account for almost 13% of the patients in the DAS responder group. Hypothetically, due to a different observed response distribution in the validation cohort, different predictive values could be the result. Therefore additional analyses were performed considering responders according to DAS, but were nonresponder according to DAS28 (n=18), as nonresponders. Regarding the number of patients classified, TPPV, TNPV, accuracy and AUC, no significant differences compared with the results in table 3 were found (data not shown).
The additional recommendation for classification of patients in the intermediate group at baseline, as reported in our previously study (achieving a decrease of $>1.2$ or $\leq 1.2$ in DAS at 3 months) (12), was left out of the validation analyses. Notably, no data was present due to alternative study design of the Swefot clinical trial.

In general, lower values for the Swefot cohort of patients found in this study (except for the number of patients classified) were anticipated. Namely, effect sizes found by association studies could appear to be smaller, but closer to a genuine effect size, than the first reported effect size (16). A well-designed meta-analysis of comparing the effect size (e.g. TPPV) with this predictive model in more cohorts of patients could provide the most optimal legitimate effect size.

Screening for markers in serum of patients with arthritis is a helpful tool and common practice in rheumatology diagnostics. For example, in a meta-analysis by Nishimura et al it was demonstrated that the sensitivity and specificity of testing for RF status within RA patients was 69% and 85% (17). Notably, this meta-analysis included studies that evaluated patients for the utility of RF for diagnosis or suspected RA. The sensitivity and specificity could be recalculated to a TPPV of 82% and a TNPV of 72% for the diagnosis of RA by testing for RF positivity. Interestingly, these predictive values, explained by RF-status, are comparable with the values regarding the predictive model’s TPPV and TNPV found in our study. In this way, testing for response to MTX monotherapy in recent-onset RA patients prior to treatment appears to be qualitatively equivalent to testing for RF status to diagnose RA in clinical practice.

In this study, a pharmacogenetic model for predicting the efficacy of MTX in patients with RA was validated. Notably, the exact role as predictive markers of response for the four genetic polymorphisms included in this model remains to be determined by future studies. Still, our results demonstrate that predicting treatment response is feasible in RA. Additional replication and (ideally) performance of prospectively designed study with this model in large cohorts is warranted to demonstrate the legitimate predictive value in rheumatology practice.
References

(1) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Wellcome Trust Case Control Consortium. Nature 2007 June 7;447(7145):661-78.


