The handle http://hdl.handle.net/1887/46254 holds various files of this Leiden University dissertation

Author: Fu, J.
Title: Systems diagnosis in chronic disease: prediction and evaluation
Issue Date: 2017-02-21
Chapter 4

Systems approach for classifying the response to biological therapies in patients with rheumatoid arthritis in clinical practice

Based on

Junzeng Fu, Herman A. van Wietmarschen, Anita van der Kooij, Mei Wang, Bart V.J. Cuppen, Eduard van Wijk, Yan Schroën, Anne Karien Marijnissen, Thomas Hankemeier, Jacqueline J. Meulman, Floris P.J.G. Lafeber and Jan van der Greef

Systems approach for classifying the response to biological therapies in patients with rheumatoid arthritis in clinical practice

Submitted for publication
Abstract

Background Biological therapies have greatly improved the treatment efficacy in rheumatoid arthritis (RA). However, in clinical practice a significant proportion of patients experience an inadequate response to treatment.

Objective To classify responders and non-responders to biological therapy in RA.

Methods Cold and Heat symptoms accessed by a Chinese medicine (CM) questionnaire and Western clinical data were collected as baseline data, before initiating biological therapy. Categorical principal components analysis with forced classification (CATPCA-FC) approach was applied to the baseline data set to classify responders and non-responders.

Results In this study, 61 RA patients were characterized using a CM questionnaire and clinical measurements. The combination of baseline symptoms (‘preference for warm food’, ‘weak tendon severity’) and clinical parameters (positive rheumatoid factor/anti-cyclic citrullinated peptide antibody, C-reactive protein, creatinine) were able to differentiate responders from non-responders to biological therapies with a positive predictive value of 82.35% and a misclassification rate of 24.59%. Adding CM symptom variables in addition to clinical data did not improve the classification of responders, but it did show 8.3% improvement in classifying non-responders.

Conclusions This study, for the first time includes CM symptoms into a classification analysis of RA patients’ response to biological therapies. The adding of symptoms to the clinical parameters could improve the classification of non-responders. Although this improvement is not significant in the current study, we consider it worthwhile to investigate further the possible potential of symptom variables for reducing inefficient treatment.
Abstract

Background

Biological therapies have greatly improved the treatment efficacy in rheumatoid arthritis (RA). However, in clinical practice a significant proportion of patients experience an inadequate response to treatment.

Objective

To classify responders and non-responders to biological therapy in RA.

Methods

Cold and Heat symptoms accessed by a Chinese medicine (CM) questionnaire and Western clinical data were collected as baseline data, before initiating biological therapy. Categorical principal components analysis with forced classification (CATPCA-FC) approach was applied to the baseline data set to classify responders and non-responders.

Results

In this study, 61 RA patients were characterized using a CM questionnaire and clinical measurements. The combination of baseline symptoms (‘preference for warm food’, ‘weak tendon severity’) and clinical parameters (positive rheumatoid factor/anti-cyclic citrullinated peptide antibody, C-reactive protein, creatinine) were able to differentiate responders from non-responders to biological therapies with a positive predictive value of 82.35% and a misclassification rate of 24.59%. Adding CM symptom variables in addition to clinical data did not improve the classification of responders, but it did show 8.3% improvement in classifying non-responders.

Conclusions

This study, for the first time includes CM symptoms into a classification analysis of RA patients’ response to biological therapies. The adding of symptoms to the clinical parameters could improve the classification of non-responders. Although this improvement is not significant in the current study, we consider it worthwhile to investigate further the possible potential of symptom variables for reducing inefficient treatment.

Introduction

Rheumatoid arthritis (RA) is an autoimmune disease that results in a systemic inflammation, affecting 0.5% to 1% population in Northern Europe and North America. The pathogenesis of RA is not defined as yet, but there is no doubt that complex immune responses are highly related to inflammation and joint erosion. In the late 1990s, tumor necrosis factor (TNF) inhibitors were introduced for the treatment of RA, followed by other kinds of biological agents including the anti-CD 20 agent, IL-6 inhibitor etc. By targeting cell-surface receptors or intracellular pathways, biological agents show powerful capabilities in the modulation of the immune response. Compared to the conventional RA medicines, biological agents can not only reduce disease activity but also decrease or prevent radiographic progression in RA.

Biological therapies can greatly improve the treatment outcome in RA, but there is a significant proportion of patients who have an inadequate response. Approximately 30% of the RA patients failed to respond after (at least) three-month biological treatment. For these non-responders this problem is compounded by high financial cost and significant side effects. Additionally in clinical practice for RA treatment, a biological agent is recommended only if the treatment outcome of non-biologic therapy is not reached with 6 months. Therefore, the opportunity to control the erosion of joints in early RA might be missed, further affecting the long-term outcomes of RA. Hence, it is necessary to better target this medication to patients who will benefit from the biological therapies, and to develop more personalized medication based a novel diagnostic principle.

As a practical medicine, Chinese medicine (CM) has a long history of development and optimization through observation in daily practice. Especially for complex diseases, CM could provide new possibilities for subtyping patients by pattern diagnosis. In CM, RA is defined as a Bi-syndrome. Bi-syndromes consist of multiple subtypes based on symptom patterns, including two basic patterns—‘Cold’ and ‘Heat’. For each pattern of RA, there is a different treatment strategy in CM. We previously reported differences in biological mechanisms between Cold and Heat RA patients, determined by metabolomics measurements of plasma and urine samples as well as gene expression analysis of CD-4 T-cells. A large literature mining study suggests that Cold type of diseases is related to hormone disturbances whereas immune systems disturbances are Heat type related. Therefore, we hypothesize that the Cold or Heat patterns may be associated with the response to biological therapies of RA patients. According to CM, Cold or Heat diagnosis is based on integrating corresponding symptoms. However, in most cases patients diagnosed with Cold RA could also show Heat
RA related symptoms, besides dominant Cold symptoms; or the other way around with RA Heat patients.

Thus, it is difficult to find patients with exclusively Cold or Heat patterns in practice. Therefore, in the present study we focused on the individual Cold and Heat symptoms instead of patterns. Symptoms, which are important for the evaluation of patients’ quality of life and disease burden, are more and more used as patient-reported outcome measures for the evaluation of treatment effects in clinical studies. Many factors from clinical and lab tests have been already reported as predictors/biomarkers of patients’ response to biologic agents, but there is no study including baseline symptoms as potential markers.

In this study we used Eastern and Western diagnostic principles, including Cold and Heat symptoms as well as other baseline disease characteristics, to classify response to biological therapy in patients with RA. If we are able to better classify non-responders and responders at the start of biological therapy, patients can be treated with the best available drugs timely, avoiding side effects and unnecessary financial cost caused by trial-and-error practice.

Materials and Methods

Study design and participants

RA patients were selected from the observational BiOCURA, in which patients with RA starting a biological therapy were recruited. There was no particular intervention in this purely observational study. After three months of treatment with one of the following biological agents: Etanercept, Adalimumab, Golimumab, Certolizumab pegol, Rituximab, Abatacept or Tocilizumab, the outcome of the therapy was assessed according to EULAR response criteria. Re-inclusion after switching to a different biological agent was possible. The type of biological agent provided to each of the patients was decided by their own clinician as is done in routine clinical practice. The study was approved by the Medical Ethical Committee of UMC Utrecht and all subjects gave their written informed consent for participation in any procedure specifically for the study. Our study was restricted to the BiOCURA patients who completed a CM questionnaire at baseline.

Symptom questionnaire

This questionnaire was designed to measure symptoms based on a CM perspective of Cold/Heat patterns on arthritis, which was developed and tested recently. The same questionnaire was used in the present study except translated into the Dutch language, which consists of 57 items separated into five categories (breathing, digestion, climate, quality of the
symptoms and pain). Most of the questions are on a Likert-scale, evaluating the severity and the frequency of symptoms from score 1 to 7. Score 1 was interpreted as never or not and score 7 was very severe or very often. The remaining questions were in yes/no format. The questionnaires were completed before initiating the biological therapy and used as baseline symptom data.

**Demographics and clinical parameters**

Before starting biological therapy, clinical parameters and demographic characteristic were obtained as baseline clinical data. The following demographics and clinical parameters of patients were collected at baseline: gender, age, body mass index (BMI), disease duration, smoking status, alcohol consumption, biological naivety, concomitant DMARDs, 28 tender joint count (TJC), 28 swollen joint count (SJC), 100 mm visual analogue score (VAS), disease activity score in 28 joints (DAS28), erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), creatinine, hemoglobin, alanine aminotransferase (ALT), leukocyte count, rheumatoid factor (RF), platelet count, and anti-cyclic citrullinated peptide antibody (anti-CCP).

After three-month treatment, DAS28 was measured again as present DAS28. By combining the present DAS28 as well as the improvement of DAS28 after three-month treatment, a patient’s response to a biological agent can be evaluated according to EULAR response criteria. A good response was defined as a present DAS28 ≤ 3.2 with a DAS28 reduction > 1.2, whereas a reduction of DAS28 ≤ 0.6 or present DAS28 > 5.1 with a reduction ≤ 1.2 was defined as non-response. In between, a reduction > 1.2 with present DAS28 > 3.2 or a reduction between 0.6 and 1.2 with present DAS28 < 5.1 was specified as a moderate response. Since both good- and moderate responders achieved sufficient response, they were combined as responders in the following data analysis.

**Data analysis**

Firstly, univariate analyses including independent student’s t-tests, chi-square tests, and Kruskal wallis H tests were applied to compare the differences of baseline clinical and demographic data between responders and non-responders. Subsequently, multivariate analysis was performed on combined data from questionnaires, clinical parameters, and demographic characteristics. Since the combined data sets contained variables with different measurement levels (nominal, ordinal, or numeric) that might be nonlinearly related to each other, categorical principal components analysis (CATPCA) was applied.
As a nonlinear principal component analysis technique, CATPCA allows different analysis levels (numeric, ordinal and nominal) for variables and is able to handle categorically and numerically measured variable. In addition, variables can be given weights—the higher a variable weights, the more the solution will be influenced by the variable. When applying a large weight to only one variable, the solution will be dominated by that variable. This feature can be used to perform forced classification (FC) by including a classification variable with a large weight in the analysis, the cases are forced into clusters according to the classification variable. Thus, variables with high component loadings indicate that they are highly related to the classification variable. In a previous study, CATPCA-FC has been used to identify the important symptoms related to Cold and Heat status of patients. Therefore, in the present study we use CATPCA-FC in an exploratory way to identify which baseline data related to treatment response, subsequently to classify responders and non-responders at baseline.

In the study, there were seven biological agents involved. Limited number of subjects for each treatment may lead to instability of CATPCA-FC model. Thus biological agents were grouped according to their mechanisms as TNF inhibitors (Etanercept, Adalimumab, Golimumab, Certolizumab pegol) and non-TNF inhibitors (Rituximab, Abatacept, Tocilizumab). RF and anti-CCP are clinically important parameters for the diagnosis and prognosis of RA patients. The combination of the two parameters may lead to a better classification of treatment outcome. Therefore, RF and anti-CCP were combined in one variable—positive RF/anti-CCP (three categories: both positive, only one positive; both negative). EULAR response (two categories: responder, non-responder) was taken as the classification variable, which is derived from the patient's baseline DAS28 and current DAS28 scores. In addition, the DAS28 scores are calculated from the variables TJC, SJC, VAS and ESR. Consequently, there is a strong correlation between DAS28, TJC, SJC, VAS and ESR and 3-month EULAR response. Therefore, these DAS28 related baseline clinical parameters were excluded from the CATPCA-FC analyses. The rest of baseline clinical parameters and demographic data (described in ‘Demographics and clinical parameters’) were taken as clinical data, which were merged with CM symptom to build a combined model. In the meantime, a symptom model and a clinical model were built separately.

After data screening and recoding, the processed data were further analyzed by CATPCA-FC. The number of components in CATPCA, which is required to restrict the component scores to groups, is set according to the classification variable. Since we have two classes (responders, non-responders) we used one component. The models were internally validated by a leave-one-out cross-validation (LOOCV) procedure. During the LOOCV procedure, a predicted
response and a predicted score were computed for each subject. Details of statistics procedures are described in Supplementary Methods.

In order to determine the predictive ability of each model, the positive prediction value (PPV), negative prediction value (NPV), and misclassification rate (MR) were calculated based on the observed responses and the LOOCV predicted responses. In addition, the receiver operator characteristic (ROC) curve for each model was built with predicted scores as test variables and observed response as the state variable. The ROC curves were compared on the basis of the area under the receiver operator characteristic curve (AUROC)\textsuperscript{33}.

All procedures were applied similarly to the combined model, symptom model, and clinical model. Statistical calculations were performed with IBM SPSS Statistics for Windows, Version 23.0 (IBM Corp., Armonk, N.Y., USA) and MedCalc for Windows, version 16.2.1 (MedCalc Software, Ostend, Belgium); the level of significance was set at 0.05.

**Results**

**Subjects**

A total of 65 RA patients, who completed 75 questionnaires, were enrolled in the study (8 patients were enrolled in the study twice, and one patient was enrolled three times). Each time before an RA patient restarted a new biological therapy, questionnaire data, demographic and clinical parameters were collected as baseline again. Consequently, those patients were regarded as new subjects in the following data analysis. Of those, 12 subjects were excluded due to lack of follow-up information in three months, and 2 subjects answering less than 50% questions in questionnaires were also removed. Finally, there were 61 valid subjects retained in the study: 37 responders and 24 non-responders.

**Baseline characteristics and parameters of patients**

As illustrated in Table 1, the majority of 61 subjects were female (77%), and all subjects had active rheumatoid arthritis (DAS28, 4.28 ± 1.32) before starting biological therapy. After 3-month of biological therapy, 37 subjects were EULAR responders, including 15 good responders and 22 moderate responders, whereas 24 subjects were non-responders. Baseline characteristics of non-responders versus responders differed in terms of DAS28 (mean = 3.64 ± 1.41 vs. 4.69 ± 1.10, p = 0.002), SJC (median = 0.5 (0.0-1.0) vs. 2.0 (0.5-4.0), p = 0.01), ESR (median = 8.0 mm/hr (4.0-23.0) vs. 15.0 mm/hr (11.0-43.0), p = 0.01) and CRP (median = 3.0 mg/mL (2.0-8.0) vs 11.5 mg/mL (3.3-24.0), p = 0.02). Although responders had a higher proportion of positive anti-CCP and positive RF compared to non-responders, there were no
statistical differences (p = 0.06). There were also no significant differences in the rest clinical parameters, demographic data, and the proportion of subjects using DMARDs (Supplementary Table S1) between responders and non-responders.

Table 1. Baseline characteristics of responders and non-responders. P-values for comparisons among groups were calculated by independent t-tests a, Kruskal wallis H tests b and chi-squared tests c based on distribution of the clinical parameter. Bold values indicate significant p-values (p < 0.05).

<table>
<thead>
<tr>
<th></th>
<th>All (n=61)</th>
<th>Non-responders (n=24)</th>
<th>Responders (n=37)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yrs, mean ± SD a</td>
<td>56 ± 13</td>
<td>53 ± 12</td>
<td>57 ± 13</td>
<td>0.30</td>
</tr>
<tr>
<td>Disease duration, yrs, median (IQR) b</td>
<td>10.0 (5.0-18.0)</td>
<td>9.5 (3.5-17.5)</td>
<td>10.0 (5.0-19.5)</td>
<td>0.75</td>
</tr>
<tr>
<td>BMI, kg/m², mean ± SD a</td>
<td>27.0 ± 5.1</td>
<td>28.0 ± 5.1</td>
<td>27.7 ± 9.8</td>
<td>0.92</td>
</tr>
<tr>
<td>Female, n(%) c</td>
<td>47 (77)</td>
<td>17 (71)</td>
<td>30 (81)</td>
<td>0.37</td>
</tr>
<tr>
<td>Biologic-naive, n (%) c</td>
<td>20 (33)</td>
<td>10 (43)</td>
<td>10 (26)</td>
<td>0.28</td>
</tr>
<tr>
<td>DAS28, mean ± SD a</td>
<td>4.28 ± 1.32</td>
<td>3.64 ± 1.41</td>
<td>4.69 ± 1.10</td>
<td>0.002</td>
</tr>
<tr>
<td>TJC, median (IQR) b</td>
<td>6.0 (2.5-11.0)</td>
<td>3.0 (0.0-10.3)</td>
<td>7.0 (4.0-11.0)</td>
<td>0.80</td>
</tr>
<tr>
<td>SJC, median (IQR) b</td>
<td>1.0 (0.0-3.5)</td>
<td>0.5 (0.0-1.0)</td>
<td>2.0 (0.5-4.0)</td>
<td>0.01</td>
</tr>
<tr>
<td>VAS, mean ± SD a</td>
<td>54 ± 24</td>
<td>50 ± 25</td>
<td>56 ± 23</td>
<td>0.41</td>
</tr>
<tr>
<td>ESR, mm/hr, median (IQR) b</td>
<td>13.0 (6.0-33.5)</td>
<td>8.0 (4.0-23.0)</td>
<td>15.0 (11.0-43.0)</td>
<td>0.02</td>
</tr>
<tr>
<td>CRP, mg/mL, median (IQR) b</td>
<td>6.0 (2.0-16.0)</td>
<td>3.0 (2.0-8.0)</td>
<td>11.5 (3.3-24.0)</td>
<td>0.01</td>
</tr>
<tr>
<td>Positive RF and anti-CCP, n (%) c</td>
<td>26 (42.6)</td>
<td>5 (21)</td>
<td>21 (57)</td>
<td>0.02</td>
</tr>
<tr>
<td>Positive anti-CCP,n (%) c</td>
<td>32 (52)</td>
<td>9 (37)</td>
<td>23 (68)</td>
<td>0.06</td>
</tr>
<tr>
<td>Positive RF, n (%) c</td>
<td>35 (57)</td>
<td>10 (43)</td>
<td>25 (67)</td>
<td>0.06</td>
</tr>
<tr>
<td>ALT, U/L, median (IQR) b</td>
<td>21.0 (17.0-31.0)</td>
<td>23.5 (19.0-30.8)</td>
<td>19.0 (17.0-35.0)</td>
<td>0.67</td>
</tr>
<tr>
<td>Platelets, 10×10⁶/L, median (IQR) b</td>
<td>260 (206.5-342.5)</td>
<td>259.0 (208.8-320.8)</td>
<td>293.0 (204.5-352.5)</td>
<td>0.85</td>
</tr>
<tr>
<td>Leukocytes, 10⁶/L, median (IQR) b</td>
<td>7.6 (6.15-9.3)</td>
<td>7.6 (4.9-9.4)</td>
<td>7.7 (6.3-9.3)</td>
<td>0.50</td>
</tr>
<tr>
<td>Creatinine, µmol/L, mean ± SD a</td>
<td>74 ± 14</td>
<td>76 ± 12</td>
<td>73 ± 15</td>
<td>0.40</td>
</tr>
<tr>
<td>Hemoglobin, mmol/L, mean ± SD a</td>
<td>8.4 ± 0.8</td>
<td>8.6 ± 0.7</td>
<td>8.3 ± 0.8</td>
<td>0.12</td>
</tr>
</tbody>
</table>

DAS28=0.56×√TJC+0.28×√SJC+0.70×ln(ESR)+0.014×VAS²²

ALT, alanine aminotransferase; Anti-CCP, anti-cyclic citrullinated peptide antibody; BMI, body mass index; CRP, C-reactive protein; DAS28, disease activity score in 28 joints; ESR, erythrocyte sedimentation rate; IQR, interquartile range; VAS, 100 mm visual analogue scale; RF, rheumatoid factor; SD, standard deviation; SJC, 28 swollen joint count; TJC, 28 tender joint.

**Systems approach for classifying the responders and non-responders to biological therapy by CATPCA**

**Data screening, recoding and discretization**

After data screening, the inconsistent answers found in the questionnaires of 4 subjects were set to missing. In the recoding procedure, 13 variables related to 6 symptoms (shortness of
breath, dryness of the throat, tender abdomen, diarrhea, cold feeling, fever) were simplified and reorganized into 6 new variables. The variable ‘color of phlegm’ was with only one category, therefore, could not be used in the analyses. In total, there were less than 10% missing data identified in the symptom data set. Clinical variables with more than 15 continuous values were discretized by grouping the values into 15 categories. In total, 49 symptoms, 7 clinical parameters, and 12 demographics were taken as the baseline data for the further CATPCA. Details of included data are shown in Supplementary Table S2.

Multivariate analysis on baseline data—CATPCA-FC

In order to search for baseline data related to response, CATPCA-FC approach was employed. Baseline data of 61 subjects were included in the analyses with optimal analysis levels; EULAR response was set with large weight as the classification variable.

Variables with absolute component loadings > 0.30 were selected and included in the final models. For the final symptom model, two symptoms ‘weak tendon severity’ and ‘preference for warm food’ remained, and the final clinical model consisted of CRP, creatinine, and positive RF/anti-CCP. The final combined model comprised of the variables from both clinical and symptom model. The transformation plots and component loadings of variables in the combined model are shown in Table 2 and Figure 1, respectively. The transformation plot represents the transformed values of a variable based on its optimal analysis level, and the component loading indicates the (positive or negative) relationship between the transformed variable and the classification variable.

Table 2. Component loadings and analysis levels of selected variables in the final combined model

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Component loading</th>
<th>Analysis level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creatinine</td>
<td>-0.349</td>
<td>Spline nominal</td>
</tr>
<tr>
<td>Preference for warm food</td>
<td>-0.312</td>
<td>Nominal</td>
</tr>
<tr>
<td>Positive RF/anti-CCP</td>
<td>0.355</td>
<td>Ordinal</td>
</tr>
<tr>
<td>Weak tendon severity</td>
<td>0.376</td>
<td>Nominal</td>
</tr>
<tr>
<td>CRP</td>
<td>0.406</td>
<td>Spline ordinal</td>
</tr>
</tbody>
</table>

CRP, C-reactive protein; RF, rheumatoid factor; anti-CCP, anti-cyclic citrullinated peptide antibody
Figure 1. Transformation plot of the classification variable EULAR response and five classifier variables in the combined model. The categories of measured values are shown on the X-axis, and the quantification using optimal analysis is displayed on the Y-axis. The symptoms ‘weak tendon severity’ and ‘preference for warm food’ are with nominal analysis level; measured value 1 was interpreted as never or not and score 6 was very severe or very often. The clinical variables, RF/anti-CCP are with ordinal analysis level; C-reactive protein are with spline ordinal analysis level (degree 2, interior knots 2); creatinine is with spline nominal analysis level (degree 2, interior knots 2).

Classifiers ‘Positive RF/anti-CCP’, ‘weak tendon severity’ and CRP were found to have positive loading, indicating that the transformed values of these variables were higher in responders. Specifically, responders scored relatively high in the upper region of ‘weak tendon severity’ (4 through 5) and CRP (> 8 mg/mL), while non-responders scored relatively low on these variables. For the positive RF/anti-CCP, the category of both positive (high score) was associated with responders while both negative or one negative (low score) were associated with non-responders. In contrast, of ‘preference for warm food’ with negative loading, responders scored in the low region (1 and 2), while non-responders had a high score in either 4 or 6. The transformation plot of creatinine (negative loading) shows a curve with a dip in the middle, to be interpreted as low, high, and intermediate values of creatinine being associated with responders, while non-responders tend to have moderate low and moderate

Table 3. Prediction parameters of the combined model, symptom model, and clinical model

<table>
<thead>
<tr>
<th>Classified</th>
<th>Observed</th>
<th>Non-responders</th>
<th>Responders</th>
<th>PPV a</th>
<th>NPV b</th>
<th>MR c</th>
<th>AUROC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>Non-responders</td>
<td>18</td>
<td>6</td>
<td>82.35</td>
<td>66.67</td>
<td>24.59</td>
<td>0.769 (0.644 to 0.867)</td>
</tr>
<tr>
<td></td>
<td>Responders</td>
<td>9</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symptom</td>
<td>Non-responders</td>
<td>12</td>
<td>12</td>
<td>68.42</td>
<td>52.17</td>
<td>37.70</td>
<td>0.574 (0.441 to 0.700)</td>
</tr>
<tr>
<td></td>
<td>Responders</td>
<td>11</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>Non-responders</td>
<td>16</td>
<td>8</td>
<td>77.78</td>
<td>64.00</td>
<td>27.87</td>
<td>0.765 (0.639 to 0.864)</td>
</tr>
<tr>
<td></td>
<td>Responders</td>
<td>9</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PPV, positive predictive value; NPV, negative predictive value; MR, misclassification rate; AUROC, area under the receiver operator characteristic curve; CI, confidence interval.

a. Positive predictive values = true positive/(true positive + false positive) × 100%
b. Negative predictive values = true negative/(true negative + false negative) × 100%
c. Misclassification rate = (false positive + false negative)/number of all subjects × 100%
high values. None of the variables has a high loading in the model, suggesting that these variables by themselves could not give a clear-cut classification between response and non-response.

On the basis of the prediction results of the LOOCV procedure, prediction parameters PPV, NPC, MR and AUC were calculated for each model. As shown in Table 3, the combined model with both clinical and symptom classifiers had the best accuracy in differentiating responders from non-responders (MR = 24.59%) and predicting responders (PPV = 82.35%). Two observed non-responders, who were incorrectly classified as responders in the clinical model, were reclassified correctly as non-responders in the combined model, whereas no improvement in classifying responders were found. The reclassification of non-responders in the combined model was 8.3% ((18-16)/24 × 100%) better than the clinical model. The AUROC of the combined model is significantly larger than that of the symptom model (p = 0.02), whereas the differences of AUROCs between clinical and combined model were not significant (Supplementary Table S3).

<table>
<thead>
<tr>
<th>Classified Model</th>
<th>Non-responders</th>
<th>Responders</th>
<th>PPV a</th>
<th>NPV b</th>
<th>MR c</th>
<th>AUROC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined model</td>
<td>Non-responders</td>
<td>18</td>
<td>6</td>
<td>82.35</td>
<td>66.67</td>
<td>24.59</td>
</tr>
<tr>
<td></td>
<td>Responders</td>
<td>9</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symptom model</td>
<td>Non-responders</td>
<td>12</td>
<td>12</td>
<td>68.42</td>
<td>52.17</td>
<td>37.70</td>
</tr>
<tr>
<td></td>
<td>Responders</td>
<td>11</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical model</td>
<td>Non-responders</td>
<td>16</td>
<td>8</td>
<td>77.78</td>
<td>64.00</td>
<td>27.87</td>
</tr>
<tr>
<td></td>
<td>Responders</td>
<td>9</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PPV, positive predictive value; NPV, negative predictive value; MR, misclassification rate; AUROC, area under the receiver operator characteristic curve; CI, confidence interval.

a Positive predictive values = true positive/(true positive + false positive) × 100%
b Negative predictive values = true negative/(true negative + false negative) × 100%
c Misclassification rate = (false positive + false negative)/number of all subjects × 100%

**Discussion**

In this study, we described the application of CATPCA-FA on baseline data to classify response to biological therapies in patients with RA. The baseline data, which included symptoms from a Chinese diagnosis questionnaire, clinical parameters, and demographics data, were collected before the starting of the biological treatment. The combination of two
symptom and three clinical classifiers in CATPCA-FA model was able discriminate responders and non-responders with good accuracy (MR = 24.59%, AUROC = 0.769).

In order to compare the classification abilities of CM symptoms and clinical data, we developed and tested three CATPCA-FC models—one model using the symptoms only, one model using the clinical data only and one model using all the baseline data (the combined model). Although the combined model seems to perform slightly better than the clinical model in PPV, NPV and MR, no statistical differences were found in AUROC. However the AUROC of the combined model is significantly larger than that of the symptom model, suggesting that the combination of CM symptoms and clinical data has higher classification ability compared to the symptoms. Additionally, adding symptoms to the clinical data seemed to slightly improve the classification of non-responders. It is an interesting observation since this indicates that symptoms provide only limited extra information on classifying non-responders, besides the information captured with the clinical data. This might be explained by the complexity of this study cohort. Subjects were selected from a pragmatic-observational study, in other words our subjects were from a highly heterogeneous cohort reflecting real-life situations regarding RA. The subjects did not express clearly distinctive Cold or Heat symptom patterns, therefore the mixed Cold and Heat symptoms in a patient could have limited their classifying value. Besides, the CM symptom patterns reflecting the dynamic health of the entire system might not be suitable to classify the response to a Western symptom treatment such as blocking a single protein. A homogenous cohort of patients with more clear Cold and Heat symptom patterns is needed for the future research, which might provide a more clear view on the role of CM symptoms for classification.

It is the first time that a Cold/Heat symptom-based questionnaire was used to classify response to biological therapy in patients with RA. It is also the first time that CATPCA-FC is used as an exploratory classification method for discriminating the response to biological therapy in RA. The combined model consisted of two key symptom variables (‘preference for warm food’, ‘weak tendon severity’) related to EULAR response to biological therapy. Based on CM theory, ‘Weak tendon severity’ is a typical Heat pattern related symptom and ‘preference for warm food’ is a Cold pattern related symptom, and these two symptoms could be present in either Cold pattern or Heat pattern RA patients. According to the component loadings (Table 2), ‘weak tendon severity’ is positively related to EULAR response whereas ‘preference for warm food’ is negatively related to response. In other words, it seems that RA patients with evident Heat symptom and less evident cold symptom might respond better to biological therapy. However, a previous study reported more apoptosis of CD4+ T cell in RA Heat patients compared to RA Cold patients. Another study showed the resistance of
synovial and inflammatory cells to apoptosis in RA patients, which is suggested to be the mechanism of RA. Therefore, enhancing apoptosis might result in clinical improvement in RA. It is furthermore suggested that the use of anti-TNF-α antibody might induce apoptosis and clinical improvement. Lu C et al. found that RA Cold patients responded better than RA Heat patients to combination therapy (using nonsteroidal anti-inflammatory drugs (NSAIDs) and DMARDs). Therefore, it is possible that patients with some obvious Heat symptoms may have more active apoptosis in the pathogenesis of RA, and might therefore respond better to biological therapy.

In the present study, we identified three clinical parameter (creatinine, positive RF/anti-CCP, CRP) related to treatment response. Several studies have used different strategies to identify clinical biomarkers that could classify/predict treatment outcome to biological therapies, in which positive RF/anti-CCP is consistently identified as predictive to response to biological therapy, as well as a high level of CRP. In addition, the clinical parameter creatinine is generally used as an indicator of renal function in drug studies. However the association between creatinine and treatment outcome of biological therapies are under-reported, hampering their biological interpretation. Up to now, none of the markers described above can be applied in clinical practice. This might be due to limitations in translating the results to the individual patient. Another reason might be that some (bio)markers are costly to measure and need equipment not readily available in each hospital. In our CM symptom questionnaire, the questions not only covered disease related aspects of rheumatoid arthritis, but also relevant symptoms about patients’ reactions to the external environment, such as food and climate. Collecting this type of information in larger patient groups might increase our understanding of patients’ responses to biological agents, and especially differences in response between patient subgroups. Furthermore, a CM symptom questionnaire is easy to implement in modern hospitals. This study suggests that a combination of Western medicine with Chinese medicine is a feasible road towards improving treatment strategies for rheumatoid arthritis and possibly other chronic diseases.

**Conclusion**

This study for the first time includes Chinese medicine symptoms into an analysis of RA patients’ response to biological therapy. Adding symptom variables (‘preference for warm food’, ‘weak tendon severity’) in addition to clinical measures (creatinine, positive RF/anti-CCP, CRP) does not improve the classification of responders, but it does improve the classification of non-responders. Although this improvement is not significant, we consider that Chinese medicine symptoms are worthwhile to further explore for the differentiation of
Acknowledgements

The authors thank A. Sloeserwij, J. Nijdeken, K. Schrijvers and M. Vianen for collection of clinical data, A. Concepcion and K. Coeleveld for bio-banking and the Society for Rheumatology Research Utrecht (SRU) for including patients. The China Scholarship Council is also gratefully acknowledged (grant to JF).

References

15. van Wietmarschen, H. *et al.* Systems biology guided by Chinese medicine reveals new markers...


36. Lügeling, A. *et al.* Infliximab induces apoptosis in monocytes from patients with chronic active


Supplementary information

Tables

Table S1. Currently used treatments of all subjects and split for responders and non-responders.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>All (n=61)</th>
<th>Non-responders (n=24)</th>
<th>Responders (n=27)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiated bDMARDs, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adalimumab</td>
<td>12 (19.7)</td>
<td>6 (25.0)</td>
<td>6 (16.2)</td>
<td>0.175</td>
</tr>
<tr>
<td>Etanercept</td>
<td>9 (14.8)</td>
<td>2 (8.3)</td>
<td>7 (18.9)</td>
<td></td>
</tr>
<tr>
<td>Golimumab</td>
<td>9 (14.8)</td>
<td>2 (8.3)</td>
<td>7 (18.9)</td>
<td></td>
</tr>
<tr>
<td>Certolizumab</td>
<td>9 (14.8)</td>
<td>6 (25.0)</td>
<td>3 (8.1)</td>
<td></td>
</tr>
<tr>
<td>Abatacept</td>
<td>7 (11.5)</td>
<td>4 (16.7)</td>
<td>3 (8.1)</td>
<td></td>
</tr>
<tr>
<td>Tocilizumab</td>
<td>8 (13.1)</td>
<td>1 (4.2)</td>
<td>7 (18.9)</td>
<td></td>
</tr>
<tr>
<td>Rituximab</td>
<td>7 (11.5)</td>
<td>3 (12.5)</td>
<td>4 (10.8)</td>
<td></td>
</tr>
<tr>
<td>Concomitant DMARDs, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methotrexate</td>
<td>39 (63.9)</td>
<td>14 (58.3)</td>
<td>25 (67.6)</td>
<td>0.418</td>
</tr>
<tr>
<td>Hydroxychloroquine</td>
<td>12 (19.7)</td>
<td>7 (29.2)</td>
<td>5 (13.5)</td>
<td>0.193</td>
</tr>
<tr>
<td>Leflunomide</td>
<td>17 (27.9)</td>
<td>10 (41.7)</td>
<td>7 (9.4)</td>
<td>0.082</td>
</tr>
<tr>
<td>Glucocorticoids</td>
<td>16 (26.2)</td>
<td>8 (33.3)</td>
<td>8 (21.6)</td>
<td>0.377</td>
</tr>
</tbody>
</table>

P values were obtained from Chi-square tests comparing the proportion of subjects using drugs between responders and non-responders.

bDMARDs: biological disease-modifying antirheumatic drugs.
Table S2. Optimal analysis level of baseline data in the multivariate analyses

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Analysis level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptoms</td>
<td>Dry mouth-Severity, Dry mouth-Frequency, Cough with phlegm, Amount of phlegm, Sticky phlegm, Swollen or distended abdomen, Fullness of abdomen, Loose stools, Thirst-Severity, Thirst-Frequency, Chills-Severity, Chills-Frequency, Warm feeling, Aversion to heat, Aversion to cold, Sweating at light exertion-Severity, Sweating at light exertion-Frequency, Preference for cold drinks, Preference for warm drinks, Preference for cold food, Preference for warm food, Heavy feeling of affected parts, Numb and cold skin, Stiff joints, Swollen joints, Red joints, Warm joints, Swellings, Weak limbs, Weak tendons, Pain, Pain aggravates with fog, Pain aggravates at night, Pain aggravates by cold, Pain with redness and swelling, Pain improves by warmth and movement, Soreness, Stabbing pain, Sharp pain, Deep pain, Heavy pain, Dull pain, Continuous light pain, Shortness of breath, Dry throat, Tender lower abdomen, Diarrhea, Cold feeling, Fever with chills</td>
<td>Nominal</td>
</tr>
<tr>
<td>Clinical parameters</td>
<td>C-reactive protein, Hemoglobin, Creatinine, Alanine aminotransferase, Plate count, Leukocyte count, Positive rheumatoid factor/anti-cyclic citrullinated peptide antibody, TNF-alpha inhibitor, Leflunomide, Methotrexate, Hydroxychloroquine, Glucocorticoids, Biological naivety, Gender, Smoke status, Alcohol consumption, Body mass index, Age, Disease duration</td>
<td>Spline ordinal</td>
</tr>
</tbody>
</table>

Table S3. Pairwise comparison of ROC curves

<table>
<thead>
<tr>
<th></th>
<th>Combined model ~</th>
<th>Combined model ~</th>
<th>Clinical model ~</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clinical model</td>
<td>Symptom model</td>
<td>Symptom model</td>
</tr>
<tr>
<td>Difference between areas</td>
<td>0.0045</td>
<td>0.195</td>
<td>0.19</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.0495</td>
<td>0.0852</td>
<td>0.11</td>
</tr>
<tr>
<td>95% Confidence Interval</td>
<td>-0.0926 to 0.102</td>
<td>0.0278 to 0.362</td>
<td>-0.0255 to 0.406</td>
</tr>
<tr>
<td>z statistic</td>
<td>0.0909</td>
<td>2.286</td>
<td>1.728</td>
</tr>
<tr>
<td>Significance level</td>
<td>0.9276</td>
<td>0.0223</td>
<td>0.0839</td>
</tr>
</tbody>
</table>
Methods

Data pre-processing

Before the analysis, the questionnaire data were screened and recoded. Firstly the consistency of answers to the same symptom were checked, such as the estimation of severity and frequency of the symptom. The inconsistent answers were set to missing. Secondly, to avoid unstable results, the frequencies of the categories of the questionnaire variables were checked. The categories with a frequency less than 8 (the square root of the number of subjects) were merged with an adjacent category.\(^5\)

In the clinical dataset, the baseline DAS28 scores are calculated from the variables ‘TJC’, ‘SJC’, ‘VAS’ and ‘ESR’. The classification variable ‘Response’ is derived from the patient's baseline DAS28 and current DAS28 scores, thereby there is a strong correlation among the baseline variables ‘DAS28’, ‘TJC’, ‘SJC’, ‘VAS’, ‘ESR’ and the classification variable ‘Response’. Therefore, these clinical variables were not included in the subsequent analysis.

Categorical principle components analysis (CATPCA) procedures

CATPCA can handle both numeric and categorical variables, as well as non-linear relationships among them. All missing values in the dataset were handled by the option passive, which means all available data are used in the analysis without imputation and deleting data (i.e. without deletion of an entire case when missing on one or more variables).

Variable discretization

The clinical and demographic variables with numeric measurement levels were discretized by grouping, which uniformly recoded variable values into 15 categories.

Specifying optimal analysis options

Specifying an optimal analysis level for each variable is one of the most important features of CATPCA. The optimal scaling level, which is also called the optimal analysis level, can be chosen independently of the measurement level of a variable.

We do not have a priori idea of which level is optimal for each variable, so analyses with different levels were performed to determine which one resulted in the best balance between variance account for (VAF) and degrees of freedom (DF). First of all, a numeric scaling level was used for all variables, which is the most restrictive level (i.e., the least number of DF), and boils down to standard (linear) PCA. Then we compared the VAF per variable and the overall VAF of the linear analysis with results of analyses with more freedom in the
transformations: ordinal analysis levels and nominal (i.e., the most number of DF); for variables with many categories (e.g. BMI, age), ordinal or nominal spline levels were used. The level that resulted in relatively high total VAF and relatively low DF was selected as the optimal one for each variable.

**Variable selection and model building**

The respective contribution of CM symptoms, demographic and clinical variables to the follow-up response was studied by three separate models: a combined model, a symptom model, and a clinical model.

Firstly, for the combined model all variables were analyzed with optimal analysis levels in CATPCA with FC. The variables with loading > 0.30 were selected for the final model. Then, with the same procedure, models including only symptom variables or only clinical data were built separately.

**Leave-one-out cross validation**

In LOOCV, one subject is left out of the model (by specifying the test case as a supplementary case) as a test case. After running the model on the remaining subjects as the training set, the model is applied to the test case. For this test case, a predicted value was computed from the variables’ transformed quantifications and component loadings. In the training model, observed scores for training cases were computed identically. This procedure is repeated N times, each time leaving out another subject and applying the training model to that case. In the meantime, in the training model a mean value (mean 1) of all responders’ observed scores and a mean value (mean 2) of all non-responders’ observed scores was calculated, and we chose the average of mean 1 and mean 2 as the cutoff value. Depending on this cutoff value and the predicted score, a test case will be predicted as a responder or a non-responder.
transformations: ordinal analysis levels and nominal (i.e., the most number of DF); for variables with many categories (e.g. BMI, age), ordinal or nominal spline levels were used. The level that resulted in relatively high total VAF and relatively low DF was selected as the optimal one for each variable.

Variable selection and model building

The respective contribution of CM symptoms, demographic and clinical variables to the follow-up response was studied by three separate models: a combined model, a symptom model, and a clinical model.

Firstly, for the combined model all variables were analyzed with optimal analysis levels in CATPCA with FC. The variables with loading > 0.30 were selected for the final model. Then, with the same procedure, models including only symptom variables or only clinical data were built separately.

Leave-one-out cross validation

In LOOCV, one subject is left out of the model (by specifying the test case as a supplementary case) as a test case. After running the model on the remaining subjects as the training set, the model is applied to the test case. For this test case, a predicted value was computed from the variables' transformed quantifications and component loadings. In the training model, observed scores for training cases were computed identically. This procedure is repeated N times, each time leaving out another subject and applying the training model to that case. In the mean time, in the training model a mean value (mean 1) of all responders' observed scores and a mean value (mean 2) of all non-responders' observed scores was calculated, and we chose the average of mean 1 and mean 2 as the cutoff value. Depending on this cutoff value and the predicted score, a test case will be predicted as a responder or a non-responder.
Chapter 5

Differences between serum lipid profiles of male and female rheumatoid arthritis patients in response to glucocorticoid treatment


Differences between serum lipid profiles of male and female rheumatoid arthritis patients in response to glucocorticoid treatment


*These authors contributed equally