Decision tree analysis for evaluating disease activity in patients with rheumatoid arthritis

Lei Wang[1] , Lifen Zhu[1], Jiahui Jiang[1],
Lijuan Wang[2] and Wanmao Ni[1,3]

Abstract

Objective: Rheumatoid arthritis (RA) is a chronic inflammatory autoimmune disease characterized by inflammatory synovitis. We developed a new disease activity evaluation system using important cytokines to help doctors better evaluate disease activity in patients with RA.

Methods: Flow cytometry was used to detect the levels of seven cytokines. Then, the results were analyzed using an R language decision tree.

Results: The levels of six cytokines, namely interleukin (IL)-2, IL-4, IL-6, IL-10, tumor necrosis factor-α, and interferon-γ, were significantly different between the active disease and remission stages. Decision tree analysis of the six cytokines with statistical significance identified two judgment rules for the remission stage and three judgment rules for the active disease stage.

Conclusion: We proposed the use of the decision tree method to analyze cytokine levels in patients with RA and obtain a more intuitive and objective RA disease activity scoring system. This method revealed the relationships of IL-6 and TNF-α levels with inflammatory characteristics in patients with RA, which can help predict disease activity.

[1]Molecular Diagnosis Laboratory, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, Hangzhou, China
[2]Department of Rheumatology and Immunology, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, Hangzhou, China
[3]Key Laboratory of Tumor Molecular Diagnosis and Individualized Medicine of Zhejiang Province, Zhejiang Provincial People's Hospital, Hangzhou, China

Corresponding authors:
Wanmao Ni and Lijuan Wang, Molecular Diagnosis Laboratory, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, No.158 Shangtang road, Gongshu District, Hangzhou, Zhejiang 310000, China; Department of Rheumatology and Immunology, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, No.158 Shangtang road, Gongshu District, Hangzhou, Zhejiang 310000, China.
Email: wm_ni@163.com; wanglijuan9281@sina.com

Introduction

Rheumatoid arthritis (RA) is a chronic inflammatory autoimmune disease characterized by inflammatory synovitis.[1] Although the pathogenesis is unknown, its close relationship with the overexpression of pro-inflammatory factors is well established. With disease progression, articular cartilage and articular capsules are damaged to varying degrees, eventually leading to joint deformity and a loss of function.[2] Studies illustrated that interleukin (IL)-6, IL-17, and tumor necrosis factor (TNF)-α play important roles in RA.[3,4] The typical histological characteristics of RA are local inflammation and osteoclast maturation and activation, which eventually lead to the destruction of cartilage and bone.[5] Inflammatory cytokines (e.g., TNF-α, IL-6) released by joint cells because of the excessive activation of Th17 cells are among the main factors that destroy immune cells in joints.[6] These inflammatory cytokines (e.g., TNF-α, IL-6, IL-17) induce osteoclast formation by inducing the expression of RANKL.[7] In patients with RA, TNF-α is the key pro-inflammatory factor.[8] Meanwhile, IL-10, as the main anti-inflammatory cytokine, plays an important role in RA, and its influence mainly depends on the release time, extent of release, and cell location.[9]

At present, the methods used to evaluate RA mainly include the 28-joint Disease Activity Score (DAS28), Simplified Disease Activity Index (SDAI), Clinical Disease Activity Index (CDAI), Rheumatoid Arthritis Disease Activity Index, Patient Activity Scale (PAS), PAS II, and Routine Assessment of Patient Index Data. DAS28,[10] combined with the erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) level,[11] is presently used to define RA in the clinic. There is evidence that DAS28 may not provide a reliable account of disease remission.[12] Significant differences and heterogeneity between patients and doctors in the assessment of disease activity often arise, which affects the formulation of treatment plans, clinical expectations, remission, and the doctor–patient relationship.[13] These assessment methods are based on doctors' subjective or empirical judgment of the disease. On the contrary, the judgment rules of disease activity based on the cytokine results of decision tree analysis are relatively simple and objective. Therefore, we aimed to develop a new disease activity evaluation system using important cytokines to reflect the actual state of the body more intuitively and objectively and potentially help doctors better evaluate disease activity in patients with RA. In this study, data on sex, age, and disease activity in patients were collected and described to provide clues for the judgment rules of disease activity based on a decision tree analysis of cytokines.

Materials and methods

Human subjects

In this observational cross-sectional study, patients with RA who were admitted to the Department of Rheumatology and Immunology of Zhejiang Provincial People's Hospital between March 2016 and June 2018 were enrolled. All patients met the new RA classification standard proposed by the American College of Rheumatology (ACR) and European

Alliance of Associations for Rheumatology (EULAR) in 2009. This research was also approved by the ethical review board of our hospital. Patients with other diseases, such as other connective tissue diseases, infectious diseases, or tumors, were excluded. Because our specimens were collected at the time of admission, treatment after admission had little effect on our results. Treatment was performed according to ACR/EULAR guidelines, with 67.5% of patients receiving disease-modifying anti-rheumatic drugs, 27.5% receiving biological treatment, and 5% receiving corticosteroids.

According to the DAS28 score, the inpatients were divided into remission and active disease groups. All samples were collected in a manner consistent with Chinese law. The requirement for informed consent was waived because we used discarded blood samples, and the study was approved by the Ethics Committee of Zhejiang Provincial People's Hospital (approval no. 2019KY232).

Cytokine sample preparation

Two milliliters of fasting blood samples were collected in the morning from the elbow veins of subjects in a sitting position. The samples were centrifuged at $500 \times g$ for 5 minutes in a separation gel accelerating tube, and serum was collected for cytokine detection.

Reagents and methods

The BD^{TM} Cytometric Bead Array Human Th1/Th2/Th17 Cytokine Kit was purchased from BD (Franklin Lakes, NJ, USA) and used as per the manufacturer's instructions.

Cytokines were detected using a Navios flow cytometer (Beckman Coulter, Brea, CA, USA). According to the EULAR recommendations, DAS28 was calculated as follows:

$DAS28 = 0.56 \times$ the number of tender joints among 28 joints assessed $+ 0.28 \times$ the number of swollen joints among 28 joints $+ 0.7 \times \ln(ESR) + 0.014 \times$ visual analog scare. $DAS28 < 2.6$ was defined as remission, and $DAS28 \geq 2.6$ indicated active disease.

Statistical analysis

SPSS Statistics 21.0 (IBM Corp., Armonk, NY, USA) was used to calculate and analyze the data. Because the data did not conform to a normal distribution, the Mann–Whitney U test was used for statistical analysis. The measurement data were expressed as the mean ± standard deviation and analyzed by a nonparametric test. $P < 0.05$ denoted statistical significance. Cytokines with significantly different levels between the remission and active disease groups were analyzed using a decision tree to calculate their sensitivity and specificity. Recursive PARTitioning (rpart, https://cran.r-project.org/web/packages/rpart/rpart.pdf) was used to build classification or regression models of an extremely general structure using a two-stage procedure, and the resulting models can be represented as binary trees.

According to the result of the Mann–Whitney U test, our code for constructing the decision tree was as follows:

fit<-rpart(State ~ IL.2 + IL.4 + IL.6 + IL.10 + TNF.α + IFN.γ, method = "class", data = d)rpart.plot(fit, extra = 106, type=3)

Results

Differences in cytokine levels between the active disease and remission groups

In total, 263 patients with RA who were admitted to the Department of Rheumatology and Immunology of Zhejiang Provincial People's Hospital from March 2016 to June 2018 were enrolled. The patients included 60 men

and 203 women with an average age of 56 ± 13 years. The mean duration of disease was 8.47 ± 33 years. The remission group included 77 patients, including 8 men and 69 women with an average age of 54 ± 13 years. Meanwhile, the active disease group included 186 patients (52 men and 134 women) with an average age of 56 ± 13 years. There was no significant difference in the age between the two groups. Cytokine levels did not differ between men and women (Table 1). First, we compared the serum levels of IL-2, IL-4, IL-6, IL-10, IL-17A, TNF-α, and IFN-γ between patients with RA and a cohort of healthy controls (34 men and 66 women; mean age, 39 ± 12 years). The levels of all seven cytokines were significantly higher in patients with RA ($P < 0.05$, Table 2). Meanwhile, the levels of all cytokines except IL-17A were significantly higher in the active disease group than in the remission group (all $P < 0.05$, Table 3).

Decision tree analysis

The enrolled patients with RA were analyzed *via* decision tree analysis. In total, 80% of patients (210 patients) were randomly assigned to the training group, and the remaining 20% of patients (53 patients) comprised the test group. The training group included 60 patients in remission and 150 patients with active disease, whereas the test group included 17 patients in remission and 36 patients with active disease. Decision tree analysis was performed using cytokines with significantly different levels between the active disease and remission groups (IL-2, IL-4, IL-6, IL-10, TNF-α, and IFN-γ) using the rpart package of R language (www.r-package.org) based on the data of patients in the training group. The decision tree is presented in Figure 1.

Two rules were used for judging the remission phase:

Table 1. Cytokine expression in men and women with rheumatoid arthritis.

Cytokine (n = 263)	Women (n = 203)	Men (n = 263)	P
IL-2	11.45 ± 19.26	16.20 ± 57.03	0.751
IL-4	7.52 ± 19.13	14.35 ± 59.13	0.395
IL-6	45.60 ± 91.94	53.93 ± 62.75	0.078
IL-10	8.22 ± 12.77	11.52 ± 38.84	0.098
IL-17A	6.00 ± 20.45	22.73 ± 146.63	0.328
TNF-α	23.64 ± 40.90	34.44 ± 86.48	0.107
IFN-γ	5.71 ± 16.53	12.86 ± 73.29	0.131

Data are presented as ng/L.
IL, interleukin; TNF, tumor necrosis factor; IFN, interferon.

Table 2. Cytokine expression in patients with RA and normal controls.

Cytokine (n = 363)	Normal (n = 100)	RA (n = 263)	P
IL-2	1.31 ± 0.90	12.53 ± 31.97	<0.001
IL-4	1.36 ± 0.87	9.08 ± 32.83	<0.001
IL-6	2.20 ± 1.46	47.50 ± 86.11	<0.001
IL-10	2.34 ± 2.75	8.97 ± 21.62	<0.001
IL-17A	1.87 ± 1.76	9.82 ± 72.21	0.040
TNF-α	1.86 ± 1.39	26.10 ± 54.72	<0.001
IFN-γ	2.13 ± 2.24	7.34 ± 37.81	0.003

Data are presented as ng/L.
RA, rheumatoid arthritis; IL, interleukin; TNF, tumor necrosis factor; IFN, interferon.

Table 3. Cytokine expression in patients with active disease and those in remission.

Cytokine (n = 263)	Active disease group (n = 186)	Remission group (n = 77)	P
IL-2	16.87 ± 37.15	2.02 ± 2.63	<0.001
IL-4	12.11 ± 38.62	1.75 ± 2.64	<0.001
IL-6	61.50 ± 97.06	13.67 ± 31.59	<0.001
IL-10	11.53 ± 25.23	2.78 ± 2.67	<0.001
IL-17A	13.10 ± 85.71	1.87 ± 1.59	0.252
TNF-α	35.10 ± 62.71	4.38 ± 8.61	<0.001
IFN-γ	9.42 ± 44.80	2.33 ± 2.53	0.033

Data are presented as ng/L.
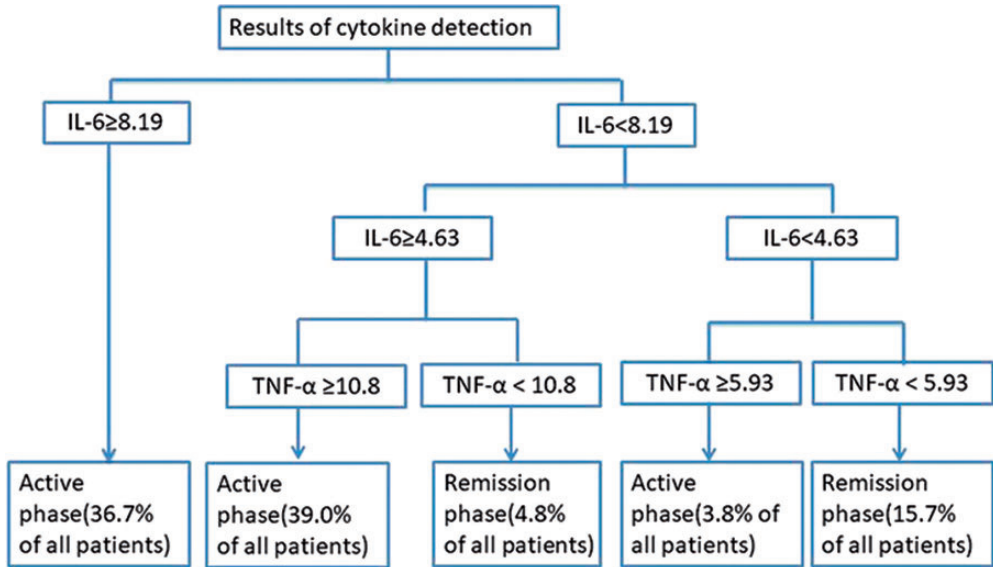IL, interleukin; TNF, tumor necrosis factor; IFN, interferon.

Figure 1. Through R language decision tree analysis, five judgment rules were identified, including two rules for judging remission and three rules for judging active disease. Satisfaction of any single rule was sufficient for identifying active disease or remission.

Table 4. Error matrix depicting accuracy of the developed decision tree for discriminating patients with active disease from those in remission.

Predicted	Actual		
	Active phase	Remission phase	
Active phase	34	5	PPV: 87.18%
Remission phase	2	12	NPV: 85.71%
	Sensitivity: 94.44%	Specificity: 70.59%	

PPV, positive predictive value; NPV, negative predictive value.

1. $4.63 \leq$ IL-6 < 8.19 and TNF-α < 10.8
2. IL-6 < 4.63 and TNF-α < 5.93

Remission was judged if patients met either of these rules.

The three rules were used for judging the active phase:

1. IL-6 ≥ 8.19
2. IL-6 ≥ 4.63 and TNF-α ≥ 10.8
3. IL-6 < 4.63 and TNF-α ≥ 5.93

Active disease was judged if patients met any of these rules.

According to these rules, the test group data were predicted, and the accuracy rate was 86.79%. In the test group, the predictive value of the decision rules was confirmed. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were 94.44%, 70.59%, 87.18%, and 85.71%, respectively (Table 4).

Discussion

Several studies have been conducted globally to understand RA, but the etiology of RA remains unclear. It has been demonstrated that abnormal cytokine expression plays an important role in RA.[14,15] Currently, the most widely used method for the clinical evaluation of RA disease activity is DAS28. Other RA assessments include CDAI, SDAI, and PAS. These scores reflect patient's own assessment of the disease. Correct evaluation of the disease by patients is affected by many factors, such as the extent of their understanding of the disease and their concerns about health. The assessment is also influenced by the existence of anxiety, depression, and other emotional factors, as well as the presence of individual differences with certain subjectivity. This study used the decision tree method to analyze the cytokine levels of patients with RA and identify some rules to judge disease activity with the exclusion of subjective factors.

Decision tree analysis is a popular supervised learning method based on simple and easy logic that consists of a prediction model, generating corresponding rules to classify the samples.[16] Decision tree analysis avoids the inaccurate assessment of disease activity caused by subjective factors.

The body can produce a variety of cytokines during the inflammatory process. According to our findings (Table 3), there were significant differences in IL-2, IL-4, IL-6, IL-10, TNF-α, and IFN-γ levels between patients with active disease and those in remission, whereas IL-17A levels did not differ between the groups. For our study, we randomly divided patients into the training (80%) and test groups (20%) and then selected IL-2, IL-4, IL-6, IL-10, TNF-α, and IFN-γ for the cytokine dataset. According to the weight of the variables, IL-6 and TNF-α, which had the strongest discriminative ability, were selected as decision variables to construct a decision tree with three-level nodes. According to these two decision variables, five rules for evaluating RA disease activity were established. We then predicted disease activity in the test group using these five rules and achieved good accuracy (86.79%), sensitivity (94.44%), and specificity (70.59%), and PPV and NPV both exceeded 85%. In addition, the F1 score was as high as 90.67%. For our study, decision tree analysis was used to judge disease activity only using a common detection index of cytokines in the laboratory. The strategy can both objectively evaluate disease activity in each patient and avoid subjectivity. At the same time, it can help improve the diagnosis and treatment of RA and reduce the economic expenditures of patients.

For the first time, we used the decision tree method to analyze cytokine levels in patients with RA and developed a more intuitive and objective RA disease activity scoring system. This method revealed relationships of IL-6 and TNF-α levels with inflammatory characteristics in patients with RA. This method can thus be used to predict disease activity in patients with RA. For future studies, we are planning to further verify the rules by increasing the sample size. To date, a variety of TNF-α antagonists[17–19] and IL-6 inhibitors[20,21] have been widely used in the clinical treatment of RA, and we would like to combine our method with clinical drug efficacy monitoring for use in drug efficacy analysis.

ORCID iD

Lei Wang https://orcid.org/0000-0003-0008-3224

References

1. Harris ED Jr. Rheumatoid arthritis: Pathophysiology and implications for therapy. *N Engl J Med* 1990; 322: 1277–1289.
2. Lafeber FPJG and Van Der Laan WH. Progression of joint damage despite control of inflammation in rheumatoid arthritis: a role for cartilage damage driven synovial fibroblast activity. *Ann Rheum Dis* 2012; 71: 793–795.
3. Paradowska-Gorycka A, Grzybowska-Kowalczyk A, Wojtecka-Lukasik E, et al. IL-23 in the Pathogenesis of Rheumatoid Arthritis. *Scand J Immunol* 2010; 71: 134–145.
4. De Brouwer SJM, Middendorp HV, Kraaimaat FW, et al. Immune responses to stress after stress management training in patients with rheumatoid arthritis. *Arthritis Res Ther* 2013; 15: R200.
5. McInnes IB and Schett G. Cytokines in the pathogenesis of rheumatoid arthritis. *Nat Rev Immunol* 2007; 7: 429–442.
6. Park YH, Kim HJ, Lee K, et al. Combination of gp130-targeting and TNF-targeting small molecules in alleviating arthritis through the down-regulation of Th17 differentiation and osteoclastogenesis. *Biochem Biophys Res Commun* 2019; 522: 1030–1036.
7. Neumann E, Lefèvre S, Zimmermann B, et al. Rheumatoid arthritis progression mediated by activated synovial fibroblasts. *Trends Mol Med* 2010; 16: 458–468.
8. Brennan FM, Maini RN and Feldmann M. TNFα - A pivotal role in rheumatoid arthritis?. *Rheumatology* 1992; 31: 293–298.
9. Gambhir D, Lawrence A, Aggarwal A, et al. Association of tumor necrosis factor alpha and IL-10 promoter polymorphisms with rheumatoid arthritis in North Indian population. *Rheumatol Int* 2010; 30: 1211–1217.
10. Van Der Heijde DM, Van 't Hof MA, Van Riel PL, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis* 1990; 49: 916–920.
11. Avouac J, Meune C, Chenevier-Gobeaux C, et al. Inflammation and disease activity are associated with high circulating cardiac markers in rheumatoid arthritis independently of traditional cardiovascular risk factors. *J Rheumatol* 2014; 41: 248–255.
12. Smolen JS, Basset SC, Boers M, et al. Clinical trials of new drugs for the treatment of rheumatoid arthritis: focus on early disease. *Ann Rheum Dis* 2016; 75: 1268–1271.
13. Sacristán JA, Dilla T, Díaz-Cerezo S, et al. Patient-physician discrepancy in the perception of immune-mediated inflammatory diseases: rheumatoid arthritis, psoriatic arthritis and psoriasis. A qualitative systematic review of the literature. *PLoS One* 2020; 15: e0234705.
14. McInnes IB, Buckley CD and Isaacs JD. Cytokines in rheumatoid arthritis — shaping the immunological landscape. *Nat Rev Rheumatol* 2016; 12: 63–68.
15. Feldmann M, Brennan FM and Maini RN. Role of cytokines in rheumatoid arthritis. *Annu Rev Immunol* 1996; 14: 397–440.
16. Ture M, Tokatli F and Kurt I. Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients. *Expert Syst Appl* 2009; 36: 2017–2026.
17. Elliot MJ, Maini RN, Feldmann M, et al. Treatment of rheumatoid arthritis with chimeric monoclonal antibodies to tumor

necrosis factor alpha. *Arthritis Rheum* 1993; 36: 1681–1690.

18. Breedveld FC, Weisman MH, Kavanaugh AF, et al. The PREMIER study: A multicentre, randomised double blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone. *Arthritis Rheum* 2006; 54: 26–37.

19. Smolen JS, Kay J, Doyle MK, et al. Golimumab in patients with active rheumatoid arthritis after treatment with tumour necrosis factor α inhibitors (GO-AFTER study): a multicentre, randomised, double-blind, placebo-controlled, phase III trial. *Lancet* 2009; 374: 210–221.

20. Choy E, Isenberg DA, Garrood T, et al. Therapeutic benefit of blocking interleukin-6 activity with an anti-interleukin-6 receptor monoclonal antibody in rheumatoid arthritis: a randomized, double-blind, placebo-controlled, dose-escalation trial. *Arthritis Rheum* 2014; 46: 3143–3150.

21. Nishimoto N, Yoshizaki K, Miyasaka N, et al. Treatment of rheumatoid arthritis with humanized anti-interleukin-6 receptor antibody: a multicenter, double-blind, placebo-controlled trial. *Arthritis Rheum* 2014; 50: 1761–1769.