Universiteit
Leiden
The Netherlands

Predictive factors for outcome of rheumatoid arthritis

Linden, M.P.M. van der

Citation
Linden, M. P. M. van der. (2011, September 15). *Predictive factors for outcome of rheumatoid arthritis*. Retrieved from https://hdl.handle.net/1887/17836

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/17836

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

CHAPTER 2

Comparison of methodology to analyze progression of joint destruction in rheumatoid arthritis

R. Knevel[1]
R. Tsonaka[2]
S. le Cessie[2,3]
M.P.M. van der Linden[1]
T.W.J. Huizinga[1]
D.M.F.M. van der Heijde[1]
J.J. Houwing-Duistermaat[2]
A.H.M. van der Helm-van Mil[1]

[1] *Department of Rheumatology, Leiden University Medical Center, Leiden, the Netherlands*
[2] *Department of Medical Statistics, Leiden University Medical Center, Leiden, the Netherlands*
[3] *Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands*

Submitted

ABSTRACT

Background

The field of genetics is reaching phenotypic disease aspects. Within rheumatoid arthritis (RA), progression of joint destruction is an important phenotypic feature. Genetic factors often have small effect sizes, making avoidance of phenotypic misclassification and discerning true effects from noise challenging. Assembling radiological measurements repeatedly in time harbors a smaller risk of misclassification than single measurements. Given serial measurements, different methods of analysis can be applied. This study evaluates different statistical methodology to analyze longitudinal data and its effect on the power of such a study.

Methods

Kruskal-Wallis, Linear Regression and Repeated Measurements Analysis (RMA) were studied, both cross-sectionally (testing for differences in joint destruction at individual time points) and longitudinally (testing for differences in progression rates). Of these tests, only RMA takes advantage of within-patient correlations in serial radiological measurements. Data of 602 early RA patients included in an inception cohort with yearly radiographs and 7-years follow-up were assessed. Genetic data of HLA-DRB1 Shared-Epitope alleles and rs675520 (*TNFAIP3-OLIG3*) were used as example.

Results

From all methods studied, cross-sectional and longitudinal RMA were most powerful. For example analyses using longitudinal RMA in the current data set yielded powers >95%, even in presence of missing radiographs. In particular in the presence of small effect sizes RMA was more powerful than linear regression. The preciseness increased with a higher number of available measurements per patient.

Conclusion

A repeated measurement analysis on subsequent radiographs provides the most powerful methodology to analyze longitudinal data.

INTRODUCTION

In medicine more than 600 genome wide association studies have been published; often revealing inconsistent findings.[1] Now the field of genetics is moving from qualitative traits (disease yes/no) to phenotypic disease aspects and disease outcomes, which are often quantitative traits. Correct determination of the phenotype is of most importance here. Within rheumatoid arthritis (RA), progression of joint destruction is a relevant outcome measure, reflecting the cumulative burden of inflammation over time. The severity of joint destruction is highly variable between patients. Thus far, little is known about the pathophysiology of this difference. In addition, several clinical and serological risk factors for a severe rate of joint destruction have been identified, but the variation explained by these factors is low (R^2 0.36).[2-4] Prediction models based on these variables could classify only ~50% of RA patients.[2,5,6] In order to increase the understanding of the mechanisms underlying joint destruction, additional risk factors need to be identified. Thus far, few identified genetic factors for joint destruction are replicated. The absence of replication can have several causes. Obviously, it may be due to false-positive results in the initial study. Secondly, the replication study could have been underpowered. It is challenging to obtain long-term radiological data of a large number of patients. Finally, differences between studies may occur when different radiological measures are studied or when different methods of analyses are applied. Since the effect sizes of genetic markers in complex diseases are often moderate to small, both sensitive measurements of joint destruction and powerful methods of analysis are necessary to prevent false negative findings.

It is discussed elsewhere that the use of a continuous method to measure the degree of joint damage is more sensitive and discriminative than usage of categorical measures such as the presence of erosions.[7] In addition it has been shown that serial measures in time per patient give a more accurate and precise estimation of the rate of joint destruction compared to single measurements. Therefore, whenever possible, RA patients are preferably studied prospectively and have radiographs made at subsequent time-points.[7] In the presence of serial quantitative measurements, different statistical methods for analysis are available and applied. The level of joint destruction can be compared between groups at individual time-points, with and without taking radiological data on other time-points into consideration. Alternatively, the progression over all time-points can be compared in one test. An additional challenge in analyzing longitudinal radiological data is how to deal with missing radiographs. Therefore, we aimed to compare currently used statistical methodology to analyze continuous data on joint destruction over time. The main outcome measure evaluated was the power. We therefore evaluated the power of analyses performed with different statistical methods on the same patients and genetic data. First the power of these methods was evaluated using data of genetic variants known to associate with joint destruction. Second, we compared the ability of the different methods to deal with missing radiological data, as well as the effect of the number of available radiographs on the power of the study.

PATIENTS AND METHODS

Patients

Radiological data were used of 602 RA patients (according to the 1987 ACR-criteria) that were included in the Leiden Early Arthritis Clinic cohort (EAC) in 1993-2006.[8] Median symptom duration at inclusion was 0.36 years. At baseline, the mean age was 56.1±15.8 years, 78% was female and 54% was ACPA-positive. Yearly follow-up data over 7-years was used. Radiographs of hands and feet were scored chronologically according to the Sharp-van der Heijde method (SHS) by an experienced reader.[9-11] 409 radiographs belonging to 60 randomly selected RA patients were rescored. The intraclass correlation coefficient was 0.91 for all scored radiographs, and 0.97 for the radiographic progression rate. Treatment strategies changed in time.[8,12] Patients included in 1993-1995 were initially treated with analgesics and subsequently with chloroquine or salazopyrin. From 1996-1998 chloroquine or salazopyrin was promptly started. From 1999-2006 patients were readily treated with methotrexate or salazopyrin. Twenty-eight of the 602 patients received anti-TNF treatment somewhere during the seven follow-up years. The frequency of anti-TNF users was equally distributed between periods of inclusion (3.3%, 4.7% and 4.7% respectively).

Methods to analyze joint destruction

The HLA-DRB1 Shared-Epitope (SE) alleles and rs675520 (*TNFAIP3-OLIG3)* are associated with joint destruction.[13-16] To compare different statistical methods, these two genetic variants were studied as example (Figure 1). Three statistical methods were studied, representing the major methods for analyses. Other not-applied methods are more or less similar to the methods applied here.

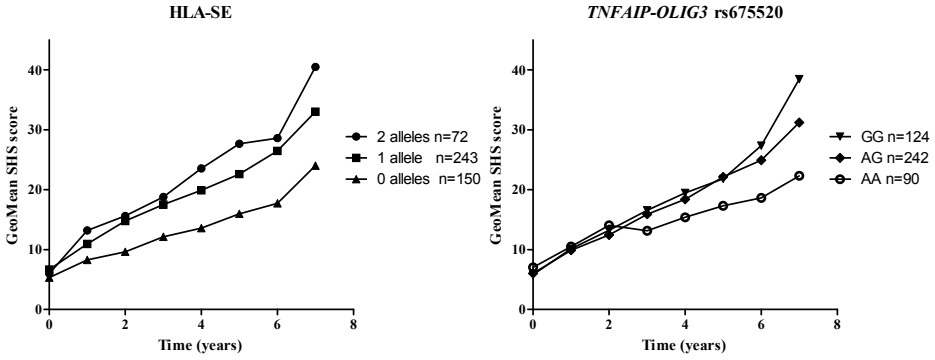

Figure 1. Sharp-van der Heijde scores during 7-years of follow-up for RA patients with 0, 1 or 2 HLA-SE alleles and with absence, presence or double presence of the minor allele of *TNFAIP-OLIG3* rs675520. Presented are the geometric means of the SHS

Cross-sectional methods studied, comparing destruction levels at individual time-points, were the Kruskal-Wallis test, linear regression analysis (LR_{cs}) and repeated measurement analysis (RMA_{cs}). The Kruskal-Wallis and LR_{cs} was performed on each time-point with SHS score as dependent variable ignoring the data of other time-points. For RMA_{cs}, a multivariate normal regression analysis was used with time as categorical variable.[17] The RMA_{cs} tested differences between SHS levels at each time-point taking radiological data on previous time-points into consideration.

The evaluated longitudinal methods, testing for differences in progression rates over time, were Kruskal-Wallis, longitudinal linear regression analysis (LR_{long}) and repeated measurements analysis (RMA_{long}). Here the Kruskal-Wallis test compared subtractions of SHS between baselines and the 7-years time-points and therefore data of only two measurements could be used. LR_{long} compared regression coefficients which are based on all available measurements, assuming them to be independent. RMA_{long} evaluated the progression rates over time considering the correlation between the measurements at all time-points within one subject. In order to have optimal comparisons of the tests, no adjustments were made in the LRs, and RMAs.

Since SHS were positively skewed, radiological scores were log-transformed to approximate normal distribution before performing any of the LRs and the RMAs. Analyses were performed using SPSS version 16.0 (SPSS Inc., Chicago, IL).[18]

Repeated measurement analysis

Detailed information on the used RMAs, a multivariate normal regression analysis, is provided in Box I, supplementary data. This analysis uses all available radiological measurements and has great flexibility to model time effects. It takes advantage of within patients' correlations and can handle missing data provided that the reason for missingness can be determined from the observed data (an assumption called missingness at random).[17,19]

The within-patient correlation of serial measurements is quantified by a covariance matrix. To determine the best-fitting covariance matrix the matrices available in SPSS were considered, using the Akaike information criteria as measure of goodness of fit. The heterogeneous first order autoregressive (ARH1) matrix was our final choice. It assumes a stronger correlation for measurements taken in a short period than taken over a longer period in time.

Power of different methods

It was hypothesized that the different methods will yield differences in power. To study this, the power to detect an association between the two genetic variants and joint destruction over 7-years was determined; both for the cross-sectional and longitudinal methods. For the Kruskal-Wallis, Quanto version 1.2.4[20] was used on the present data assuming that the effect of HLA-SE and rs675520 increased with respectively 1.3 and 1.2 times per year. The power of LR and RMA were computed by simulating the RMA model. The baseline characteristics of the patients, the sample size and parameter values were sampled such that they correspond to the original EAC

data. In order to also study the impact of missingness to the power, the percentage of missing radiographs was varied from 0 to almost 90% for the last visit. For the remaining visits, missingness was created with the same percentage as in the original dataset. More detailed description on the power analyses are described in the supplement. Power analyses were performed using R statistical software.[21]

Effect of number of radiological measurements

The number of measurements available per subject can differ between different study designs. Here we studied the influence of the number of measurements per subject on the preciseness of the estimation expressed as the 95% confidence interval (95%CI) of the effect size. To this end 107 patients with complete yearly follow-up over 7-years were studied. By simulation an increasing number of radiographs were left out between baseline and the 7-years time-point. In this way analyses were repeated with a lower number of radiological measurements per patient. Analyses were done on HLA-SE and joint destruction analyzed with both LR_{long} and RMA_{long}.

Missing radiological data in relation to different methods

The presence of missing data in longitudinal cohort studies is inevitable. Exclusion of the patients with missing data will generate bias in case missingness is related to the outcome of interest.[22] From the methods evaluated here, RMA is able to deal with missing data provided that the missingness is 'at random' or 'completely at random' and that the correlation structure (expressed by the covariance matrix) of the patients with missing data is comparable to that of patients with complete radiological data. Therefore the characteristics of missing radiological data in the studied cohort were evaluated.

RESULTS

Methods to analyze joint destruction

The cross-sectional and longitudinal methods of analysis were compared using radiological data of RA patients with different numbers of HLA-SE alleles. The various methods all resulted in significant outcomes at individual time-points (cross-sectional analyses) as well as on progression over time (longitudinal methods). The width of the 95%CI differed between the methods (see Table I).

Power and preciseness of different methods

The power to detect an association of HLA-SE with levels of joint destruction at the individual time-points from baseline till 7-years with Kruskal-Wallis in the present dataset were 0.52, 0.37, 0.40, 0.34, 0.36, 0.41, 0.48, 0.47. For rs675520, the power were 0.53, 0.31, 0.29, 0.22, 0.21, 0.19, 0.20, 0.18 from baseline till 7-years. Comparing differences in SHS between baseline and 7-years with Kruskal-Wallis had a power of 0.92 and 0.25 for HLA-SE and rs675520 respectively. The effect of missingness on the power of LR and RMA, both cross-sectional and longitudinal, are

illustrated by a simulation for different frequencies of missingness in Figure 2. The power to detect a difference in the cross-sectional analyses of HLA-SE groups was approximately 100% if the data at 7-years were complete. With increasing missingness the power of LR_{cs} diminished to <80%, whereas the power of RMA_{cs} remained >95%, even in case of a large percentage of missingness (Figure 2A). Although the power to detect a difference was lower in the analysis of rs675520, again it was observed that the power of RMA_{cs} remained higher than of LR_{cs}. Also for the longitudinal analyses, RMA had a higher power compared to LR (Figure 2B), for both HLA-SE and rs675520.

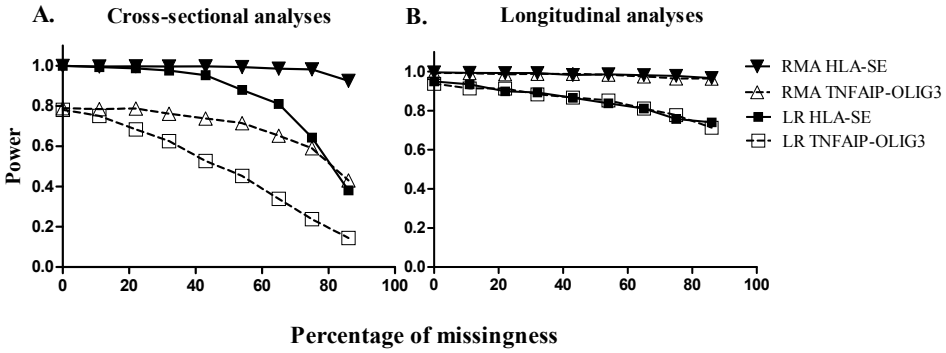

Figure 2. Power to detect differences in joint destruction with (A) cross-sectional and (B) longitudinal methods (LR and RMA) for different percentages of missing radiographs at the last time-point. Depicted is the power (y-axis) to detect an association between two different genetic variants, HLA-SE and *TNFAIP-OLIG3* (rs675520) and the rate of joint destruction in the present RA patients at the 7-years time-point.[24] The power was calculated (A) cross-sectional with linear regression (LR) and repeated measurement analysis (RMA) *at* 7-years and (B) longitudinally with LR and RMA *over* 7-years with different percentages of missing radiographs at the 7-year time-point (x-axis)

Effect of number of radiological measurements

With an increasing number of available radiographs the 95%CI of the estimation of the progression rate decreased, indicating a more precise estimation in the presence of more measurements per subject (see Figure 3).

Missing radiological data

Three major causes were identified that together accounted for >90% of all missing follow-up data: sustained DMARD-free remission (n=64), death (n=74), and not having complete follow-up data because of recent inclusion. Patients without sustained DMARD-free remission had a 2.35 (95%CI 1.83-3.19 p<0.001, RMA_{long}) times larger increase in SHS per 7-years. Patients had a constant 2.09 (95%CI 1.65-2.65 p<0.001, RMA_{long}) times larger joint damage over 7-years compared to those who stayed alive. For both reasons of missing data the missingness related to the outcome (missingness at random).

Table I: Risk estimates and p-values resulting from different statistical analyses associating the number of HLA-SE alleles with Sharp-van der Heijde scores during 7-years

		β (95% CI) *P-value*							
		Baseline	1	2	3	4	5	6	7
Cross-sectional analyses	Kruskal-Wallis#	N/A *0.244*	N/A *0.006*	N/A *0.002*	N/A *0.011*	N/A *0.007*	N/A *0.019*	N/A *0.037*	N/A *0.047*
	Linear regression † (LR_{cs})	1.10 (0.95-1.26) *0.20*	1.27 (1.1-1.48) *0.002*	1.32 (1.12-1.56) *0.001*	1.29 (1.08-1.54) *0.005*	1.34 (1.11-1.61) *0.003*	1.33 (1.10-1.63) *0.006*	1.29 (1.04-1.62) *0.02*	1.31 (1.01-1.66) *0.03*
	Repeated Measures Analysis (RMA_{cs}) †‡	N/A∞	1.25 (1.1-1.42) *0.004*	1.25 (1.11-1.40) *<0.001*	1.27 (1.15-1.42) *<0.001*	1.25 (1.14-1.37) *<0.001*	1.22 (1.13-1.32) *<0.001*	1.19 (1.12-1.25) *<0.001*	1.23 (1.07-1.41) *<0.001*
Longitudinal analyses	Kruskal-Wallis	N/A *0.008*							
	Linear regression* (LR_{long})	1.16 (0.94-1.43) *0.18*							
	Repeated measures Analysis* (RMA_{long})	1.25 (1.09-1.43) *0.002*							

The cross-sectional analyses compare the differences in SHS level at each time-point separately. The longitudinal analyses compare the differences in progression rates over time

\# This analysis does not result in a risk estimate.

† The β of the LR_{cs} and RMA_{cs} indicates the relative increase in SHS per risk allele at the individual time-points.

∞ Baseline was the reference in this analysis, therefore no risk estimate and p-value are present.

‡ The β of the RMA_{cs} indicates the relative increase in SHS per risk allele at the individual time-points.

* The β of the LR_{long} and RMA_{long} indicates the relative increase in SHS-progression per risk allele for the progression *over* 7-years.

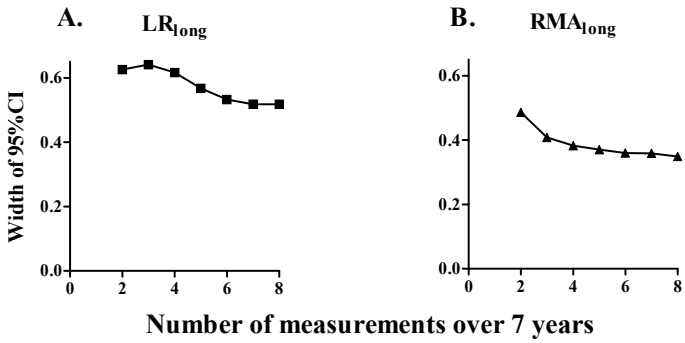

Figure 3. Width of 95% confidence interval (95%CI) for different number of measurement over 7-years of follow-up for (A) Linear regression analysis and (B) Repeated measurement analysis. Depicted is the 95%CI width (y-axis) of the analyses of the association between HLA-SE and joint destruction. The analysis was performed on 107 patients with complete follow-up yearly over 7-years. First only baseline and 7-years data was used, additional time-points were added to test the effect of the number of measurement used over the same time-period. A) The width of the 95%CI analyzing HLA-SE with LR_{long} demonstrates the advantage of adding more measurements to the analyses. B) The width of the 95%CI analyzing HLA-SE with RMA_{long} demonstrates the advantage of more measurements plus taking the correlation into account

DISCUSSION

The field of genetics is moving from disease susceptibility studies to studies addressing disease outcomes. Since genetic risk factors generally have small effect sizes, it is crucial to measure the outcome sensitively and to apply powerful statistical methodology. Given the presence of repeated radiologic measurements in time, different statistical tests can be used. We aimed to derive optimal statistical methodology. We considered commonly used methods but did not intend to give a complete overview of all possible statistical methods. We observed that, among the methods tested, a RMA is most powerful and least susceptible to bias. The increased power is the result of taking advantage of the high within-patient correlation in repeated measurements. We also observed that effect estimates were more precise in the presence of a higher number of measurements, an effect which is not specific for RMA. A RMA can compare absolute differences in SHS levels at a single time-point and rates of progression over time; the choice between these two may depend on whether one is interested in identifying associations with the level of joint destruction at a specific time-point or in identifying associations with the speed of progression of radiological joint damage over time.

We considered commonly used methods but did not intend to give a complete overview of all possible statistical methods. Advantages and disadvantages of the methods studied are presented in Table II. Advantageous of RMA is that all patients, also those who had missing radiographs, are included. This is done assuming that missing radiological scores can be estimated using available measurements and complete datasets of patients with similar characteristics, a situation called 'missingness at random'. Identified causes for missing radiographs in the present study

Table II: Advantages and disadvantages of the tested statistical methods to identify risk factors for joint destruction in RA

		Description	Can adjust for interfering factors	Includes serial measurements in one analysis	Includes patients with missing radiographs	Reliably deals with missing data	Uses within patient correlation (covariance matrix)
Cross-sectional analyses	Kruskal-Wallis	Tests the difference in rank of a continuous variable among the categories of a second variable	–	–	–	–	–
	Linear regression (LR_{cs})	Tests mean differences of continuous variable among the categories of a second variable	+	–	–	–	–
	Repeated Measures Analysis (RMA_{cs})	Idem previous	+	+	+	+	+
Longitudinal analyses	Kruskal-Wallis (progression between baseline & 7-year)	Tests the difference in rank of a continuous variable among the categories of a second variable	–	–/+	–	–	–
	Linear regression (LR_{long})	Tests differences of mean evolution of continuous variable among the categories of a second variable	+	+	+	–	–
	Repeated measures Analysis (RMA_{long})	Idem previous	+	+	+	+	+

– No, it does not

+ Yes, it does

were assumed to be missing at random, a requirement for adequate handling of missing data by the RMA. The RMA takes into account the uncertainty of the estimation for patients with missing radiographs. In other words, patients with complete datasets are weighted more heavily in the analysis than patients with missing radiographs. The RMA is the only studied method that did not exclude patient with missing data, which prevents certain bias.

Another, simple and frequently used method to deal with missing radiographs is a completers only analysis. Here all patients with missing observations are excluded. This is used when comparing differences in SHS between 7-years and baseline with Kruskal-Wallis tests and can lead to conflicting results at different time-points. An alternative is the last-observation-carried forward approach; this uses the last observation for every subsequent missing. Both methods can create bias since we observed that patients that are more inclined to have missing radiographs have relatively severe or relative mild joint destruction.[23]

The longitudinal LR studied compared the regression coefficients of SHS with time between groups. An advantage of LR above Kruskal-Wallis is that it gives an effect size and allows adjustment of correction variables. A drawback of LR is that it ignores the correlation between serial measurements; accounting for this would have resulted in a smaller standard error and therefore a more sensitive analysis. An alternative LR analysis over time is a two-step approach;[24] first a regression coefficient of SHS over time for each individual is estimated, which are then compared between groups. Although this method takes into account the correlation of the serial measurements within one subject, it ignores the standard error of these individual coefficients. Therefore, standard errors obtained with this approach are generally too small, introducing the risk of false-positive findings.

The RMA used in this manuscript is a multivariate normal regression analysis.[17] An alternative statistical method to analyze repetitive measurements is Generalized Estimating Equations (GEE),[25] which is occasionally used in clinical trials.[26,27] Advantages of GEE are that the data do not have to be normally distributed and the correlation structure does not have to fit the data. A disadvantage of GEE is that it assumes that missingness is 'completely at random', which is often not the case.[28] An extension of GEE, GEE with inverse probability weights,[29,30] can deal with missing data that is not completely at random, but this extension is not readily available in standard software packages. Since for GEE the correlation structure does not have to fit the data, GEE is often less precise than a multivariate normal regression. Since in the present cohort missingness was not 'completely at random', we preferred multivariate normal regression over GEE.

In the present study, no adjustments were made in the LRs and RMAs in order to increase the comparability of the tests. However, in studies evaluating associations with risk factors, it will be relevant to adjust for factors that interfere with or modify levels of joint damage, such as treatment. Adjustments generally result in a more precise estimation since the residual variance is decreased.

In conclusion, identification of new risk factors for RA severity is important. Genetic risk factors generally have moderate to small effect sizes. Therefore it is important to differentiate true

effects from noise and to have powerful methods of analysis. The present study demonstrated that a repeated measurement analysis on subsequent radiographs provides a sensitive method to analyze associations with joint destruction over time in longitudinal cohort studies.

REFERENCES

1. Rienhoff HY Jr. Genomewide association studies and assessment of risk of disease. N Engl J Med 2010 18;363: 2077.
2. De Vries-Bouwstra J, Le Cessie S, Allaart C, et al. Using predicted disease outcome 289 to provide differentiated treatment of early rheumatoid arthritis. J Rheumatol 290 2006;33: 1747-53.
3. Syversen SW, Goll GL, Ødegård S, et al. Cartilage and bone biomarkers in rheumatoid arthritis (RA): prediction of 10-year radiographic progression [abstract]. Ann Rheum Dis 2007;66 Suppl II: 92.
4. de Rooy DP, van der Linden MP, Knevel R, et al. Predicting arthritis outcomes—what can be learned from the Leiden Early Arthritis Clinic? Rheumatology (Oxford) 2011;50: 93-100.
5. Visser K, Goekoop-Ruiterman Y, De Vries-Bouwstra J, et al. A matrix risk model for prediction of rapid radiographic progression in rheumatoid arthritis patients receiving different dynamic treatment strategies [abstract]. Ann Rheum Dis 2009;68 Suppl 3: 402.
6. Vastesaeger N, Xu S, Aletaha D, et al. Matrix risk model for prediction of rapid radiographic progression in rheumatoid arthritis [abstract]. Ann Rheum Dis 2008; 67: 63.
7. Van der Helm-van Mil AHM, Knevel R, van der Heijde D, et al. How to avoid phenotypic misclassification in using joint destruction as an outcome measure for Rheumatoid Arthritis? - Sensitive measuring of joint destruction in RA. Rheumatology (Oxford) 2010;49: 1429-35.
8. van Aken J, van Bilsen JH, Allaart CF, et al. The Leiden Early Arthritis Clinic. Clin Exp Rheumatol. 2003;21: S100-5.
9. Van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000; 27: 261-263.
10. Van der Heijde DM. Plain X-rays in rheumatoid arthritis: overview of scoring methods, their reliability and applicability. Baillieres Clin Rheumatol. 1996;10: 435-53.
11. Landewé R, van der Heijde D. Radiographic progression in rheumatoid arthritis. Clin Exp Rheumatol 2005;23: S63-S68.
12. Lard LR, Boers M, Verhoeven A et al. Early and aggressive treatment of rheumatoid arthritis patients affects the association of HLA class II antigens with progression of joint damage. Arthritis Rheum 2002;46: 899- 905.
13. Combe B, Dougados M, Goupille P, et al. Prognostic factors for radiographic damage in early rheumatoid arthritis: a multiparameter prospective study. Arthritis Rheum 2001;44: 1736-43.
14. Goronzy JJ, Matteson EL, Fulbright JW, et al. Prognostic markers of radiographic progression in early rheumatoid arthritis. Arthritis Rheum 2004;50: 43-54.
15. Mewar D, Marinou I, Coute A, et al. Association between radiographic severity of rheumatoid arthritis and shared epitope alleles: differing mechanisms of susceptibility and protection. Ann Rheum Dis 2008;67: 980-983.
16. Scherer HU, van der Linden MP, Kurreeman FA, et al. Association of the 6q23 region with the rate of joint destruction in rheumatoid arthritis. Ann Rheum Dis 2010 Mar; 69(3): 567-70.
17. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. Chapter 3 New York: John Wiley & Sons; 2004.
18. SPSS for Windows, Rel. 16.0 2001. Chicago: SPSS Inc.
19. Little RJA, Rubin DB. Statistical analysis with missing data. 334 New York: John Wiley & Sons; 2002, 2nd 335 edition.
20. WJ Gauderman (2003), "Candidate gene association studies for a quantitative trait, using parent-offspring trios", Genetic Epidemiology. 25: 327-338. http://hydra.usc.edu/gxe.
21. R Development Core Team 2009 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
22. Baron G, Ravaud P, Samsom A, et al. Missing data in randomized controlled trials of rheumatoid arthritis with radiographic outcomes: a simulation study. Arthritis Rheum 2008; 59:25–31.
23. Van der Heijden GJMG, Donders ART, Stijnen T, et al. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example. Journal of Clinical Epidemiology 2006;59: 1102-1109.
24. Landewé R, Geusens P, Boers M, et al. Markers for type II collagen breakdown predict the effect of disease-modifying treatment on long-term radiographic progression in patients with rheumatoid arthritis. Arthritis Rheum 2004;50: 1390-9.

25. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models, Biometrika 73 (1986), pp. 13–22.
26. Ødegård S, Landewé R, van der Heijde D, et al. Association of early radiographic damage with impaired physical function in rheumatoid arthritis: a ten-year, longitudinal observational study in 238 patients. Arthritis Rheum 2006;54: 68-75.
27. van der Heijde D, Landewé R, van Vollenhoven R, et al. Level of radiographic damage and radiographic progression are determinants of physical function: a longitudinal analysis of the TEMPO trial. Ann Rheum Dis 2008;67: 1267-70.
28. Fitzmaurice GM, Molenberghs G, et al. Regression model for longitudinal binary responses with informative drop-outs. Journal of the Royal Statistical Society; Series B 1995 57: 691–704.
29. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association 1995;90: 106 –121.
30. Van Ness PH, Allore HG, Fried TR, et al. Inverse Intensity Weighting in Generalized Linear Models as an Option for Analyzing Longitudinal Data with Triggered Observations. Am J Epidemiol 2010;171: 105-112.

SUPPLEMENTARY DATA

Methods of power calculations

To show the power loss in detecting genetic effects when the within-patient correlation is ignored, we simulated data from the RMA (I) and (II) for the longitudinal and cross-sectional analyses, respectively. In the following section we will discuss the simulation setup in terms of the RMA (I).

First the RMA (I) is fitted to the EAC data in order to obtain estimates for the regression coefficients (namely intercept, $\beta_{\_j}$, γ and δ) and the variance components Σ. Then baseline characteristics for 602 patients are simulated based on the EAC patients' information. In particular, regarding sex 68% are women and 32% men and their age has been simulated from a normal distribution with mean 56 and standard deviation 16. In addition, the patients were assumed to have enrolled at different inclusion periods, i.e., 18% in the first, 35% in the second and 13% in the third. Regarding the genotypic information, genotypes have been simulated such that the minor allele frequency equals 0.41 and 0.37 (similar to the HLA-SE and *TNFAIP-OLIG3* in the EAC study). Finally, 8 yearly measurements are assumed to have been scheduled for all the patients. Using the baseline characteristics longitudinal responses Y_{ij} are simulated under model (I). To induce missingness at the last visit we randomly deleted 0-85% of the recorded values. The simulation of the longitudinal responses (for each missingness percentage) has been repeated 2000 times. In each of the 2000 simulated datasets both the RMA model (I) and LR_{long} are fitted and for each model we counted the number of times (out of the 2000) that the null hypothesis $\delta = 0$ is rejected. Thereby we compute the power to detect a genetic effect with effect size equal to that estimated for the EAC patients for different missingness percentages at the last visit. The same procedure is followed when model (II) is considered.

BOX I: Formula of RMA's

General formula of multivariate normal regression:

$Y_{ij} = \text{intercept} + \beta_1 x_{ij1} + \ldots. + \beta_p x_{ijp} + \varepsilon_{ij}$, $\quad i = 1, \ldots, n, j = 1, \ldots, T$

Y_{ij}= outcome from patient *i* at time-point *j*.

β_p = coefficient of P

P = covariate / interfering variables

ε_i = error terms, we assumed a multivariate normal distribution with mean vector zero and variance-covariance matrix Σ.

Here, the outcome is written as a linear function of a set of P covariates x_{ijp}

So the RMA_{long} concerns the following formula:

$Y_{ij} = \text{intercept} + \beta_{\_j} * [time_{ij}=t_j] + \gamma * \text{risk factor}_i + \delta * time_{ij} * \text{risk factor}_i + \varepsilon_{ij}$, (I)I = 1, …, 602, j = 1, …, 8

γ = the main group effect not changing over time

δ = the difference in increase of the outcome per year.

$time_{ij}=t_j$, time as factor, this allows the mean increase in response to diminish over time.[20,21]

For the RMA_{cs} the risk factor was entered with an interaction of time as categorical variable:

$Y_{ij} = \text{intercept} + \beta_{\_j} * [time_{ij}=t_j] + \gamma * \text{risk factor}_i + \delta_j * [time_{ij}=t_j] * \text{risk factor}_i + \varepsilon_{ij}$, (II)i = 1, …, 602, j = 1, …, 8