Universiteit Leiden
The Netherlands

Genetics, autoantibodies and clinical features in understanding and predicting rheumatoid arthritis

Helm-van Mil, A.H.M. van der

Citation
Helm-van Mil, A. H. M. van der. (2006, October 26). *Genetics, autoantibodies and clinical features in understanding and predicting rheumatoid arthritis*. Retrieved from https://hdl.handle.net/1887/4929

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/1887/4929

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

Chapter 14

Genetics and clinical characteristics to predict rheumatoid arthritis. Where are we now and what are the perspectives?

A.H.M. van der Helm - van Mil
T.W.J. Huizinga

Future Rheumatology 2006;1(1):79-89.

ABSTRACT

In medicine, the perspective of prediction contains three major issues: the variability of the host, the characteristics of the disease-causing agent and the interaction between these factors (the disease process itself). The current review focuses on the prediction of RA in patients with undifferentiated arthritis. Data from inception cohorts have revealed that in about a third of the patients presenting with recent onset arthritis to a rheumatologist no diagnosis can be made, so-called undifferentiated arthritis. In a portion of these patients RA develops, whereas a substantial portion remits spontaneously. Current epidemiological data on RA incidence rates are compatible with host (genetic) factors being an independent predictor of susceptibility to RA. The majority of the genetic factors have still to be identified but HLA-alleles and PTPN22-alleles have now been identified as risk factors in a number of populations. Disease characteristics such as anti-CCP antibodies and erosions on X-rays are identified as being of high predictive value. In the future, models that take into account both genetic and clinical characteristics will have to be evaluated in patient groups with undifferentiated arthritis in order to establish the accuracy of these models in predicting with an 80% probability the chance on progression to RA. Such prediction models will be of help in treatment decisions in these patients.

INTRODUCTION

A number of different items are used in clinical prediction models. These items vary from variability of the host, the variability of the causing agent, and variability in the disease process and finally the dynamics of the interaction that takes time into account as well. Examples of the relevance of measuring pure host characteristics are e.g. the BRCA-1/2 carriership in families with a history of breast/ovarian cancer. Examples of the relevance of the characteristics of the disease-causing agent in sensu strictu are the causing micro-organisms in infectious diseases such as in community-acquired pneumonia. An example in rheumatology when both host and causing agent determine susceptibility is that certain micro-organisms (e.g. Chlamydia) particularly cause disease (reactive arthritis) in hosts that are HLA-B27 positive. Examples to detect differences in prognosis by studying the mode of interaction of host and disease-inducing process are the study of characteristics of diseased tissue such as micro-array studies in breast cancer. An example when the evolution of the disease over time is included in the determination of a response parameter is cervical abnormalities detected by a slightly abnormal cervical smear test. The value of such a test can be that the cervical smear test is advised to be repeated within three months to see whether natural regression of the abnormalities has occured or whether progression to abnormalities indicative of precancerous characteristics has occurred.

The outcome of a prediction model can be the development of a disease or disease severity. In rheumatology and particularly in rheumatoid arthritis (RA), physicians generally want to avoid morbidity and disability. Existing prediction models are therefore built to predict chronicity and erosiveness. Prediction models are of importance as they might help in treatment decisions. They may guide the choice in treatment options between wait and see, start with a relative mild treatment or initiate aggressive treatment directly. The current evidence on (early) treatment of RA is based on large trials with RA patients, in which RA is defined according to the ACR criteria. Inception cohorts include patients in who during the first visits with the current methodology a diagnosis can directly be made (about 60% of patients). About forty percent of patients in inception cohorts have a form of arthritis in which no definite diagnosis can be made; these patients are identified as undifferentiated arthritis (UA). These UA patients can go into remission, develop RA or develop other diagnosis (see Figure 1)(1). At present no data on the effects of treatment of UA patients are available. Early treatment of the patients with UA that will develop RA might be beneficial, whereas treatment of the group that will remit spontaneously is potentially harmful. The spontaneous remission rate of UA patients is about 40% (1). As current knowledge on the effects of treatment in RA is based on patients with RA classification according to the ACR-criteria it will be helpful to have a model to predict development of RA in UA patients. The prediction of RA is

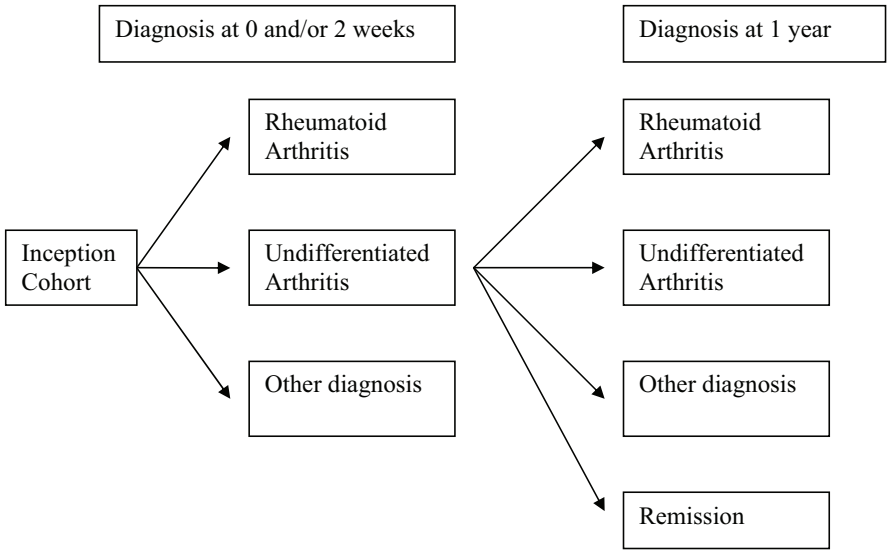

Figure 1. Flow diagram of inception cohort.

in most cases synonymous with prediction of disease persistency as the remission rate of RA is about 10-18% (2,3).

The current review is focussed on prediction of RA. The following sections will review the evidence that pure host characteristics are informative, pure disease characteristics are relevant and which evidence is available that a combination of disease characteristics, host characteristics and natural course allows prediction. The applicability of prediction models is also determined by basic epidemiological rules and the predictive value of a test/model is dependent on the prevalence of the disease in a given population. For the question on genetic testing "where we are now and what are the perspectives" this is relevant because most current data on genetic factors describe the comparison between cases and healthy controls. These are relevant data to undrape pathogenesis, but are difficult to be interpreted with regard to relevance in diagnostic testing in individual patients.

VARIABILITY IN THE HOST DETERMINES INCIDENCE OF RA

This hypothesis has a number of relevant implications for RA. First, it assumes that the trigger or triggering events for RA are common and therefore the opportunity to encounter these triggering events is not the limiting or determining factor. Second, it has implications on the thoughts whether RA is one disease or an assembly of truly different diseases and third it assumes that a number of DNA variants are associated with disease and that a DNA fingerprint might predict disease susceptibility.

Does RA have a common trigger?

The assumption that RA has a common or a combination of commonly available triggers imply that such triggers can be encountered in most populations. An argument in favour of this assumption is that RA is a worldwide condition indicating that the triggers leading to RA are worldwide available. The incidence of RA is age related, with a higher RA incidence with higher age. This age-related incidence of RA provides some evidence how many triggers are needed to get RA. Roberts-Thompson et al. studied population data obtained from the Australian Bureau of Statistics in order to assess the number of events necessary leading to RA (4). By computer modelling in which the age-specific incidence rates, the proportion of population at risk and the age at onset are included, the number of random events that must occur for the disease to manifest (given a stochastic model) was calculated. This number varied somewhere between 4 and 6 events. In inception cohorts, the patients with UA at inclusion that after 1 year of follow-up had persistent UA were significantly younger than the UA patients that during the first year developed RA (1). The difference in age between the groups that do and do not progress to RA might reveal a difference in time period and subsequent chance to acquire sufficient numbers of triggering events.

RA has a variable incidence in different populations. In Pima Indians the incidence rates in the same time periods were 10 times higher than in the Caucasian US population and 5 times higher than in the Japanese population (5). This implies that the frequency of the events leading to RA is different in the different populations and/or that the host factors in the different populations differ. The relative contribution of genetic or environmental factors is difficult to determine, but based on studies of populations that have migrated to different environments, it is likely that the majority of the difference in rates of RA in different populations can be explained by genetic factors (6). Moreover, the differences in the frequency of the identified genetic risk factor for RA, the HLA-alleles encoding the shared epitope, associate with the frequency of RA in the respective populations (7). An additional argument that RA is caused by commonly available triggers is the absence of geographical clustering of RA incident cases (8). In the aggregate, these data suggest that in modern lifestyle a combination of triggers is common in a variety of cultures but that host characteristics determine whether these triggers can lead to RA.

Further understanding can be achieved by studying populations among whom RA is not prevalent. A study in indigenous people (Aboriginals) in Australia found no paleopathological or ethnographical evidence to support the existence of RA before white settlement (9). Similarly in a rural Nigerian population RA was not observed (10) indicating that either the common trigger was not present at that time or that these populations are genetically protected. Arguments for this last statement are the much lower frequency of HLA-DR alleles encoding RA risk alleles in these populations. Careful studies have now identified RA in Aboriginals. However, in all the Aboriginal RA patients some evidence of

prior interracial marriage was found. This indicates that genetic admixture is necessary for development of RA. Yet, a contribution of changing lifestyles that is concomitant to racial admixtures cannot be easily excluded.

In a study from Minnesota that studied RA incidence rates from 1955 to 1995, the incidence rate fell progressively over the 4 decades of study, from 61.2/100,000 in 1955-1964, to 32.7/100,000 in 1985-1994 (11). A Japanese study showed that the incidence rates fell in Japan as well and the falling incidence rates over time have occurred in diverse populations such as Indians and Finnish (12). There are several possible explanations for the decrease in RA incidence. As this decrease is apparent in various populations, an explanation is likely a factor that has an identical effect in all populations throughout the world in the birth cohorts from 1890 to 1950. It is proposed that this is caused by a change in the population genome (13). The explanation for this genetic drift is that in previous times, human reproductive success was very unevenly distributed, with a minority of fertile women who gave birth to the majority of newborns. For example, in the 1912 Australian census, 50% of the children were the offspring of 1 in 7 of the women (14). However, in recent times this predominance steadily decreased since both fertile and less fertile women have equally contributed to the next generation. There are also other explanations for the decrease in RA incidence rates over time. Besides a real time dependent decline in RA, changing methodology in classification may be also important (15). In addition to a decrease in RA incidence, a decrease in RA severity over time is also reported. This decline seems to be contributable to earlier and more aggressive treatment (16).

In summary, the overview of the studies on incidence rates of RA are compatible with the notion that host characteristics are the major factors that drive whether a patient will develop RA. We suggest that most persons nowadays will encounter those 4-6 triggering events and that host factors are the driving force to explain differences in incidence rates.

What is the evidence in support of the assumption that RA is one disease versus an assembly of different diseases?

At present, RA is formally diagnosed when patients fulfil the criteria that were formulated by the ACR in 1987. Whether the patients that have RA according to these criteria, all have the same disease -characterised by an identical pathogenesis-, is questionable. Recently it was observed in a European and American population that RA patients carrying antibodies to citrullinated proteins (anti-CCP antibodies) have an association with different genetic risk factors than patients that lack these antibodies. The shared epitope encoding HLA-alleles only conferred risk to anti-CCP positive RA and not to anti-CCP negative RA (17). Anti-CCP antibodies are reported to have high disease specificity and are often present before the clinical presentation (18,19); anti-CCP antibodies are therefore thought to play a role in RA pathogenesis. The finding that the shared epitope alleles only correlate

with anti-CCP positive disease strongly suggests that RA patients that have anti-CCP antibodies have differences in the pathophysiological pathway compared to RA patients that are anti-CCP negative. This consequently induces the question whether anti-CCP positive and negative RA are different disease entities with distinct clinical characteristics. In a recent study RA patients with and without anti-CCP antibodies were extensively compared with regard to clinical characteristics. No differences were found in the characteristics on disease presentation between these two patient groups: among others the age of disease onset, the type of initial symptoms, the distribution of initial symptoms, the presence and duration of morning stiffness, and the number and distribution of painful or swollen joints was similar in anti-CCP positive and negative RA patients (20). From these data it can be concluded that different pathophysiological pathways end in one phenotypical presentation of the disease. Specific characteristics of the host such as the presence of anti-CCP antibodies subsequently associate with the course of the disease.

Which potential genetic risk factors for RA are known?

The HLA Class II molecules are the most powerful recognized genetic factors so far for RA, contributing to at least 30% of the total genetic effect. The HLA-DRB1 alleles *0101, *0102, *0401, *0404, *0405, *0408, *1001, *1402 share a conserved amino acid sequence at position 70-74 in the third hypervariable region of the DRβ1 chain. These residues constitute an α-helical domain forming one side of the antigen presenting binding site. The Shared Epitope hypothesis postulates that the shared epitope motif itself is directly involved in the pathogenesis of RA by allowing the presentation of an arthritogenic peptide. Extensive evidence exists showing associations between the shared epitope encoding alleles and susceptibility to RA. The presence of shared epitope encoding alleles is associated with odds of 3-4 to develop RA (21,22).

The second genetic risk factor is a risk allele of a haematopoietic-specific protein tyrosine phosphatase, PTPN22. This allele was identified in 17% of the North American white controls and 28% of the RA patients and confers odds of about 2 to develop RA (23-26). This allele changed the function of the protein that functions as negative regulator of T-cell activation, leading to T-cells with a lower threshold for T-cell activation. This mutation is apparently leading to several autoimmune diseases since this mutation also conferred risk for SLE, type 1 diabetes and Graves disease (27,28). The last years an increasing number of SNPs associated with RA have been identified. Some of them have not been replicated and some show different results in different populations. A genetic risk factor that is currently under investigation and seems to be associated with RA, diabetes and myocardial infarction, is MHC2TA. This SNP associates with a lower expression of MHC molecules and in a Swedish cohort of 1288 RA patients and 709 controls; this SNP conferred a 1.3 higher risk to develop RA (29). The findings on this SNP await replication. In Japanese patients and controls an association between haplotypes (combinations of SNPs on one chromosome

that tend to inherit together) of the gene encoding PADI4 with increased susceptibility to RA was observed (30). The RA-susceptible PADI4 variant was shown to produce a more stable transcript than the non-susceptible variant, implying an increased production of PADI4 and therefore higher level of citrullination by the RA-susceptible variant. Unless a higher level of citrullination, the described PADI4 haplotypes did not correlate with (the level of) anti-CCP antibodies (30). The association of PADI4 with RA is shown in the Japanese population. Data from Caucasians from France and the UK showed no association with PADI4 haplotypes and RA (31,32). Susceptibility genes can interact such that the resulting predisposition of carrying both genes is larger than the summed odds ratios of the individual genes. The presence of such interaction is important with regard to prediction. In 820 Japanese RA patients and 620 controls a risk for RA of 1.3 was identified for a risk allele in the organic cation transporter gene SLC22A4 9(33). Intriguingly, the identified SNP affects the transcriptional efficiency of SLC22A4 in vitro by altering the binding affinity of a haematopoietic transcription factor, called RUNX1. A small but significant association was observed with the minor allele in the RUNX1-gene. Importantly homozygosity for both susceptibility alleles (SLC22A4 and RUNX1) showed a high odds ratio of 9, indicative for a gene-gene interaction (33). Recently the effects of this RUNX-1 SNP was not found in a Caucasian population. (34). A SNP in the promoter region of FCRL3 is recently shown to be associated with susceptibility to RA (35). For a large number of other genes suggested to be relevant in the pathophysiology of RA, association was observed in only one study without replication. These studies concerned Beta-adrenergic receptor gene SNPs, RANKL, ICAM-1, VEGF, PDCD-1 and IL1-RA gene (36-40).

Besides genetic risk factors that confer a higher risk to develop RA, there also a genetic risk factors that protects to RA. This concerns particularly the HLA-DRB1 alleles that encodes for the amino acids DERAA (DRB1*0103, *0402, *1102, *1103, *1301, *1302, *1304). Interestingly, the HLA-DRB1 alleles can encode for different alleles with an opposite effect on disease susceptibility. The protective effect of the DERAA-encoding alleles is independent from the shared epitope encoding alleles with predisposing effects (22,41,42). Both in the presence and in the absence of shared epitope encoding alleles the DERAA-encoding alleles confer significant lower odds of 0.6 to develop RA (22).

In summary, the current knowledge on well-validated genetic risk factors to be included in a DNA fingerprint is limited to HLA-DRB1 and PTPN22. HLA-DRB1 is estimated to account for ~30% of the genetic component of this autoimmune disease (43) while the contribution of PTPN22 is much smaller. This implies that a significant part of the genetic contribution is still to be identified. In a number of whole genome scans (44) a considerable number of peaks of linkage have been identified. In a study to estimate the number of true RA gene regions that took into account both the heterogeneity of RA and the performance of a dense genome scan, it was found that 8 +/- 4 regions (mean +/- SD) were true-positives and evidence for 3 additional regions was provided from covariate-based

analysis (45). One of those regions is the HLA-DRB1 locus, meaning that at least 10 +/- 4 additional genes will be identified with each quite modest effect. Technical progress such as SNP-based linkage analysis has been demonstrated to allow loci to be defined more precisely (46). The chance that this will lead to the identification of the majority of the genetic risk factors is larger when RA is caused by a dozen common genetic variants than when RA is the result of many rare mutations. Given the fact that HLA and PTPN22 have already been identified, we speculate that RA is caused by a dozen common genetic variants. The statistical methods to evaluate many gene variants with disease status, as in candidate-gene case-control studies are still in its infancy especially for the low effect sizes of the individual disease loci and the sometimes low frequencies of the disease allele(s). The standard methods to evaluate the association of multiple markers with disease status are based on multimarker multivariate analyses. For such multimarker multivariate analyses, one typically uses logistic regression to test simultaneously the main effects (and possibly interactions) of multiple markers. For each marker, a covariate can be created, such as the number of rare alleles at each marker. When this type of coding is used in logistic regression, the resulting score statistic for each marker implies many degrees of freedom, implying that the overall model suffers from weak power. Moreover complex models tend to overfit the data stressing the necessity for replication in independent cohorts. Despite these difficulties the outlook is that the genetic contribution to RA is about 50-60% (47). This number is estimated by the comparison between the concordance rates in monozygotic twins and the prevalence in the respective populations (47). This high percentage implies that measurement of genetic host characteristics is likely to have a role in a predictive test.

Which environmental risk factors for RA are known?

So far smoking has been shown the only plausible environmental risk factor for RA. An association with smoking in RA is particularly found for rheumatoid factor positive RA compared to rheumatoid factor negative disease (48,49). Current or ex-smokers have a risk for autoantibody positive RA with an odds ratio of 1.7-1.9. The risk increases with cumulative dose of smoking (48). A recent report investigated whether smoking is primarily associated with the development of rheumatoid factor or anti-CCP antibodies. This study nicely revealed a gene-environment interaction by showing that in the presence of HLA-shared epitope alleles smoking significantly contributes to the development of anti-CCP antibodies (50).

A predictive effect of oral anticonceptives on RA has been claimed (51). This finding has however not been replicated in the Nurses' Health study (52).

Predictive value of a DNA fingerprint-test

A large problem in transferring the data on genetic risk factors to prediction models is that the most current studies compared patients and controls, revealing odds ratios that are determined on group levels. The value of these genetic risk factors for individual predictive testing may be limited. Compare for example the statistical models to predict the pre-test probability on BRCA-1/2 genes. BRCA1 or 2 carriers have a very strong risk for ovarian/breast cancer. The statistical models to predict the presence of a BRCA1/2 risk allele are only informative in a selected population with affected family members (53). This example underlines that findings for a whole group cannot be automatically used for prediction in subgroups of patients or for individuals. Genes may confer risk to subgroups of RA patients. For example, the well-known HLA-shared epitope alleles predispose particularly to anti-CCP positive RA (17). The BRCA- example also elucidates that the predictive value of a test depends on the prevalence of a disease in a population. For UA, a number of inception cohorts of patients with recent onset arthritis have identified patients with a form of arthritis that has the potential for a persistent course, without fulfilling the classification criteria of other rheumatic disorders (54). In nine cohorts, the proportion of patients with UA that evolved into RA within 1 year varied from 17% to 32%. Thus, in this group of UA patients the pre-test probability of developing RA varies between 17-32%. Given the dynamics of development of UA to either remission or progression to RA, the evaluation of predictive models for this patient group is highly relevant.

CHARACTERISTICS OF THE DISEASE PROCESS AND PREDICTION

The theoretical background of this section is the assumption that the expression of the disease in an initial phase allows prediction of the outcome. The genomic revolution has fuelled much optimism that gene expression profiles allow such outcome measures. Gene expression profiles are currently used in breast cancer to select the patients that would benefit from adjuvant therapy (55). However, others warned that the prognostic value of the published micro-array results in cancer studies should be considered with caution as the list of genes identified as predictors of prognosis was highly unstable and the molecular signatures depended strongly on the selection of patients (56). The prognostic value of the several micro-arrays in oncology therefore needs replication.

Disease characteristics present at the presentation of UA that predict progression to RA

The most important and best-validated disease characteristics with regard to prediction are auto-antibodies (anti-CCP and rheumatoid factor, RF) and the presence of erosions on the radiographs of hands and feet at initial presentation. In univariate analysis, the

presence of anti-CCP antibodies in patients with a UA conferred an odds ratio of 38 to develop RA, compared to anti-CCP negative patients with UA (57). A logistic regression model showed odds of 16 for anti-CCP antibodies in the prediction of RA (58). Raza et al. followed 124 patients with synovitis for less than 3 months for 72 weeks and assessed the prognostic value of anti-CCP antibodies and RF (59). In this study the combination of anti-CCP antibodies and RF had a positive predictive value of 100% and a negative predictive value of 88% for a diagnosis of RA (59).

Clinical disease characteristics of 329 UA patients that presented to the Leiden Early Arthritis Clinic differed between the UA patients that developed RA versus those who did not develop RA. Disease characteristics associated with RA development were a higher age (55 versus 46), female sex (66% versus 50%), duration of morning stiffness (60 minutes versus 15 minutes), longer duration of symptoms (131 versus 81 days) and a higher number of swollen joints (4 versus 2)(1).

Visser et al. developed a clinical model for the prediction of three forms of arthritis outcome: self-limiting disease, persistent non-erosive disease and persistent erosive disease (60). For the development of this model the first 524 consecutive patients referred to the Leiden Early Arthritis Clinic were studied and the arthritis outcome was recorded after two years of follow-up. The developed prediction model consisted of 7 variables: symptom duration at first visit, morning stiffness for ≥ 1 hour, arthritis in ≥ 3 joints, bilateral compression pain in the metatarsophalangeal joints, RF positivity, anti-CCP positivity and the presence of erosions at study entry. The ROC area under the curve for discrimination between self-limiting and persistent non-erosive arthritis was 0.84 and for discrimination between persistent nonerosive and erosive arthritis 0.91 (60). After the addition of predisposing HLA-Class II alleles, the discriminative ability of the model was not significantly improved (60). The model derived by Visser used all patients of the Leiden Early Arthritis Clinic instead of only the patients with UA. The advantage of the model of Visser is that it can be used for a "random" patient with arthritis that visited an outpatient clinic of a rheumatologist, the disadvantage is that it also predicts occurrence of RA in patients that already fulfilled the classification criteria for RA. Currently it is analyzed whether the clinical characteristics used in this model also have predictive value in patients with UA.

To what extent can clinical observation be used in prediction?

Clinical observation of the natural course is the best way of predicting what the subsequent course will be. From a retrospective viewpoint, the history of the patients can be used as illustrated in the model proposed by Visser in which a long duration of complaints was associated with higher odds for chronic and erosive disease (60). The policy to include a "wait and see" policy can only be taken in the context when the possible disadvantages are considered. The progress form UA to RA is characterized by the acquisition of certain

phenotypic characteristics that form the ACR classification criteria, including joint destruction with subsequent deformities, and extra-articular features such as nodules. Given the accumulating evidence that appropriate therapy might prevent the development of a detrimental RA phenotype, observation without treatment is in our view only justified when the patients doesnot fullfill the ACR criteria. Specific studies that compare the initiation of treatment as a function of disease duration are scarce. Valuable data were obtained in a 5-year follow-up study by Egsmose et al. in which early treatment with intramuscular gold was compared with a delayed treatment strategy (61). The early treatment group showed improvement with respect to signs and symptoms, physical function and radiographic progression, thus supporting the hypothesis of a therapeutic 'window of opportunity'. In another trial by van der Heide et al. immediate versus delayed introduction of disease-modifying anti-rheumatic drug (DMARD) therapy were compared in patients with recently diagnosed RA (62). Early introduction of DMARDs showed greater patient improvement with regards to signs and symptoms, physical function and radiographic progression. Van Aken et al. compared the conventional pyramid strategy, consisting of sequential institution of NSAIDs and subsequent DMARD therapy, with immediate initiation of DMARD therapy in an observational study: again, the early treatment group showed less radiographic progression (63). Finally, an observational study performed at the Norfolk Arthritis Register provided evidence that patients in whom DMARD therapy was initiated within 6 months of diagnosis of RA had a better 5-year radiographic outcome than patients starting DMARD therapy 6 months after RA diagnosis (64). All the aforementioned studies have investigated the importance of timing of treatment with regard to diagnosis.

In conclusion, a role for clinical observation in prediction for development of RA seems only justified in UA patients with a low probability to develop RA.

CONCLUSION

In the search to predict RA, current data indicate that host characteristics are relevant to predict RA. The identification of these host characteristics have yielded HLA alleles as risk and as protective alleles and identified PTPN22 as the second risk gene. Progress to identify genetic risk factors is slow due to the fact that each gene most likely has a very small effect and the lack of good statistical models to analyse combinations of genetic risk factors. Clinical factors to be included in a predictive model are most likely the duration of morning stiffness, the presence of an anti-CCP response and the presence of erosive abnormalities on X-rays of hands and feet.

SUMMARY AND FUTURE PERSPECTIVE

Prediction of the future is impossible but a model can provide a probability to an individual patient. Such a prediction model should guarantee a clinician and patient enough certainty (e.g. 80%) that a patient is assigned to the correct category. In the context of UA were about a third will develop RA and two third will not develop RA, it is not exactly known what the minimum value of R^2 has to be to get a valuable prediction model. The R^2, the fraction of explained variation, is a measure for the model's ability to predict. It compares the mean squared error of the prognostic model to the mean squared error of the model without any prognostic variables and does not have a dimension. Some indication for an acceptable R^2 can be obtained from a similar problem, the prediction of the severity of joint destruction in RA. De Vries et al. recently determined the adequacy of clinical parameters in the prediction of joint destruction (64). This model had a R^2 of 0.36 and correctly classified 62% of patients. Furthermore it was calculated that to get a correct classification of 80% of the patients, such a hypothetical model should have a R^2 of 0.9 (65). A model that predicts joint damage scores gives an estimate for a continuous variable and is therefore different from a model that predicts the absence or presence of development of RA. Nevertheless the data as presented by De Vries et al. indicate the requirements for a model to adequately predict disease development in patients with UA. Currently, to our knowledge there are no prediction models analysed that are able to determine with at least 80% certainty whether an individual patient will or will not develop RA. However, given that more genetic factors that associate with susceptibility to RA will most likely be identified in the next decade, we expect that the results of the genetic variants will be included in future prediction models. The predictive value of disease characteristics such as anti-CCP antibodies has already been identified and given its very large and very specific effect, this will be included in prediction models. Clinical characteristics have not yet been defined in great detail but we expect that with the current inclusion of many patients in different early arthritis initiatives, these data will become available in the next decade. Given the expectation that genetic, serological and clinical data each contain independent information, it should be possible to combine these datasets to gain more accurate prognostic information. Hopefully the R^2 of such a model will be sufficiently large to allow prediction at the patient level.

This review is focussed on prediction of the diagnosis of RA but not on prediction of the prognosis of RA. This is caused by the lack of epidemiological data whether severity of RA such as rate of joint destruction is caused by e.g. genetic factors. In the next decade we expect that these basic epidemiological data will become available leading to development of predictive tests for outcome of RA as well.

REFERENCES

1. Van Aken J, van Dongen H, LeCessie S, Allaart CF, Breedveld FC, Huizinga TWJ. Long-term outcome of rheumatoid arthritis that presented with undifferentiated arthritis compared to rheumatoid arthritis at presentation - an observational cohort study. Ann Rheum Dis. 2006;65(1):20-5.
2. Linn-Rasker SP, Allaart CF, Kloppenburg M, Breedveld FC, Huizinga TWJ. Sustained remission in a cohort of patients with RA: association with absence of IgM-rheumatoid factor and absence of anti-CCP antibodies. Int J Advances Rheumatology. 2004:2(4): 4-6.
3. van der Helm-van Mil AH, Dieude P, Schonkeren JJ, Cornelis F, Huizinga TW. No association between tumour necrosis factor receptor type 2 gene polymorphism and rheumatoid arthritis severity: a comparison of the extremes of phenotypes. Rheumatology (Oxford). 2004:43(10):1232-4.
4. Roberts-Thomson PJ, Jones ME, Walker JG, Macfarlane JG, Smith MD, Ahern MJ. Stochastic processes in the causation of rheumatic disease. J Rheumatol. 2002;29(12): 2628-34.
5. Del Puente A, Knowler WC, Pettitt DJ, Bennett PH. High incidence and prevalence of rheumatoid arthritis in Pima Indians. Am J Epidemiol. 1989;129(6):1170-8.
6. Silman AJ, Pearson JE. Epidemiology and genetics of rheumatoid arthritis. Arthritis Res 2002;4:S265-72.
7. Ferucci ED, Templin DW, Lanier AP.Rheumatoid arthritis in American Indians and Alaska Natives: a review of the literature Semin Arthritis Rheum. 2005;34(4):662-7.
8. Silman A, Harrison B, Barrett E, Symmons D. The existence of geographical clusters of cases of inflammatory polyarthritis in a primary care based register. Ann Rheum Dis. 2000;59:152-154.
9. Rachel A Roberts-Thomson,[a] P J Roberts-Thomson Rheumatic disease and the Australian aborigine. Ann Rheum Dis. 1999;58:266-70.
10. Silman AJ, Ollier W, Holligan S, Birrell F, Adebajo A, Asuzu MC, et al. Absence of rheumatoid arthritis in a rural Nigerian population. J Rheumatol. 1993;20: 618-622.
11. Doran MF, Crowson CS, O'Fallon WM, Hunder GG, Gabriel SE. Trends in incidence and mortality in rheumatoid arthritis in Rochester, Minnesota, over a forty-year period. Arthritis Rheum. 2002;46(3):625-31.
12. Shichikawa K, Inoue K, Hirota S, Maeda A, Ota H, Kimura M, Ushiyama T, Tsujimoto M. Changes in the incidence and prevalence of rheumatoid arthritis in Kamitonda, Wakayama, Japan, 1965-1996. Ann Rheum Dis. 1999;58(12):751-6.
13. Huizinga TWJ, Linn-Rasker S, Lard RG, Westendorp RG. Genetic drift as an explanation for the reduced incidence of RA, Arthritis Rheum. 2002;46(11):3107.
14. Cummins J. Evolutionary forces behind human infertility. Nature 397: 557-8 (1993).
15. Uhlig T, Kvien TK. Is rheumatoid arthritis disappearing? Ann Rheum Dis. 2005:64(1): 7-10.
16. Welsing PMJ, Fransen J, Van Riels PLCM. Is the disease course of rheumatoid arthritis becoming milder?: Time trends since 1985 in an inception cohort of early rheumatoid arthritis. Arthritis Rheum. 2005;2(9):2616-24.
17. Huizinga TWJ, Amos CI, Van der Helm- van Mil AH, et al. Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins, Arthritis Rheum. 2005;52(11):3433-8.
18. Rantapaa-Dahlqvist S, de Jong BA, Berglin E, et al. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. Arthritis Rheum. 2003;48:2741-2749.

19. Nielen MM, van Schaardenburg D, Reesink HW, et al. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. Arthritis Rheum. 2004;50:380-386.
20. Van der Helm-van Mil AH, Verpoort KN, Breedveld FC, et al. Antibodies to citrullinated proteins and differences in clinical progression of rheumatoid arthritis Arthritis Res Ther 2005;7(5):R949-58.
21. MacGregor A. Ollier W, Thomson W, et al. HLA-DRB1*0401/0404 genotype and rheumatoid arthritis: increased association in men, young age at onset, and disease severity. J Rheumatol 1995;22:1032-1036.
22. Van der Helm-van Mil AH, Huizinga TW, Schreuder GM, et al An independent role for protective HLA Class II alleles in rheumatoid arthritis severity and susceptibility, Arthritis Rheum 2005;52(9):2637-44.
23. Begovich AB, Carlton VE, Honigberg LA, et al. A Missense Single-Nucleotide Polymorphism in a Gene Encoding a Protein Tyrosine Phosphatase (PTPN22) Is Associated with Rheumatoid Arthritis. Am J Hum Genet. 2004;75(2):330-337.
24. Van Oene M, Wintle RF, Liu X, et al. Association of the lymphoid tyrosine phosphatase R620W variant with rheumatoid arthritis, but not Crohn's disease, in Canadian populations. Arthritis Rheum 2005;52(7):1993-8.
25. Simkins HM, Merriman ME, Highton J, et al. Association of the PTPN22 locus with rheumatoid arthritis in a New Zealand Caucasian cohort. Arthritis Rheum 2005;52(7):2222-2225.
26. Wesoly J, Chokkalingam AP, Van der Helm-van Mil AH, et al. Association of the PTPN22 C1858T SNP with RA phenotypes in an inception cohort. Arthritis Rheum 2005;52(9):2948-50.
27. Bottini N, Musumeci L, Alonso A, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet. 2004;36(4):337-338.
28. Criswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, Moser KL, Begovich AB, Carlton VE, Li W, Lee AT, Ortmann W, Behrens TW, Gregersen PK. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. Am J Hum Genet 2005;76(4):561-71.
29. Swanberg M, Lidman O, Padyukov L, et al. MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. Nat Genet. 2005;37(5):486-94.
30. Suzuki A, Yamada R, Chang X, et al. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat Genet. 2003;34(4):395-402.
31. Caponi L, Petit-Teixeira E, Sebbag M, et al. A family-based study shows no association between rheumatoid arthritis and the PADI4 gene in a French Caucasian population. Ann Rheum Dis 2004;64(4):587-93.
32. Barton A, Bowes J, Eyre S et al. A functional haplotype of the PADI4 gene associated with rheumatoid arthritis in a Japanese population is not associated in a United Kingdom population. Arthritis Rheum 2004;50(4):1117-1121.
33. Tokuhiro S, Yamada R, Chang X, et al. An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. Nat Genet 2003;35(4):341-348.
34. Wesoly JZ, Toes REM, Slagboom PE, Huizinga TWJ. RUNX1 intronic SNP is not associated with rheumatoid arthritis susceptibility in Dutch Caucasians. Rheumatology 2005;44(9):1196.

35. Kochi Y, Yamada R, Suzuki A, Harley JB, Shirasawa S, Sawada T, Bae SC, Tokuhiro S, Chang X, Sekine A, Takahashi A, Tsunoda T, Ohnishi Y, Kaufman KM, Kang CP, Kang C, Otsubo S, Yumura W, Mimori A, Koike T, Nakamura Y, Sasazuki T, Yamamoto K. A functional variant in FCRL3, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. Nat Genet. 2005;37(5):478-85.
36. Wu H, Khanna D, Park G, et al. Interaction between RANKL and HLA-DRB1 genotypes may contribute to younger age at onset of seropositive rheumatoid arthritis in an inception cohort. Arthritis Rheum 2004;50(10):3093-3103.
37. Lee EB, Kim JY, Kim EH, et al. Intercellular adhesion molecule-1 polymorphisms in Korean patients with rheumatoid arthritis. Tissue Antigens 2004;64:473-477.
38. Han SW, Kim GW, Seo JS, et al. VEGF gene polymorphisms and susceptibility to rheumatoid arthritis. Rheumatology 2004;43(9):1173-1177.
39. Prokunina L, Padyukov L, Bennet A, et al. Association of the PD-1.3A allele of the PDCD1 gene in patients with rheumatoid arthritis negative for rheumatoid factor and the shared epitope Arthritis Rheum 2004;50(6):1770-1773.
40. Lee YH, Kim HJ, Rho YH, et al. Interleukin-1 receptor antagonist gene polymorphism and rheumatoid arthritis. Rheumatol Int. 2004;24(3):133-136.
41. Mattey DL, Dawes PT, Gonzalez-Gay MA, Garcia-Porrua C, Thomson W, Hajeer AH, et al. HLA-DRB1 alleles encoding an aspartic acid at position 70 protect against development of rheumatoid arthritis. J Rheumatol. 2001;28(2):232-9.
42. Wagner U, Kaltenhäuser S, Pierer M, Seidel W, Trölzsch, Häntzschel, et al. Prospective analysis of the impact of HLA-DR and –DQ on joint destruction in recent-onset rheumatoid arthritis. Rheumatology 2003;42:553-62.
43. Deighton CM, Walker DJ, Griffiths ID, et al. The contribution of HLA to rheumatoid arthritis. Clin Genet 1989;36:178-82.
44. Huizinga TW. Genetics in rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2003;17(5):703-16.
45. Osorio Y Fortea J, Bukulmez H, Petit-Teixeira E, Michou L, Pierlot C, Cailleau-Moindrault S, Lemaire I, Lasbleiz S, Alibert O, Quillet P, Bardin T, Prum B, Olson JM, Cornelis F. Dense genome-wide linkage analysis of rheumatoid arthritis, including covariates. Arthritis Rheum. 2004;50(9):2757-65.
46. John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, et al.. Whole-Genome Scan, in a Complex Disease, Using 11,245 Single-Nucleotide Polymorphisms: Comparison with Microsatellites. Am. J. Hum. Genet. 2004;75:54-64.
47. MacGregor AJ, Snieder H, Rigby AS, et al. Characterising the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 2004;43:31-37.
48. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, Klareskog L, Alfredsson L; EIRA study group. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. Ann Rheum Dis.2003;62(9):835-41.
49. Padyukov L, Silva C, Stolt P, Alfredsson L, Klareskog L. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. Arthritis Rheum. 2004;50(10):3085-92.
50. Linn-Rasker SP, van der Helm-van Mil AH, Van Gaalen FA, Kloppenburg M, de Vries R, le Cessie S, Breedveld FC, Toes RE, Huizinga TW. Smoking is a risk factor for anti-CCP antibodies only in RA patients that carry HLA-DRB1 Shared Epitope alleles. Ann Rheum Dis. 2006;65(3):366-71.
51. Doran MF, Crowson CS, O'Fallon WM, Gabriel SE. The effect of oral contraceptives and estrogen replacement therapy on the risk of rheumatoid arthritis: a population based study. J Rheumatol. 2004;31(2):207-13.

52. Karlson EW, Mandl LA, Hankinson SE, Grodstein F. Do breast-feeding and other reproductive factors influence future risk of rheumatoid arthritis? Results from the Nurses' Health Study. Arthritis Rheum. 2004;50(11):3458-67.
53. de la Hoya M, Diez O, Perez-Segura P, et al Pre-test prediction models of BRCA1 or BRCA2 mutation in breast/ovarian families attending familial cancer clinics. J Med Genet 2003;40(7):503-10.
54. Verpoort KN, van Dongen H, Allaart CF, Toes RE, Breedveld FC, Huizinga TW. Undifferentiated arthritis--disease course assessed in several inception cohorts. Clin Exp Rheumatol. 2004;22(5 Suppl 35):S12-7.
55. Van 't Veer LJ, Dai H, Van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530 - 536.
56. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet 2005;5;365(9458):488-92.
57. Van Gaalen FA, Linn-Rasker SP, van Venrooij WJ, et al. Autoantibodies to cyclic citrullinated peptides predict progression to rheumatoid arthritis in patients with undifferentiated arthritis: a prospective cohort study. Arthritis Rheum 2004;50(3):709-15.
58. Berglin E, Padyukov L, Sundin U, et al. A combination of autoantibodies to cyclic citrullinated peptide (CCP) and HLA-DRB1 locus antigens is strongly associated with future onset of rheumatoid arthritis. Arthritis Res Ther 2004;6(4):R303-8.
59. Raza K, Breese M, Nightingale P, et al. Predictive value of antibodies to cyclic citrullinated peptide in patients with very early inflammatory arthritis. J Rheumatol 2005;32(2): 231-8.
60. Visser H, le Cessie S, Vos K, et al. How to diagnose rheumatoid arthritis early: a prediction model for persistent (erosive) arthritis. Arthritis Rheum 2002;46(2):357-65.
61. Egsmose C, Lund B, Borg G, et al. Patients with rheumatoid arthritis benefit from early 2nd line therapy: 5 year followup of a prospective double blind placebo controlled study. J Rheumatol 1995;22: 2208–2213.
62. van der Heide A, Jacobs JW, Bijlsma JW, et al. The effectiveness of early treatment with "second-line" antirheumatic drugs. A randomized, controlled trial. Ann Intern Med 1996;124: 699–707.
63. van Aken J, Lard LR, le Cessie S, et al Radiological outcome after four years of early versus delayed treatment strategy in patients with recent onset rheumatoid arthritis. Ann Rheum Dis 2004;63: 274–279.
64. Bukhari MA, Wiles NJ, Lunt M, et al. Influence of disease-modifying therapy on radiographic outcome in inflammatory polyarthritis at five years: results from a large observational inception study. Arthritis Rheum 2003;48: 46–53.
65. de Vries-Bouwstra JK, Le Cessie S, Allaart CF, Breedveld FC, Huizinga TWJ. Using predicted disease outcome to provide differentiated treatment of early rheumatoid arthritis, submitted.