RHEUMATOLOGY

Rheumatology 2016;55:6–15
doi:10.1093/rheumatology/keu287
Advance Access publication 4 August 2014

# RA: from risk factors and pathogenesis to prevention

## Strategies to predict rheumatoid arthritis development in at-risk populations

**Elizabeth W. Karlson[1], Dirkjan van Schaardenburg[2] and Annette H. van der Helm-van Mil[3]**

REVIEW

### Abstract

The development of RA is conceived as a multiple hit process and the more hits that are acquired, the greater the risk of developing clinically apparent RA. Several at-risk phases have been described, including the presence of genetic and environmental factors, RA-related autoantibodies and biomarkers and symptoms. Intervention in these preclinical phases may be more effective compared with intervention in the clinical phase. One prerequisite for preventive strategies is the ability to estimate an individual's risk adequately. This review evaluates the ability to predict the risk of RA in the various preclinical stages. Present data suggest that a combination of genetic and environmental factors is helpful to identify persons at high risk of RA among first-degree relatives. Furthermore, a combination of symptoms, antibody characteristics and environmental factors has been shown to be relevant for risk prediction in seropositive arthralgia patients. Large prospective studies are needed to validate and improve risk prediction in preclinical disease stages.

**Key words:** rheumatoid arthritis, epidemiology, arthralgia, (genetic) risk, autoantibodies, ACPA, prediction, preclinical.

**Rheumatology key messages**

- The highest-risk groups for RA are first-degree relatives and seropositive arthralgia patients.
- For both groups, validated prediction rules are needed to identify persons at risk for RA.
- Proof-of-concept intervention studies will reveal the efficacy of intervening in persons at risk for RA.

### Introduction

Numerous studies have been undertaken to seek risk factors for RA. For several factors it remains unclear whether they contribute to the development of RA. For example, many viruses have been studied, but consistent evidence that infections contribute to the development of RA is lacking. For other factors such as autoantibodies, smoking and genetic variants, associations with RA have been convincingly shown. Because RA is a complex disease, these risk factors are assumed to contribute to the development of RA in a multiple hit model. Knowledge of risk factors and consideration of time course have led to the conception of several preclinical stages of RA. The European League Against Rheumatism (EULAR) Study Group for Risk Factors for RA has recommended that, in prospective studies, individuals without RA but at risk for RA can be described as having genetic risk factors for RA, environmental risk factors for RA, systemic autoimmunity associated with RA, symptoms without clinical arthritis and unclassified arthritis [1]. Although not every patient necessarily passes through all of these preclinical phases [1], an advantage of defining different phases is that it provides a framework for defining risk factors over a time course.

This review focuses on risk prediction in persons within preclinical phases who might have an increased risk of developing RA. Medical literature databases (PubMed,

[1]Division of Rheumatology, Allergy, and Immunology, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA, [2]Jan van Breemen Research Institute | Reade, Amsterdam and [3]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands

Submitted 24 March 2014; revised version accepted 23 May 2014

Correspondence to: Dirkjan van Schaardenburg, Reade Postbus 58271, 1040 HG Amsterdam, The Netherlands.
E-mail: d.v.schaardenburg@reade.nl

Elizabeth W. Karlson and Dirkjan van Schaardenburg contributed equally to this work.

Embase, Web of Science) were searched. Terms in our search were RA, arthralgia, risk, prediction, models, genes, familial RA, family history and environmental factors. Additional articles were identified through hand searches. In this review we will not use the term pre-RA since it can only be determined in retrospect after the patient has developed clinically apparent RA. From a prospective point of view, many persons with multiple risk factors will never develop RA and it would be inappropriate to classify these persons as being in a predisease stage of RA.

## Risk prediction using genetic and environmental risk factors

Genetic factors are clearly important in RA susceptibility: having a family history of RA increases the risk of RA by 3- to 9-fold [2]. Pedigree-based studies estimate heritability at ~50% for seropositive RA [3, 4]. The HLA-DRB1 region shows the strongest association with RA, with shared epitope alleles (HLA-SE) associated with a 3-fold increased risk [5]. Amino acids at positions 11, 13, 71 and 74 explain much of the HLA-DRB1 association and account for 12.7% of heritability [6]. There are now >100 confirmed non-HLA RA susceptibility alleles from genome-wide association studies [7, 8]. Genes identified outside of the major histocompatibility complex (MHC) region explain 5.5% and 4.7% of heritability in Europeans and Asians, respectively [8].

Multiple environmental, lifestyle and behavioural risk factors have been studied for association with the development of RA, however, cigarette smoking is the strongest and most consistent factor identified [9] with a clear dose response [10–12]. As with genetic studies, environmental factors are most strongly associated with the seropositive RA phenotype including smoking, potentially due to heterogeneity among the seronegative phenotype. It is estimated that 25% of all RA and 35% of seropositive RA risk can be attributed to smoking [11, 13]. Factors with moderate evidence for association with higher RA risk include lower educational level, high birth weight and obesity, and factors associated with lower risk include moderate alcohol intake and breastfeeding [14, 15]. Other exposures such as silica dust, solvents, air pollution and ultraviolet (UV) light have shown modest associations with the risk of RA, while reproductive and hormonal factors, dietary factors and periodontitis have shown the least consistent results, especially when prospective studies where exposure is assessed prior to outcomes are compared with case–control studies, which are subject to recall bias [14, 15]. It is likely that environmental factors interact with genetic factors in complex networks that are yet to be elucidated. However, an interaction between the strongest genetic risk factor (HLA-SE) and smoking and seropositive RA has been demonstrated [16] and replicated with a dose effect for both the number of HLA-SE alleles and pack-years of smoking [13, 17], supporting the biological relevance of this interaction.

Risk models for RA have been developed in cohort studies where subjects are followed prospectively for the development of RA and RA serological phenotypes [14, 18–20]. In these studies, healthy, asymptomatic populations include subjects without RA but at risk for RA due to genetic risk factors and exposed to

**Table 1** Measures to assess the model fit and predictive ability of RA risk models

| Aspect | Measure | Visualization | Characteristic |
|---|---|---|---|
| Overall performance | Nagelkerke's $R^2$ | — | Quantification of variance explained by the model [24] |
| Goodness-of-Fit | Hosmer-Lemeshow $\chi^2$ | — | How well observed fit predicted results [25] |
| Discrimination | AUC | ROC | Rank order statistic for a pair of patients with and without the outcome |
| Discrimination | IDI | — | Per cent improvement in overall sensitivity and 'one minus specificity' of the new model compared with the baseline model; is less dependent on the AUC of the baseline model [21, 26, 27] |
| Reclassification | NRI | Reclassification table | Summary measure quantifying the correct upward *vs* downward movement in the model based on predicted probabilities for events and non-events when adding a new predictor to a baseline model [21, 26, 27] |
| Quantification of clinical usefulness | Net benefit | — | The relative weights of harms and benefits of overtreatment is integrated when assessing the number of true and false positives; this can be done for several cut-offs of a model |

AUC: area under the curve, ROC: receiver operating characteristic; NB: net benefit; NRI: net reclassification index; IDI: integrated discrimination index.

environmental risk factors (phases A and B of the EULAR classification). Risk prediction models with and without the new predictor(s) are often compared using analysis of the area under the receiver operating curve (AUC), a measure of discrimination. Improvement in discrimination can be quantified with novel statistical methods including the integrated discrimination index (IDI) and the continuous net reclassification index (Table 1) [17, 21–23].

Studies of environmental predictors of the development of RA-related systemic autoimmunity have focused on cohorts of high-risk subjects who are first-degree relatives (FDRs) of RA probands [e.g. Studies of the Etiology of RA (SERA) in the USA [28] and the North American Native (NAN) populations in Canada] followed for the development of autoantibodies and inflammatory arthritis [29]. Among FDRs in SERA, smoking, the absence of exposure to oral contraceptives [30] and antibodies against peptidylarginine deiminase type 4 (PAD-4) [31] were associated with the development of RF. Among FDRs in the NAN population, reproductive factors were associated with the development of RF and ACPA [32], and anti-*Porphyromonas gingivalis* antibodies [33], a marker of periodontitis, but not PAD-4 antibodies [34] were associated with development of ACPA.

To estimate the cumulative impact of multiple genetic loci, a weighted RA genetic risk score (GRS) in which the weight of each risk allele is the log of published odds ratios (ORs) has been developed for risk prediction in other diseases and applied to RA studies. Among the Nurses' Health Study (NHS) cohorts (female only) and the Swedish Epidemiologic Investigation in RA (EIRA) study (male and female), adding GRS-22 (with 8 HLA-SE alleles and 14 non-HLA alleles) to models with age and smoking significantly improved model discrimination [AUC 0.57–0.66 in the NHS and 0.63– 0.75 in the EIRA (both $P = 0.0001$)] [19] (Table 1). Further analyses in these cohorts adding age, smoking, alcohol, parity (to the female model), weighted GRS-39 (8 HLA-SE alleles, 31 non-HLA alleles) as well as an HLA-SE $\times$ smoking interaction term had AUCs of 0.72 in the NHS and 0.72 in EIRA females and 0.76 in EIRA males. Models with an expanded set of epidemiological variables including region and reproductive and occupational factors produced AUCs of 0.738 in the NHS and 0.724 in EIRA females and 0.769 in EIRA males. After stratification for family history, women with a family history of RA or SLE had an AUC of 0.85 in the full NHS model for seropositive RA and women with a family history of RA had an AUC of 0.85 for ACPA + RA in the EIRA [35]. The joint effect of high GRS-39 and family history was an OR of 6.63 (range 3.30–13.31) in the NHS and 8.24 (4.64–14.64) in the EIRA. This work suggests that prediction models applied to high-risk subjects such as those with a positive family history produce the optimal discrimination.

Other studies employing the GRS cumulative score include an electronic health records (EHRs)–based cohort that demonstrated an AUC of 0.71 using GRS-29 (1 HLA-SE allele, 28 non-HLA alleles) [20]. Another study of European cohorts demonstrated an AUC of 0.716 using GRS-45 (imputed amino acids at positions 11, 71, 74 of HLA-DRB1, 45 non-HLA alleles) that was improved to 0.724 by adding environmental factors and gene–environmental interaction terms in a subset with smoking data [23]. Finally, simulation population analyses applied to individual-level data using data from two large case–control studies from the UK demonstrated AUCs of 0.796 and 0.756 based on GRS-31 (15 four-digit/10 two-digit HLA-DRB1 alleles, 31 non-HLA alleles). After limiting to male-only subjects, the AUCs improved to 0.837 and 0.857 ever-smoking status was added [36]. An overview is presented in Table 1.

Prediction models consistently demonstrate improved discrimination with a GRS that includes HLA alleles compared with non-HLA alleles, reflecting the stronger association of HLA-SE with RA and improvements in discrimination and reclassification when including smoking and other environmental factors. However, the role of including autoantibodies or early symptoms such as arthralgias in these models has not been studied. While prediction models among FDRs have the highest discrimination, these models are not necessarily helpful for screening, as the background prevalence of RA is very low, estimated to be 3.6% for women and 1.7% for men, with a lifetime risk of RF + RA of 2.4% for women and 1.1% for men, with a 4- to 9-fold elevation of absolute risk among FDRs [2, 4, 37]. Thus the absolute risk conferred by positive family history, multiple environmental factors and high GRS may be too low to consider a prevention trial of medication that has significant side effects. An alternative high-risk population to target for prevention trials may be subjects with markers of systemic autoimmunity.

## Risk prediction using markers of systemic autoimmunity

The notion that RA can have a long preclinical phase was first recognized through studies in Finland showing that RF and anti-keratin antibodies precede the diagnosis of RA by many years [38, 39]. Subsequently it was shown that the anti-keratin test measured autoreactivity to proteins at the site of a post-translational modification of arginine to citrulline [40, 41]. For unknown reasons the formation of antibodies to citrulline is highly specific for RA and is likely to be involved in its pathogenesis [42]. To enhance the ability to engage autoantibodies in patients' serum, a CCP was developed that became known as the anti-CCP1 test [40, 41]. Later it became apparent that many citrullinated proteins can be the target of these antibodies, which as a group are now called ACPAs. The more sensitive second-generation anti-CCP2 test is based on several reactivities. With this test it was shown in blood donors who later developed RA that RF and ACPA can be found in serum of 28–34% (RF) and 34% (CCP2)–41% (CCP1) of later patients [43, 44]. Anti-CCP2 is positive in up to two-thirds of RA patients, and a similar frequency was noted in pre-RA blood donors shortly before the onset of symptoms [45]. There is no

evidence for a certain order of appearance of ACPA fine specificities, nor of a differential risk for RA of ACPA fine specificities [46–47].

Meanwhile, new autoantibodies have been detected that are related to RA and are also often present in the preclinical stage, such as anti-carbamylated protein antibodies (anti-CarP), anti-IgG hinge antibodies and anti-PAD-4 [31, 33, 50, 51]. Antibodies to *P. gingivalis* are found at increased levels in RA patients and FDRs [33], but they were not associated with arthritis development in seropositive arthralgia patients [52].

The development of RA during the preclinical period is characterized by the appearance of ACPA, anti-CarP and anti-PAD-4, later followed by RF and anti-IgG hinge. Over time, the number of ACPA reactivities increases as well as their concentrations, [45, 48, 49] and the ACPAs acquire a more pro-inflammatory profile [53, 54]. There are also elevations of cytokines, chemokines and acute phase reactants 2–12 years before RA onset [48, 54–60], with one cohort demonstrating that increasing numbers of cytokines and chemokines were predictive of decreased time to RA diagnosis [48, 59]. Among FDRs in SERA, higher numbers of cytokines were associated with RF positivity [61]. The timing of elevation in cytokines relative to the development of autoantibodies is unclear. Among FDRs in NANs, cytokines were elevated compared with controls, but RA probands, FDRs and controls had unique profiles [62]. However, elevated levels of acute phase reactants were found to occur simultaneously with the appearance of ACPA [55, 63].

How can these serological markers be used to predict RA? In asymptomatic blood donors, RF as a single autoantibody was not associated with risk of RA. Only ACPA-positive donors had an absolute risk of 5% of RA within 5 years, and those with a combination of RF and ACPA all developed RA within 5 years [44]. However, when these data were combined with data from families with two or more members with RA [64], it was calculated that an ACPA-positive family member would have a 69% chance of developing RA within 5 years, which underlines the value of combining hereditary and serological data.

The most obvious group to screen for autoantibodies would therefore be FDRs of RA probands. However, despite the 3- to 9-fold increased risk of RA in FDRs of RA patients [2], the yield of such screening is quite low [43, 65–67] and should perhaps be restricted to populations with a very high prevalence of RA, such as NANs, whose FDRs have a 8.5% rate of positive ACPA [62]. One reason for the low prevalence of autoantibodies in FDRs of RA patients is that the mean duration of the serological window before clinical RA is only 5 years. An alternative method would be to use environmental risk factors (such as smoking or being overweight [68]) and GRS to identify the highest-risk FDRs of RA probands and measure autoantibodies; however, this screening approach has yet to be tested.

A problem with the prediction of RA based on autoantibodies, apart from discarding seronegative RA, is that the antibody status is usually unknown before the disease is diagnosed. In the Amsterdam health care region, with >1 million inhabitants, seropositive RA is newly diagnosed ~300 times/year, whereas persons with arthralgia and a positive test for ACPA, a group that we actively look for, are detected ~35 times/year. After 3–4 years, 50% of these ACPA-positive persons develop RA. This means that in the Amsterdam region, ~17 of 300 (6%) new patients with seropositive RA are identified during the serological window. This figure may vary by region depending on the possibility for and inclination of general practitioners to perform an ACPA test. Screening the general population for ACPA would be very expensive given that the rate of ACPA positivity is <3% [67].

Not all people with ACPA develop RA, and most people that have been identified with ACPA also have some kind of symptoms, therefore it makes sense to make use of these symptoms to enhance predictive ability. This is dealt with in the next section. In addition, prediction models can be further improved by adding environmental risk factors, more biomarkers, such as genetic information, and cytokine or IFN activity [48, 69, 70].

## Risk prediction using symptoms

### Symptoms as a preclinical phase

The finding that symptoms might be present in the preclinical phase has been widely recognized in clinical practice. Many rheumatologists have encountered patients with recent-onset inflammatory arthritis who had suffered from joint pain prior to joint swelling. A recent qualitative study among patients with recent-onset RA evaluated the type of symptoms prior to RA diagnosis [71]. RA patients recalled consistent levels of pain at symptom onset that progressed to more intense levels before diagnosis, often associated with joint stiffness, tingling or burning sensations [71]. Although the location of the symptoms can migrate, symptoms often involved the hands [71]. Patients presenting with joint pain are common in primary and secondary care settings, and the diagnostic and prognostic value of these symptoms is yet to be explored. Since the clinical presentation influences the actions taken by physicians, prospective studies including thorough investigations of the predictive value of clinical symptoms are warranted to maximize the benefit from a patient's clinical history.

### Which symptoms are characteristic of RA in a preclinical phase?

No large-scale prospective studies have explored which complex of symptoms is characteristic of RA in a preclinical phase. Relying on the term inflammatory arthralgia is controversial. This term means that the arthralgia is caused by inflammation, but scientific data on the types of symptoms that constitute inflammatory arthralgia are lacking; there is also no uniformly accepted definition for this term. However, there are data indicating that the type of arthralgia is relevant for the outcome. In patients without symptoms but with ACPA, the absolute risk of RA is estimated to be 5.3% or 16% [43, 44]. In a study with

patients with ACPA and/or RF and arthralgia of any type, 20% progressed to RA during a median follow-up of 28 months [72]. Importantly, among the subgroup of ACPA-positive or RF-positive patients who had symmetric arthralgia in small joints and morning stiffness, 60% progressed towards RA [72]. Although the latter subgroup was small (6 of 10 patients developed RA), these data suggest that when RA-related autoantibodies are present, both the presence and the type of symptoms are helpful to identify patients at risk for RA. This notion is supported further by the prediction rule that was developed by van de Stadt *et al*. [47] for patients with RA-related autoantibodies and arthralgia, which assigns one point to each symptom type: symptoms $<$12 months, intermittent symptoms, arthralgia in the upper and lower extremities, morning stiffness $\geqslant$1 h and self-reported joint swelling. Of all the ACPA-positive/RF-positive patients included in this study, 35% developed arthritis after a median follow-up of 12 months; of the patients in the highest risk category ($\geqslant$7 points), this was 43% [47]. Thus, with autoantibody positivity, the type of symptoms is relevant to consider for risk stratification. Another recent study of ACPA-positive arthralgia patients produced a model for progression to inflammatory arthritis. This model consists of four variables, with two variables related to symptoms (tenderness of hand or foot joints and morning stiffness $\geqslant$30 min) [73]. For use in rheumatology or general practice it is helpful to be able to discriminate the type of arthralgia found in those who later develop RA before ordering autoantibody testing: first, because patients present with symptoms, and second, because 70% of early unclassified arthritis patients and 40% of early RA patients are autoantibody negative [74]. These patients are missed when the presence of autoantibodies is used as the starting point for risk stratification.

In order to further elucidate the symptoms that are characteristic of the preclinical phase of RA, other longitudinal studies are necessary. Examples are the prospective studies on FDRs of RA patients designed in the USA and Canada [28, 29] as well as the clinically suspect arthralgia (CSA) cohort that was initiated in Leiden, The Netherlands [75]. Patients with CSA have no clinically detectable arthritis, but have arthralgia of the small joints of recent onset that, according to the judgement of their rheumatologist, is suspected to progress towards RA over time. Results of laboratory investigations are not required. The presence of symptoms or signs consistent with non-RA diagnoses rules out CSA. Since the type of symptoms that are characteristic of imminent RA are not yet known, this approach of including the rheumatologist's expert opinion was chosen. The proportion of arthralgia patients that is diagnosed as CSA is rather small. For many patients, rheumatologists identified non-RA symptoms, and our data revealed that only 7% of patients with unexplained arthralgia had CSA. The main reasons provided by rheumatologists to attribute recent-onset arthralgia of small joints to a risk of RA development were joint pain that was worst in the early morning and improved with movement during the day, the presence of morning stiffness of $\geqslant$60 min and/or a positive family history of RA. We observed that almost half of the CSA patients had subclinical inflammation on 1.5 T extremity MRI, that about one-third of the latter patients progressed towards RA during the 4 months of follow-up and that only 7% of the CSA patients that progressed to RA had no subclinical inflammation at the baseline MRI. Therefore the presence of CSA and subclinical joint inflammation may be a valuable method of detecting patients in a preclinical phase. More research is needed to better define the clinical items relevant for classifying a patient as CSA and more follow-up is needed to determine the prognostic relevance.

### How to identify RA patients in a preclinical phase using symptoms as starting point?

Supposing that adequate risk stratification for patients with arthralgia is possible, a subsequent question is how to identify patients that are developing RA in a preclinical phase? Villeneuve *et al.* [76] reviewed strategies for promoting early referral and reducing delays in diagnosing inflammatory arthritis and showed that primary care educational programmes and rapid access clinics were most efficacious [76]. An example is the immediate access clinic that was started in Vienna, Austria, where a preliminary triage decision was made in a short visit [77]. This led to a substantial reduction in wait time. Importantly, $>$75% of the diagnoses of suspect inflammatory rheumatological diseases were correct, illustrating the usefulness of clinical experience in risk stratification. A slightly different approach are the early arthritis recognition clinics that have been initiated in several places in The Netherlands, aimed at reducing general practitioner delay in patient referral [77–79]. Here, the frequency of patients presenting with symptoms for $<$3 months increased from 31% to 62% [78]. The method that is most optimal to identify symptomatic patients in a preclinical phase of RA will depend on the health care system and other local factors. Nonetheless, approaches like these, including educational programmes and easy access facilities, will be relevant to identify patients while they are in a preclinical disease phase.

## Intervention in preclinical phases of RA

Identifying individuals at risk for RA is beneficial since it has been shown that intervention in the preclinical phase results in better outcomes than when intervening in clinically apparent arthritis. The first randomized prevention trials had negative results [80, 81], but more trials in patients with autoantibody-positive arthralgia are being conducted. Once we have an effective intervention, screening strategies will need to show cost-effectiveness and the ability to target subjects with the highest absolute risk.

## Conclusion

Most RA patients pass through several phases of increased risk, from genetic risk through exposure, to environmental, lifestyle and behavioural factors, to autoimmunity and elevated cytokines, concluding with a

TABLE 2 Prediction models for risk of seropositive or ACPA-positive RA including genetic risk scores and environmental factors

| Cohort | Outcome | AUC[a] | GRS epidemiological factors |
|---|---|---|---|
| Phase A and B, genetic and environmental risk factors | | | |
| Nurses' Health Study | Seropositive RA: 289 cases/481 controls | 0.66 (F) | 8 HLA + 14 SNPs |
| EIRA [19] | ACPA-positive RA: 629 cases/ 623 controls | 0.75 (M/F)[b] | Age, smoking |
| EHR cohort [20] | ACPA-positive RA: 871 cases/ 1229 controls | 0.71 (M/F) | 1 HLA + 28 SNPs |
| | ACPA-negative RA: 378 cases/ 1229 controls | 0.55 (M/F) | Ancestry |
| Nurses' Health Study | Seropositive RA: 317 cases/551 controls | 0.716 (F) | 8 HLA + 31 SNPs |
| EIRA [83] | ACPA-positive RA: 987 cases/ 958 controls | 0.716 (F)[b] 0.756 (M) | Age, smoking, alcohol, education, parity (F only) HLA × smoking interaction |
| Nurses' Health Study | Seropositive RA: 317 cases/551 controls | 0.738 (F) | 8 HLA + 31 SNPs + GSTT1 + HMOX1 |
| EIRA [83] | ACPA-positive RA: 987 cases/ 958 controls | 0.724 (F)[b] 0.769 (M) | Age, smoking, alcohol, education, reproductive factors (F only), occupational factors (M only), region HLA × smoking, GSTT1 × smoking, HMOX1 × smoking interactions, silica × smoking, solvents × smoking, mineral oil × smoking interactions (M only) |
| European cohorts [23] | All RA: 11 366 cases/15 489 controls | 0.738 (M/F) | GRS-45 + HLA amino acids |
| | ACPA-positive RA: 6370 cases/ 15 489 controls | 0.801 (M/F) | Gender, smoking |
| Wellcome Trust Case Control Consortium | ACPA-positive RA: 1516 cases/ 1476 controls | 0.796(M/F) | 15 HLA alleles + 31 SNPs |
| UK RA Genetics Group Consortium [36] | ACPA-positive RA: 294 cases/ 573 controls | 0.756(M/F)[b] | |
| Wellcome Trust Case Control Consortium | ACPA-positive RA: 239 cases/ 739 controls | 0.837 (M) | 15 HLA alleles + 31 SNPs |
| UK RA Genetics Group Consortium [36] | ACPA-positive RA: 294 cases/ 573 controls | 0.857 (M)[b] | Smoking |
| Phases C and D, RA-related autoantibodies and symptoms | | | |
| Amsterdam Reade RF and/or ACPA-positive arthralgia cohort [47] | Arthritis: 131 cases/243 controls All RA: 121 cases/253 controls | 0.82 0.79[c] | Family history, shared epitope, alcohol use, symptoms, antibody status |

HER: electronic health records; EIRA: epidemiologic investigation in RA; GRS-45: genetic risk score using 45 SNPs; SNPs: single nucleotide polymorphisms; GSTT1: glutathione S-transferase theta 1; HMOX1: heme oxygenase (decycling) 1. [a]AUC: area under the receiver operating curve (ROC) analysis. [b]AUC from replication in independent population. [c]AUC from internal cross-validation.

relatively short period of symptoms. The two groups with clearly increased risk of RA that have been identified are FDRs of RA patients and persons with RA-like symptoms, with or without autoantibodies. Persons from both groups may want to know their absolute risk of RA. For patients with a positive family history, questions about lifetime or absolute risks are difficult to answer since population studies provide standardized incidence ratios [2] and cohort and case–control studies provide estimates of risk as relative risks or ORs, not as absolute risk [19, 82]. Although the developed prediction models based on family history and genetic variants as risk factors have a discriminative ability (Table 2), the absolute risk among FDRs is low, and family history is negative in the majority of RA patients, making screening for RA risk challenging. Screening for autoantibodies is costly given the low prevalence in the general population, although subjects with positive autoantibodies have high absolute risk. In subjects with arthralgia and autoantibodies, the risk of RA can be estimated using the prediction rule developed by van de Stadt *et al*. [47]. This prediction rule still needs to be validated in other cohorts. Also, additional biomarkers that have been shown to be associated with RA risk need to be incorporated into such models to achieve their full potential. Furthermore, ~40% of RA patients are autoantibody negative at the onset of clinical arthritis; these RA patients are also not detected with this prediction rule.

Thus screening should be targeted to groups with a large enough absolute risk to warrant intervention. We recommend that in future analyses of predictive models,

**Table 3** A research agenda for the prediction of RA in persons at risk

| |
|---|
| **Phase A and B, genetic and environmental risk factors** |
| Which combination of genes or genes and environmental risk factors provide high absolute risk of RA in the population that can be used in selecting high-risk individuals for autoantibody, cytokine or other biomarker testing? |
| To what extent does the chance of a false positive test affect decisions regarding preventive therapy? |
| Do prediction models in the general population predict the development of arthralgia and inflammatory arthritis among first-degree relative cohorts? |
| **Phase C, RA-related autoantibodies** |
| Are the RA-related autoantibodies pathogenic and if so, what determines that some persons with RA-related autoantibodies do not get RA whereas in others it is a preclinical phase of the disease? |
| What is the absolute risk of RA in the presence of (a combination of) RA-related autoantibodies in the general population (to be determined in prospective population studies) in first-degree relatives or among subjects with arthralgias? |
| Should prevention interventions be targeted at autoantibody-positive subjects before the development of arthralgias or inflammatory arthritis? |
| **Phase D, symptoms** |
| Which proportion of RA patients has a discernable preclinical phase of symptoms without clinically apparent arthritis? |
| Which complex of symptoms is characteristic for a preclinical phase of RA? Which symptoms or signs determine the clinical expertise/clinical impression of rheumatologists to discriminate patients in a preclinical phase of RA from those with non-RA-related symptoms? |
| Which symptom complexes presenting in primary care should prompt additional activities, such as referral to a rheumatologist or autoantibody testing? |
| To what extent can a combination of genetic factors, environmental factors, autoantibodies or imaging characteristics improve the predictive ability compared with risk stratification based on symptoms alone? |
| Validate the prediction rule for patients with ACPA-positive arthralgia in independent cohort studies. |
| Improve the prediction rule for patients with ACPA-positive arthralgia by incorporating new biomarkers with predictive ability for RA. |
| **Phases A, B, C and D** |
| What is the threshold of absolute RA risk that individuals without any RA symptoms (phases A–C) or with symptoms (phase D) would consider high enough to outweigh potential side effects of preventive treatment? |
| How do high-risk individuals view the risk–benefit of preventive treatments in the spectrum of lifestyle–behavioural interventions through high-dose immunosuppressive drugs? |

metrics of model fit, discrimination and reclassification (these are described in Table 1) should be assessed. There are also several areas in risk prediction that are in need of further research; some of these are summarized in Table 3. An important issue is that research regarding methods to communicate absolute risk to high-risk individuals is needed. During the next decade we will hopefully develop high-quality, inexpensive predictive tests for the assessment of RA risk and preventive treatments. Since the preclinical symptomatic phase is relatively short, some RA patients may be missed in this phase. This increases the importance of modifying already identified lifestyle factors (such as smoking cessation, reduction of overweight and the use of small quantities of alcoholic beverages). Especially for relatives of RA patients, lifestyle interventions should receive more attention.

## Acknowledgements

E. W.K. is supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (grants AR049880, AR052403 and AR047782). A. H. v.d.H.-v.M.'s work is supported by a Vidi grant from the Netherlands Organisation for Scientific Research.

*Funding*: No specific funding was received from any funding bodies in the public, commercial or not-for-profit sectors to carry out the work described in this article.

*Disclosure statement*: The authors have declared no conflicts of interest.

## References

1 Gerlag DM, Raza K, van Baarsen LG *et al*. EULAR recommendations for terminology and research in individuals at risk of rheumatoid arthritis: report from the Study Group for Risk Factors for Rheumatoid Arthritis. Ann Rheum Dis 2012;71:638–41.

2 Hemminki K, Li X, Sundquist J, Sundquist K. Familial associations of rheumatoid arthritis with autoimmune diseases and related conditions. Arthritis Rheum 2009;60:661–8.

3 Silman AJ, MacGregor AJ, Thomson W *et al*. Twin concordance rates for rheumatoid arthritis: results

from a nationwide study. Br J Rheumatol 1993;32:903–7.

4 Frisell T, Holmqvist M, Kallberg H *et al*. Familial risks and heritability of rheumatoid arthritis: role of rheumatoid factor/anti-citrullinated protein antibody status, number and type of affected relatives, sex, and age. Arthritis Rheum 2013;65:2773–82.

5 Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum 1987;30:1205–13.

6 Raychaudhuri S, Sandor C, Stahl EA *et al*. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet 2013;44:291–6.

7 Eyre S, Bowes J, Diogo D *et al*. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet 2013;44:1336–40.

8 Okada Y, Wu D, Trynka G *et al*. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506:376–81.

9 Sugiyama D, Nishimura K, Tamaki K *et al*. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. Ann Rheum Dis 2010;69:70–81.

10 Criswell LA, Merlino LA, Cerhan JR *et al*. Cigarette smoking and the risk of rheumatoid arthritis among postmenopausal women: results from the Iowa Women's Health Study. Am J Med 2002;112:465–71.

11 Costenbader KH, Feskanich D, Mandl LA, Karlson EW. Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. Am J Med 2006;119:503–11.

12 Mikuls TR, Sayles H, Yu F *et al*. Associations of cigarette smoking with rheumatoid arthritis in African Americans. Arthritis Rheum 2010;62:3560–8.

13 Kallberg H, Ding B, Padyukov L *et al*. Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke. Ann Rheum Dis 2011;70:508–11.

14 Karlson EW, Deane K. Environmental and gene-environment interactions and risk of rheumatoid arthritis. Rheum Dis Clin North Am 2012;38:405–26.

15 Lahiri M, Morgan C, Symmons DP, Bruce IN. Modifiable risk factors for RA: prevention, better than cure? Rheumatology 2012;51:499–512.

16 Klareskog L, Stolt P, Lundberg K *et al*. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. Arthritis Rheum 2006;54:38–46.

17 Karlson EW, Chang SC, Cui J *et al*. Gene-environment interaction between HLA-DRB1 shared epitope and heavy cigarette smoking in predicting incident rheumatoid arthritis. Ann Rheum Dis 2010;69:54–60.

18 Criswell LA, Saag KG, Mikuls TR *et al*. Smoking interacts with genetic risk factors in the development of rheumatoid arthritis among older Caucasian women. Ann Rheum Dis 2006;65:1163–7.

19 Karlson EW, Chibnik LB, Kraft P *et al*. Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. Ann Rheum Dis 2010;69:1077–85.

20 Kurreeman F, Liao K, Chibnik L *et al*. Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. Am J Hum Genet 2010;88:57–69.

21 Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011;30:11–21.

22 Karlson EW, Ding B, Keenan BT *et al*. Association of environmental and genetic factors and gene-environment interactions with risk of developing rheumatoid arthritis. Arthritis Care Res 2013;65:1147–56.

23 Yarwood A, Han B, Raychaudhuri S *et al*. A weighted genetic risk score using all known susceptibility variants to estimate rheumatoid arthritis risk. Ann Rheum Dis 2013, Oct 3. doi: 10.1136/annrheumdis-2013-204133 [Epub ahead of print].

24 Nagelkerke N. A note on a general definition of the coefficient of determination. Biometrika 1991;78:691–2.

25 Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med 1997;16:965–80.

26 Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157–72; discussion 207–12.

27 Pencina MJ, D'Agostino RB Sr, Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. Stat Med 2011;31:101–13.

28 Kolfenbach JR, Deane KD, Derber LA *et al*. A prospective approach to investigating the natural history of preclinical rheumatoid arthritis (RA) using first-degree relatives of probands with RA. Arthritis Rheum 2009;61:1735–42.

29 El-Gabalawy HS, Robinson DB, Hart D *et al*. Immunogenetic risks of anti-cyclical citrullinated peptide antibodies in a North American Native population with rheumatoid arthritis and their first-degree relatives. J Rheumatol 2009;36:1130–5.

30 Bhatia SS, Majka DS, Kittelson JM *et al*. Rheumatoid factor seropositivity is inversely associated with oral contraceptive use in women without rheumatoid arthritis. Ann Rheum Dis 2007;66:267–9.

31 Kolfenbach JR, Deane KD, Derber LA *et al*. Autoimmunity to peptidyl arginine deiminase type 4 precedes clinical onset of rheumatoid arthritis. Arthritis Rheum 2010;62:2633–9.

32 Peschken CA, Robinson DB, Hitchon CA *et al*. Pregnancy and the risk of rheumatoid arthritis in a highly predisposed North American Native population. J Rheumatol 2012;39:2253–60.

33 Hitchon CA, Chandad F, Ferucci ED *et al*. Antibodies to *Porphyromonas gingivalis* are associated with anticitrullinated protein antibodies in patients with rheumatoid arthritis and their relatives. J Rheumatol 2010;37:1105–12.

34 Ferucci ED, Darrah E, Smolik I *et al*. Prevalence of anti-peptidylarginine deiminase type 4 antibodies in rheumatoid arthritis and unaffected first-degree relatives in indigenous North American populations. J Rheumatol 2013;40:1523–8.

35 Sparks JA, Chen CY, Jiang X *et al*. Improved performance of epidemiologic and genetic risk models for rheumatoid arthritis serologic phenotypes using family history. Ann Rheum Dis 2014, Mar.

36 Scott IC, Seegobin SD, Steer S *et al*. Predicting the risk of rheumatoid arthritis and its age of onset through modelling genetic risk variants with smoking. PLoS Genet 2013;9: e1003808.

37 Crowson CS, Matteson EL, Myasoedova E *et al*. The lifetime risk of adult-onset rheumatoid arthritis and other inflammatory autoimmune rheumatic diseases. Arthritis Rheum 2013;63:633–9.

38 Aho K, Heliovaara M, Maatela J, Tuomi T, Palosuo T. Rheumatoid factors antedating clinical rheumatoid arthritis. J Rheumatol 1991;18:1282–4.

39 Kurki P, Aho K, Palosuo T, Heliovaara M. Immunopathology of rheumatoid arthritis. Antikeratin antibodies precede the clinical disease. Arthritis Rheum 1992;35:914–7.

40 Schellekens GA, de Jong BA, van den Hoogen FH, van de Putte LB, van Venrooij WJ. Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. J Clin Invest 1998; 101:273–81.

41 Schellekens GA, Visser H, de Jong BA *et al*. The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. Arthritis Rheum 2000;43:155–63.

42 Kuhn KA, Kulik L, Tomooka B *et al*. Antibodies against citrullinated proteins enhance tissue injury in experimental autoimmune arthritis. J Clin Invest 2006;116: 961–73.

43 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;48:2741–9.

44 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;50:380–6.

45 van de Stadt LA, de Koning MH, van de Stadt RJ *et al*. Development of the anti-citrullinated protein antibody repertoire prior to the onset of rheumatoid arthritis. Arthritis Rheum 2011;63:3226–33.

46 Brink M, Hansson M, Mathsson L *et al*. Multiplex analyses of antibodies against citrullinated peptides in individuals prior to development of rheumatoid arthritis. Arthritis Rheum 2013;65:899–910.

47 van de Stadt LA, Witte BI, Bos WH, van Schaardenburg D. A prediction rule for the development of arthritis in seropositive arthralgia patients. Ann Rheum Dis 2013;72: 1920–6.

48 Sokolove J, Bromberg R, Deane KD *et al*. Autoantibody epitope spreading in the pre-clinical phase predicts progression to rheumatoid arthritis. PLoS One 2012;7: e35296.

49 Arkema EV, Goldstein BL, Robinson W *et al*. Anti-citrullinated peptide autoantibodies, human leukocyte antigen shared epitope and risk of future rheumatoid arthritis: a nested case-control study. Arthritis Res Ther 2013;15: R159.

50 Shi J, van de Stadt LA, Levarht EW *et al*. Anti-carbamylated protein (anti-CarP) antibodies precede the onset of rheumatoid arthritis. Ann Rheum Dis 2013;73:780–3.

51 van de Stadt LA, de Vrieze H, Derksen NI *et al*. Antibodies to IgG4 hinge can be found in rheumatoid arthritis patients during all stages of disease and may exacerbate chronic antibody-mediated inflammation. Arthritis Rheum 2014, Dec 24. doi: 10.1002/art.38335. [Epub ahead of print].

52 de Smit M, van de Stadt LA, Janssen KM *et al*. Antibodies against *Porphyromonas gingivalis* in seropositive arthralgia patients do not predict development of rheumatoid arthritis. Ann Rheum Dis 2014;73:1277–9.

53 Suwannalai P, van de Stadt LA, Radner H *et al*. Avidity maturation of anti-citrullinated protein antibodies in rheumatoid arthritis. Arthritis Rheum 2012; 64:1323–8.

54 Rombouts Y, Ewing E, van de Stadt LA *et al*. Anti-citrullinated protein antibodies acquire a pro-inflammatory Fc glycosylation phenotype prior to the onset of rheumatoid arthritis. Ann Rheum Dis 2013, Oct 8. doi: 10.1136/ annrheumdis-2013-203565. [Epub ahead of print].

55 Nielen MM, van Schaardenburg D, Reesink HW *et al*. Increased levels of C-reactive protein in serum from blood donors before the onset of rheumatoid arthritis. Arthritis Rheum 2004;50:2423–7.

56 Rantapaa-Dahlqvist S, Boman K, Tarkowski A, Hallmans G. Up regulation of monocyte chemoattractant protein-1 expression in anti-citrulline antibody and immunoglobulin M rheumatoid factor positive subjects precedes onset of inflammatory response and development of overt rheumatoid arthritis. Ann Rheum Dis 2007;66: 121–3.

57 Jorgensen KT, Wiik A, Pedersen M *et al*. Cytokines, autoantibodies and viral antibodies in premorbid and postdiagnostic sera from patients with rheumatoid arthritis: case-control study nested in a cohort of Norwegian blood donors. Ann Rheum Dis 2008;67:860–6.

58 Karlson EW, Chibnik LB, Tworoger SS *et al*. Biomarkers of inflammation and development of rheumatoid arthritis in women from two prospective cohort studies. Arthritis Rheum 2009;60:641–52.

59 Deane KD, O'Donnell CI, Hueber W *et al*. The number of elevated cytokines and chemokines in preclinical seropositive rheumatoid arthritis predicts time to diagnosis in an age-dependent manner. Arthritis Rheum 2010;62: 3161–72.

60 Kokkonen H, Soderstrom I, Rocklov J *et al*. Up-regulation of cytokines and chemokines predates the onset of rheumatoid arthritis. Arthritis Rheum 2010;62: 383–91.

61 Hughes-Austin JM, Deane KD, Derber LA *et al*. Multiple cytokines and chemokines are associated with rheumatoid arthritis-related autoimmunity in first-degree relatives without rheumatoid arthritis: Studies of the Aetiology of

Rheumatoid Arthritis (SERA). Ann Rheum Dis 2013;72:901–7.

62 El-Gabalawy HS, Robinson DB, Smolik I *et al*. Familial clustering of the serum cytokine profile in the relatives of rheumatoid arthritis patients. Arthritis Rheum 2012;64:1720–9.

63 Nielen MM, van Schaardenburg D, Reesink HW *et al*. Simultaneous development of acute phase response and autoantibodies in preclinical rheumatoid arthritis. Ann Rheum Dis 2006;65:535–7.

64 Silman AJ, Hennessy E, Ollier B. Incidence of rheumatoid arthritis in a genetically predisposed population. Br J Rheumatol 1992;31:365–8.

65 Deane KD, Striebich CC, Goldstein BL *et al*. Identification of undiagnosed inflammatory arthritis in a community health fair screen. Arthritis Rheum 2009;61:1642–9.

66 Arlestig L, Mullazehi M, Kokkonen H *et al*. Antibodies against cyclic citrullinated peptides of IgG, IgA and IgM isotype and rheumatoid factor of IgM and IgA isotype are increased in unaffected members of multicase rheumatoid arthritis families from northern Sweden. Ann Rheum Dis 2012;71:825–9.

67 Young KA, Deane KD, Derber LA *et al*. Relatives without rheumatoid arthritis show reactivity to anti-citrullinated protein/peptide antibodies that are associated with arthritis-related traits: Studies of the Etiology of Rheumatoid Arthritis. Arthritis Rheum 2013;65:1995–2004.

68 de Hair MJ, Landewe RB, van de Sande MG *et al*. Smoking and overweight determine the likelihood of developing rheumatoid arthritis. Ann Rheum Dis 2012;72:1654–8.

69 van Baarsen LG, Bos WH, Rustenburg F *et al*. Gene expression profiling in autoantibody-positive patients with arthralgia predicts development of arthritis. Arthritis Rheum 2010;62:694–704.

70 Lubbers J, Brink M, van de Stadt LA *et al*. The type I IFN signature as a biomarker of preclinical rheumatoid arthritis. Ann Rheum Dis 2013;72:776–80.

71 Stack RJ, van Tuyl LH, Sloots M *et al*. Symptom complexes in patients with seropositive arthralgia and in patients newly diagnosed with rheumatoid arthritis: a qualitative exploration of symptom development. Rheumatology 2014, Apr 11. [Epub ahead of print].

72 Bos WH, Wolbink GJ, Boers M *et al*. Arthritis development in patients with arthralgia is strongly associated with anti-citrullinated protein antibody status: a prospective cohort study. Ann Rheum Dis 2010;69:490–4.

73 Rakieh C, Nam JL, Hunt L *et al*. Predicting the development of clinical arthritis in anti-CCP positive individuals with non-specific musculoskeletal symptoms: a prospective observational cohort study. Ann Rheum Dis 2014, Apr 12. doi: 10.1136/annrheumdis-2014-205227 [Epub ahead of print].

74 de Rooy DP, van der Linden MP, Knevel R, Huizinga TW, van der Helm-van Mil AH. Predicting arthritis outcomes—what can be learned from the Leiden Early Arthritis Clinic? Rheumatology 2011;50:93–100.

75 van Steenbergen HW, van Nies JA, Huizinga TW *et al*. Characterising arthralgia in the preclinical phase of rheumatoid arthritis using MRI. Ann Rheum Dis 2014, in press.

76 Villeneuve E, Nam JL, Bell MJ *et al*. A systematic literature review of strategies promoting early referral and reducing delays in the diagnosis and management of inflammatory arthritis. Ann Rheum Dis 2013;72:13–22.

77 Gartner M, Fabrizii JP, Koban E *et al*. Immediate access rheumatology clinic: efficiency and outcomes. Ann Rheum Dis 2012;71:363–8.

78 van Nies JA, Brouwer E, van Gaalen FA *et al*. Improved early identification of arthritis: evaluating the efficacy of early arthritis recognition clinics. Ann Rheum Dis 2013;72:1295–301.

79 van Nies JA, Brouwer E, de Rooy DP *et al*. Reasons for medical help-seeking behaviour of patients with recent-onset arthralgia. Ann Rheum Dis 2013;72:1302–7.

80 Karlson EW, Shadick NA, Cook NR, Buring JE, Lee IM. Vitamin E in the primary prevention of rheumatoid arthritis: the Women's Health Study. Arthritis Rheum 2008;59:1589–95.

81 Bos WH, Dijkmans BA, Boers M, van de Stadt RJ, van Schaardenburg D. Effect of dexamethasone on autoantibody levels and arthritis development in patients with arthralgia: a randomised trial. Ann Rheum Dis 2010;69:571–4.

82 Kraft P, Wacholder S, Cornelis MC *et al*. Beyond odds ratios—communicating disease risk based on genetic profiles. Nat Rev Genet 2009;10:264–9.

83 Karlson EW, Ding B, Keenan BT *et al*. Association of environmental and genetic factors and gene-environment interactions with risk of developing rheumatoid arthritis. Arthritis Care Res 2013;65:1147–56.