Predictive factors for outcome of rheumatoid arthritis

Linden, M.P.M. van der

CHAPTER 4

Towards a data-driven evaluation of the 2010 ACR/EULAR criteria for rheumatoid arthritis: Is it sensible to look at levels of rheumatoid factor?

M.P.M. van der Linden[1]
M.R. Batstra[2]
L.E. Bakker-Jonges[2]
J. Detert[3]
H. Bastian[3]
H.U. Scherer[1,3]
R.E.M. Toes[1]
G.R. Burmester[3]
M.D. Mjaavatten[4]
T.K. Kvien[4]
T.W.J. Huizinga[1]
A.H.M. van der Helm-van Mil[1]

[1] *Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands*
[2] *Department of Medical Laboratories, Reinier de Graaf Group, Delft, The Netherlands, on behalf of SKML-HIM (WGHAS)*
[3] *Department of Rheumatology and Clinical Immunology, Charité-University Medicine Berlin, Berlin, Germany*
[4] *Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway*

Arthritis Rheum 2011; 63 (5): 1190-1199

ABSTRACT

Objective

Recently new classification criteria for Rheumatoid Arthritis (RA) have been devised by methodology that used first a quantitative approach (data from databases), then a qualitative approach (consensus-based on paper patients) and finally a common sense based approach (evaluation of the former phases). Now these criteria are being evaluated to assess characteristics of the individual items. This study analyzed characteristics of the item autoantibodies, in particular RF level.

Methods

Three separate cohorts with a total of 972 undifferentiated arthritis patients were studied for RA development (according to the 1987 ACR criteria) and arthritis persistency. Positive and negative predictive values (PPV, NPV) and likelihood ratios (LR) were compared between different levels of RF and the presence of ACPA. A similar comparison was made in 686 RA patients for the rate of joint destruction during 7 years of follow-up and achievement of sustained DMARD-free remission. The variation in RF levels obtained by different measurement methods in the same RF-positive sera was explored.

Results

Presence of ACPA had a better balance between LR+/LR- and PPV/NPV than high RF levels for RA development. The additive value of ACPA assessment after high level RF testing was higher than vice versa. High level RF was less strongly associated with RA severity than ACPA antibodies. The RF level obtained by different methods in the same patients' sera varied considerably.

Conclusion

Level determination of RF is subject to large variation; high level RF has limited additive prognostic value compared to ACPA positivity. Thus, omitting RF level and using RF presence, ACPA presence and ACPA level may improve the 2010 criteria for RA.

INTRODUCTION

Recently, the American College of Rheumatism (ACR) classification criteria for rheumatoid arthritis (RA) dating back from 1987,[1] have been subjected to a process of rejuvenation by a joint taskforce of both the ACR and the EULAR (European League Against Rheumatism). The aim of these criteria is to classify RA in an earlier disease stage compared to the 1987 ACR criteria and the development of these criteria is an important step forwards.

The development of the 2010 ACR/EULAR criteria comprised three phases. First, a data driven phase using data from 3115 patients from Europe and Canada. Next, a phase incorporating the expertise of 39 rheumatologists and finally a consensus phase by the same group.[2-4] It is foreseen that in the next years the criteria will be studied in cohorts with different ethnic backgrounds and dissimilar healthcare systems in which the pretest probability for RA in new patients visiting rheumatologists differs.

The 2010 criteria are the first that include anti-citrullinated peptide antibodies (ACPA), in addition to RF. Presence of these auto-antibodies can contribute substantially to the classification of RA for which ≥6 points are required; presence of ACPA or RF yields 2 points and high levels of ACPA or RF yields 3 points. In the data driven phase of the development of the criteria, using data of several early arthritis cohorts, ACPA and RF were recognized as a theme in a factor analysis. Then, ACPA and RF were summarized as 'serology'. Subsequently the importance of serology, independent of other variables, was determined using a multivariate regression analysis. It was observed that within the patients with a positive serology, patients with a level higher than median received a higher weight than patients with a level lower than median. After the expert-phase and consensus-phase a high level was redefined as ≥ three times the reference value.

The present study aimed to provide two main characteristics of the items serology, particularly the RF level criterion, in the 2010 ACR/EULAR criteria for RA. The first characteristic was the discriminative ability of high levels of RF compared to ACPA for early RA. Several studies observed an increased specificity for RA of a higher RF level compared to RF positivity.[5,6] However, an increased specificity for RA has also been observed for presence of ACPA compared to the presence of RF.[7] Thus far extensive comparisons of the prognostic performance for RA development of increased RF levels in comparison to the presence of ACPA, notably anti-CCP antibodies, have not been made. In three separate prospective cohorts with undifferentiated arthritis (UA) patients of recent onset from three different countries, RA development was studied in relation to baseline RF levels and ACPA. RA was defined by the 1987 ACR criteria.[1] To verify that the results were not different when other outcome measures were used, analyses in UA patients were repeated with arthritis persistency as outcome. Furthermore, in RA patients the same analyses were performed with the rate of joint destruction and the achievement of sustained disease modifying antirheumatic drugs (DMARD)-free remission as outcome.

The second characteristic was the capacity of different assays to uniformly define a high RF level. Despite the presence of international units for RF, RF level measurement is not adequately

standardized between different measurement methods. Subsequent variations in RF levels may yield differences in classifying or diagnosing RA between laboratories. Therefore we determined the degree of variation in RF levels obtained when the same RF-positive serum samples were tested by the methods that are currently most frequently applied (ELISA, nephelometry, turbidimetry). Although older studies evaluated the correlations between results of the Rose-Waaler method and ELISA,[8] data on a head-to-head comparisons of currently applied methods are to the best of our knowledge not available.

PATIENTS AND METHODS

Patients

Development of RA in UA patients

UA patients of three separate cohorts were studied for RA development, comprising an overall total of 972 UA patients (Figure 1). UA was defined as not fulfilling any of the existing classification criteria for a rheumatic disease diagnosis 2 weeks after the first presentation when the results of laboratory and radiological examinations were known.[9] Patients were followed up for one year, where after the final diagnosis was established. Patients were categorized as RA (according to the 1987 ACR criteria)[1] or as non-RA (all other diagnoses).

The Leiden EAC is a large prospective cohort that started in 1993, which has been described previously.[10] Patients with confirmed arthritis were included when the symptom duration was less than 2 years. At baseline, blood samples were taken for routine diagnostic laboratory screening (including testing for IgM-RF) and stored for determining other auto-antibodies later on (anti-CCP2). Follow-up visits (including radiographs) were performed yearly. Between 1993 and 2006, 625 patients were diagnosed with UA at baseline. Almost all patients had a follow-up duration longer than one year and 30% of the UA patients had developed RA after one year and 4% later than one year of follow-up.[11]

The Berlin EAC was started in January 2004, and patients were included if they had synovitis in at least 2 joints and a duration of symptoms between 4 weeks and 12 months. This Berlin cohort has been described previously.[12] At first presentation, 154 patients had UA. Fulfillment of the 1987 ACR criteria[1] for RA was assessed after 1 year of follow-up.

The third cohort consisted of 193 UA patients from Oslo, Norway, included in the Norwegian very early arthritis (NOR-VEAC).[13] This cohort included patients with at least one swollen joint of 16 weeks duration. During the first year patients were seen after 3, 6, and 12 months and the development of RA was classified after one year of follow-up.

In the first, data driven phase of developing the new 2010 ACR/EULAR criteria, patients from the Leiden EAC (n=213) and from the NORVEAC (n=193) were used.[3] All studies were approved by the local ethics committees. All patients gave their written informed consent.

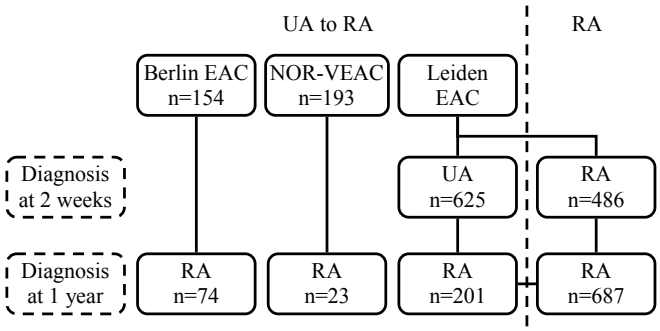

Figure 1. Flowchart of the cohorts used in this study. Left Part: Patients initially diagnosed with UA were studied for development of RA. For the Berlin data, all of the 154 UA patients had data on RF as well as ACPA. Also in the NOR-VEAC, both RF and ACPA were determined in all 193 UA patients. For the Leiden EAC, data was available for RF in 623 and for ACPA in 624 out of 625 UA patients. Right Part: In the Leiden EAC, a total of 687 patients were diagnosed with RA after one year. In these patients 686 had data on radiographic data and/or the achievement of sustained DMARD-free remission. RF and ACPA were in measured in 663 and 658 patients respectively

Arthritis persistency in UA patients

In order to determine whether results differed when another outcome measure was used, analyses were repeated with arthritis persistency as outcome in the Leiden dataset. A generally accepted definition for persistency is lacking and its frequency depends on the observation period. We defined persistent arthritis as the absence of sustained remission, which was defined as the absence of swollen joints for at least one year after cessation of eventual DMARD therapy. When remission was not obtained after 5 years of disease, a patient was classified as having persistent arthritis. With this definition, 61.3% of UA patients had persistent arthritis.

Severity of disease course in RA patients

Patients who fulfilled the ACR 1987 criteria for RA during the first year and were included in the Leiden EAC between 1993 and 2006 were studied. Of the total of 687 RA patients, 486 had already fulfilled the 1987 ACR criteria for RA at baseline and 201 developed RA within the first year of follow-up (Figure 1).

672 RA patients had radiographs of hands and feet taken at baseline and on consecutive years. These were scored chronologically by an experienced reader (MPMvdL) according to the Sharp/van der Heijde method.[14] Intraobserver intraclass correlation coefficients (ICC) were 0.91 for all radiographs, 0.84 for baseline radiographs, and 0.97 for the radiographic progression rate. To encompass a reliable sample size, radiographic follow-up data were restricted to a maximum of 7 years (median 5, IQR 2-7). Treatment strategies for RA had changed over time and became more aggressive in subsequent inclusion periods (1993-1996, 1996-1998 and 1999-2006), see reference.[15]

A second outcome measure for the severity of the disease course was the achievement of sustained DMARD-free remission. Remission was defined in a stringent form as the persistent absence of synovitis, e.g. no swollen joints, for at least one year after cessation of DMARD therapy and the identification of remission by the patient's rheumatologist.[16] Here, corticosteroids (both oral and intra-articular) were considered as DMARDs; NSAIDs were allowed. Most patients who achieved remission had a follow-up after cessation of DMARDS longer than one year. The remission status could be reliably ascertained in 641 RA patients using medical files. The frequency of DMARD-free remission in these RA patients was 12.3%.

Autoantibody testing

In the Leiden EAC, RF was determined by enzyme-linked immunosorbent assay (ELISA) (IgM-RF, in-house ELISA),[17] using a standard cutoff value of 5 arbitrary units. Anti-CCP2 autoantibodies (total IgG) were measured by ELISA (Immunoscan RA Mark 2; Euro-Diagnostica, Arnhem, The Netherlands). The cutoff level for anti-CCP2 autoantibody positivity was set at 25 arbitrary units, according to the manufacturer's instructions.

In the Berlin cohort, RF was determined by ELISA (Autostat II, Hycor Biomedical, Edinburgh, UK), using a reference value of >24 IU/l Units for a positive test result. Anti-CCP 2 was determined by ELISA (Immunoscan CCPlus, Euro-Diagnostica, Malmö, Sweden), using a reference cutoff of >25 U/l for autoantibody positivity.

In the NOR-VEAC, sera frozen at inclusion were used to analyze anti-CCP2 (Inova Inc., San Diego, USA) and IgM-RF (in-house ELISA) levels in one batch. Cutoffs used to define a positive status were as recommended by the local laboratory: anti-CCP2 25 units/ml and IgM-RF 25 units/ml.

Considering the absence of agreement on a uniform definition of high level RF, two definitions of high RF level were evaluated. These were three times the reference cutoff value, the definition of a high RF level that is used in the 2010 ACR/EULAR criteria, and a RF level of 50 U/ml (RF_{50}), as RF_{50} is the definition of high RF levels used in previous studies on this subject.[5,6]

Variation in RF measurements

In order to facilitate laboratories in quality control in the Netherlands, the SKML - section HIM (Stichting Kwaliteitsbewaking Medische Laboratoria – section Humoral Immunology) organizes external quality assessment schemes for rheumatoid factor testing twice a year. In each scheme six patient samples are sent to 78 participating laboratories. These six patient samples consist of three RF-negative samples, two RF-positive samples and one standard serum (RELARES). This is a commercially available standard serum, consisting of pooled serum of RF-positive patients, which was previously standardised to correspond with 100 International Units using the Rose-Waaler agglutination test.[18,19] For this paper the results of the spring 2008 scheme are used of the two RF-positive patient sera and the standard serum. The sera were tested according to local

protocols and reported in local units and as a ratio compared to the local cutoff value by the participants.

Statistical analysis

Development of RA in UA patients

Different test characteristics (sensitivity, specificity, positive (LR+) and negative (LR-) likelihood ratio) were determined. The likelihood ratio incorporates both the sensitivity and specificity of the test and provides an estimate of how much a test result will change the odds of having a disease. In addition, absolute post test changes on RA after 1 year of follow were determined (positive predictive value (PPV) and negative predictive value (NPV)). Analyses were performed using two descriptions of a high RF level (three times the reference cutoff level and a RF level of 50 U/ml (RF_{50}), and the resulting data were compared with the data for ACPA positivity. RA development was analyzed after 1 year of follow-up and arthritis persistency was classified after 5 years of follow-up.

Severity of disease course in RA patients

Associations with the rate of joint destruction during 7 years of follow-up were assessed using a repeated measurement analysis (RMA) on log-transformed radiological data, because of skewness. The RMA is performed using a multivariate normal regression model that, on longitudinal data, evaluates the progression rates over time and takes into account the correlation between the measurements within one subject. Adjustments were made for age, gender and applied treatment strategy as previously described.[20]

Analysis of sustained DMARD-free remission was performed by comparing Kaplan Meier curves and by Cox regression analysis, correcting for age and gender, taking into account the differences in follow-up times among patients. For patients who achieved remission, the dependent variable was "time-to-event", indicating the time until reaching remission. For non-remission patients the time to last follow-up was used.

Variation in RF measurements

To test for correlations between the different methods that are used for measurement of the RF level, non-parametric Spearman correlation coefficients (ρ) were determined.

SPSS version 17.0 (SPSS Inc., Chicago, IL, USA) was used. P-values <0.05 were considered significant. All reported p-values are two-sided.

RESULTS

Development of RA in UA patients

Baseline characteristics of UA patients included in the three cohorts are presented in Table 1. The percentages of UA patients that developed RA within the first year were 32%, 48% and 12% in the Leiden EAC, Berlin EAC and NOR-VEAC respectively.

First, the predictive values for high RF levels and presence of ACPA antibodies were determined for each cohort separately (Table 2). Increasing the cutoff value for a high RF level yielded an increased PPV and decreased NPV. Similarly, the specificity increased but the sensitivity decreased. For example, in the Leiden EAC data, the PPV increased from 62% (RF positivity) to 69% (three times the reference value) and 72% (RF_{50}) and the NPV decreased from 78% to 75 % and 71% respectively. Also, the specificity increased from 86% (RF positivity) to 93% (three times the reference value) and 97% (RF_{50}) but the sensitivity decreased from 48% to 33% and 14% respectively. In addition, the LR+ increased at the expense of an increased LR-. This indicates that the odds on RA increased in case of a high RF level, but that the odds on RA in case of the absence of a high RF level increased as well. The percentage of UA patients that had a high RF level was 15% (three times the reference value) or 6% (RF_{50}) compared to 25% that was RF positive. The observed effects were comparable for all three cohorts (Table 2).

Second, the results for a high RF level were compared to that of ACPA positivity. In all three cohorts, the 95% confidence intervals (95% CI's) overlapped. Nevertheless the balance between PPV (preferably high) and NPV (preferably high) tended to be better for ACPA than for high level RF. In addition, the balance between LR+ (preferably high) and LR- (preferably low) was better for ACPA presence than for high RF level in all three cohorts. These effects were less compelling in the NOR-VEAC than in the Berlin EAC and Leiden EAC. However, the findings in the NOR-VEAC are more difficult to interpret because of large confidence intervals. These larger

Table 1. Baseline characteristics patients with early undifferentiated arthritis included in the different cohorts

Characteristics	Leiden EAC (n=625)	Berlin EAC (n=154)	NOR-VEAC (n=193)
Age at inclusion, (yrs)	51.0 (16.9)	51.2 (14.5)	46.1 (14.5)
Female, N (%)	368 (58.9)	110 (71.9)	114 (59.1)
Symptom duration at first presentation, days	170 (181)	137.4 (96.1)	35 (30)
Swollen joint count	5.5 (6.0)[§]	2.7 (4.5)[‡]	3.9 (6.8)[£]
CRP (mg/l), median (IQR)	17.0 (7.0-43.0)[$]	6.2 (2.0-16.8)[¥]	14.0 (5.0-32.0)[$]
RF positive, N (%)	154 (24.7)	79 (51.3)	18 (9.3)
ACPA-positive, N (%)	149 (23.9)	44 (28.6)	19 (9.8)

Values are the mean ± SD except where indicated otherwise. CRP: C-reactive protein; SJC: swollen joint count; [§]44 swollen joint count; [‡]28 swollen joint count; [£]66 swollen joint count; [$]Used cutoff for abnormal CRP ≥10 mg/l; [¥]Used cutoff for abnormal CRP >5 mg/l; RF = rheumatoid factor; ACPA = antibodies to cyclic citrullinated peptide

Table 2. Comparison of different high level cutoffs for RF and the reference ACPA for predicting progression from UA to RA in three different cohorts

Patient Cohort	Autoantibody test (cutoff value)	No. of UA patients with a positive test result (%)	PPV	NPV	Likelihood Ratio		Sensitivity	Specificity
			% (95% CI)	% (95% CI)	Pos. (95% CI)	Neg. (95% CI)	% (95% CI)	% (95% CI)
Leiden EAC n=625	RF (5.0)[§]	154 (24.8)	61.7 (54.0-69.4)	77.8 (74.1-81.6)	3.45 (2.60-4.53)	0.61 (0.53-0.70)	47.7 (40.8-54.7)	86.1 (82.8-89.4)
	RF (15.0)[‡]	96 (15.4)	68.8 (59.5-78.0)	74.9 (71.2-78.6)	4.71 (3.17-7.01)	0.72 (0.65-0.79)	33.3 (26.8-39.9)	92.9 (90.5-95.4)
	RF (50.0)[¥]	39 (6.3)	71.8 (57.7-85.9)	70.8 (67.2-74.5)	5.45 (2.77-10.72)	0.88 (0.83-0.94)	14.1 (9.3-19.0)	97.4 (95.9-98.9)
	ACPA[†]	149 (23.9)	67.1 (59.6-74.7)	78.9 (75.3-82.6)	4.33 (3.21-5.83)	0.57 (0.49-0.65)	50.0 (43.1-56.9)	88.4 (85.4-91.5)
Berlin EAC n=154	RF (24.0)[§]	54 (35.3)	68.4 (58.1-78.6)	73.3 (63.3-83.3)	2.34 (1.64-3.33)	0.39 (0.26-0.59)	73.0 (62.9-83.1)	68.8 (58.6-78.9)
	RF (50.0)[¥]	39 (25.3)	72.2 (60.3-84.2)	65.0 (55.7-74.3)	2.81 (1.70-4.66)	0.58 (0.45-0.76)	52.7 (41.3-64.1)	87.3 (72.7-89.8)
	RF (72.0)[‡]	34 (22.1)	79.1 (66.9-91.2)	64.0 (55.0-72.9)	4.08 (2.10-7.93)	0.61 (0.49-0.76)	45.9 (34.6-57.3)	88.8 (81.8-95.7)
	ACPA[†]	41 (26.6)	93.2 (85.7-100.6)	70.0 (61.4-78.6)	14.77 (4.78-45.68)	0.46 (0.36-0.60)	55.4 (44.1-66.7)	96.3 (92.1-100.4)
NOR-VEAC n=193	RF (25.0)[§]	11 (5.7)	61.1 (38.6-83.6)	93.1 (89.4-96.9)	11.61 (5.01-26.95)	0.54 (0.37-0.81)	47.8 (27.4-68.2)	95.9 (92.9-98.9)
	RF (50.0)[¥]	9 (4.7)	75.0 (50.5-99.5)	92.3 (88.4-96.2)	22.17 (6.47-76.01)	0.62 (0.45-0.86)	39.1 (19.2-59.1)	98.2 (96.3-100.2)
	RF (75.0)[‡]	6 (3.1)	85.7 (59.8-111.6)	90.9 (86.7-95.0)	44.35 (5.59-352.06)	0.74 (0.59-0.95)	26.1 (8.1-44.0)	99.4 (98.3-100.6)
	ACPA[†]	14 (7.3)	73.7 (53.9-93.5)	94.8 (91.5-98.1)	20.70 (8.22-52.12)	0.40 (0.24-0.67)	60.9 (40.9-80.8)	97.1 (94.5-99.6)

ACPA: anti-citrullinated-peptide-antibodies; RF: rheumatoid factor; UA: undifferentiated arthritis; RA: rheumatoid arthritis; PPV: positive predictive value; NPV: negative predictive value; 95% CI: 95% confidence interval. [§]Reference cutoff for RF-positivity. [‡]High RF cutoff level: three times reference level. [¥]High RF cutoff level: an absolute level of 50 U/ml. [†]Reference cutoff for ACPA-positivity

confidence intervals may be related to the low percentage of UA patients with a high RF level in this very early cohort (3% for three times the reference value and 5% for RF_{50}). When arthritis persistency was used as outcome measure instead of RA development comparable observations were made (Supplementary Table 1).

Subsequently, the additive value of performing a second autoantibody test was investigated for predicting RA development. In other words, the additive value of performing an ACPA test in UA patients without high level RF was determined, as well as the additive value of testing RF levels in ACPA negative UA patients. As shown in Table 3, the PPVs and NPVs of performing an ACPA test in patients without a high level RF were about twice as large compared to the PPVs and NPVs of RF level testing in ACPA negative patients. This observation was done for different definitions of high level RF and in the different cohorts. The LR+ for additional ACPA testing in patients without a high level RF ranged between 3.6 and 12.4 and the LR- ranged between 0.63 and 0.77 in the Leiden and Berlin EACs. RF level testing in ACPA negative patients resulted in marginal LR+ and LR- (around 1) in these cohorts. This contrast was less evident in the NOR-VEAC but also here the number of ACPA negative UA patients that developed RA that had high levels of RF was very low (n=1). Overall, for the prediction of RA development in early UA patients, performing an ACPA test in addition to a RF level testing seems more valuable than determining the RF level after assessments on the presence of ACPA antibodies.

Severity of disease course in RA patients

The predictive ability for the severity of RA was assessed and compared for high level RF and presence of ACPA. The rate of joint destruction for patients with high RF levels (both for RF_{50} and three times the reference value) and ACPA positive RA patients are depicted in Figure 2A. To compare the effect sizes of the three groups, the estimates obtained from the repeated measurement analyses performed on log-transformed data were back-transformed to the original scale. This yielded a 1.13, 1.05 and 1.04 times greater progression rate per year for the presence of ACPA, three times the reference value of RF and RF_{50} respectively compared its the absence. Over a total followup period of seven years this resulted in 2.41 (95%CI 2.06-2.83, $p<0.001$), 1.45 (95%CI 1.24-1.70, $p<0.001$) and 1.29 (95%CI 1.05-1.59, $p=0.015$) times larger progression rates for ACPA, three times the reference value of RF and RF_{50}.

To further substantiate the findings on RA severity, the analyses were performed with the achievement of sustained DMARD-free remission as outcome (Figure 2B). Presence of ACPA or high RF levels was associated with a worse disease outcome, reflected by an increased hazard ratio (HR) for not achieving DMARD-remission. The observed HRs for not achieving DMARD-free remission were respectively 11.3 (95%CI 5.6-22.7, $p<0.001$), 5.7 (95%CI 2.9-11.4, $p<0.001$) and 3.1 (95%CI 1.2-7.6, $p=0.016$) for ACPA, three times the reference value of RF and RF_{50}. Similar to joint destruction, the effect sizes for high level RF (RF_{50} as well as three times the reference value) were lower than that for the presence of ACPA antibodies.

Table 3. Additional value of testing for high level RF or ACPA in case of a negative test result for the other autoantibody test in predicting development of RA from UA

Patient Cohort	Primary Test result	Additional Test performed	PPV	NPV	Likelihood Ratio		Sensitivity	Specificity	Add. no. of RA-patients with a positive test result (%)
			% (95% CI)	% (95% CI)	Pos.(95% CI)	Neg.(95% CI)	% (95% CI)	% (95% CI)	
Leiden EAC n=625	RF_{15}-‡	ACPA†	54.3 (42.6-66.0)	79.6 (75.9-83.3)	3.57 (2.33-5.47)	0.77 (0.69-0.87)	29.0 (21.2-36.8)	91.9 (89.2-94.6)	38 (6.1)
	RF_{50}-¥	ACPA†	63.5 (54.7-72.3)	79.4 (75.8-83.1)	4.25 (3.04-5.94)	0.63 (0.55-0.72)	43.2 (35.7-50.7)	89.8 (86.9-92.7)	73 (11.7)
	ACPA-†	RF_{15}‡	23.5 (3.4-43.7)	79.6 (75.9-83.3)	1.19 (0.40-3.57)	0.99 (0.95-1.04)	4.1 (0.2-8.1)	96.5 (94.7-98.4)	4 (0.6)
	ACPA-†	RF_{50}¥	20.0 (-15.1-55.1)	79.4 (75.8-83.1)	0.97 (0.11-8.55)	1.00 (0.98-1.02)	1.0 (-1.0-3.0)	98.9 (79.9-100.0)	1(0.2)
Berlin EAC n=154	RF_{50}-¥	ACPA†	84.6 (65.0-104.2)	72.4 (63.0-81.8)	10.21 (2.40-43.52)	0.71 (0.56-0.89)	31.4 (16.0-46.8)	96.6 (92.7-101.1)	11 (7.1)
	RF_{72}-‡	ACPA†	87.5 (71.3-103.7)	72.6 (63.7-81.6)	12.43 (2.97-51.92)	0.67 (0.53-0.84)	35.0 (20.2-49.8)	97.2 (93.3-101.0)	14 (9.1)
	ACPA-†	RF_{50}¥	39.1 (19.2-59.1)	72.4 (63.0-81.8)	1.50 (0.72-3.12)	0.89 (0.70-1.12)	27.3 (12.1-42.5)	81.8 (73.2-90.4)	9 (5.8)
	ACPA-†	RF_{72}‡	46.7 (21.4-71.9)	72.6 (63.7-81.6)	2.04 (0.81-5.17)	0.88 (0.73-1.07)	21.2 (7.3-35.2)	89.6 (82.8-96.4)	7 (4.5)
NOR-VEAC n=193	RF_{50}-¥	ACPA†	54.5 (25.1-84.0)	95.3 (92.1-98.5)	14.31 (4.99-41.07)	0.59 (0.37-0.93)	42.9 (16.9-68..8)	97.0 (94.4-99.6)	6 (3.1)
	RF_{75}-‡	ACPA†	64.3 (39.2-89.4)	95.3 (92.2-98.5)	17.89 (6.76-47.34)	0.48 (0.29-0.80)	52.9 (29.2-76.7)	97.0 (94.5-99.6)	9 (4.7)
	ACPA-†	RF_{50}¥	25.0 (-17.4-67.4)	95.3 (92.1-98.5)	6.11 (0.70-53.07)	0.91 (0.72-1.14)	11.1 (-9.4-31.6)	98.2 (96.1-100.2)	1 (0.5)
	ACPA-†	RF_{75}‡	50.0 (-19.3-119.3)	95.3 (92.2-98.5)	18.33 (1.25-269.92)	0.89 (0.71-1.13)	11.1 (-9.4-31.6)	99.4 (98.2-100.6)	1 (0.5)

ACPA: anti-citrullinated-peptide-antibodies; RF: rheumatoid factor; UA: undifferentiated arthritis; RA: rheumatoid arthritis; PPV: positive predictive value; NPV: negative predictive value; 95% CI: 95% confidence interval; Add. no. of RA patients with a positive test result (%): number of RA patients identified by the second/additional test that was performed, the % is calculated with the total number of patients available in each cohort; ‡High RF cutoff level: three times reference level; ¥High RF cutoff level: an absolute level of 50 U/ml; †Reference cutoff for ACPA-positivity

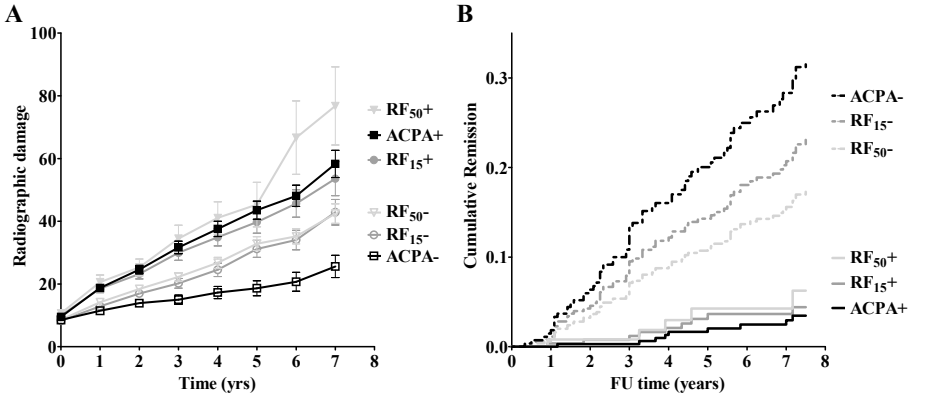

Figure 2. Comparison of high level RF and ACPA for outcome of disease severity in RA patients. Association of outcome with positive (versus negative) test results for two different high level RF cutoffs and ACPA. (A) Sharp/van der Heijde scores for radiographic progression over 7 years of followup (mean (±SEM)). (B) Achievement of DMARD-free remission. In A, the numbers of patients in each group were as follows: for RF_{15}-positive and negative, n=378 and n=271 respectively, for RF_{50}-positive and negative, n=123 and n=526 respectively and for ACPA positive and negative, n=342 and n=289 respectively. In B, the numbers of patients in each group were as follows: for RF_{15}-positive and negative, n=370 and n=252 respectively, for RF_{50}-positive and negative, n=122 and n=500 respectively and for ACPA positive and negative, n=336 and n=270 respectively. RF_{15}: three times the standard cutoff of 5.0; RF_{50}: cutoff of 50.0 U/ml; ACPA: cutoff of 25.0 arbitrary units

Variation in RF measurements

In order to evaluate whether and to what extent the method of measuring the RF level influences the test outcomes, the RF levels determined in the same serum samples by different methods were studied. The serum levels measured are shown in Figure 3A. Large variation in absolute levels was observed. In general the highest levels were measured by nephelometry, followed by turbidimetry and the lowest levels were measured by ELISA. The correlation coefficient between the absolute levels determined by nephelometry and ELISA was 0.470 (p=0.007), between nephelometry and turbidimetry was 0.531 (p=0.002) and between ELISA and turbidimetry was 0.402 (p=0.022). Since the two RF-positive sera used contained high RF levels, all of the measurements done by nephelometry and turbidimetry had an absolute RF level >50 Units. With ELISA, a measurement of <50 Units was found once. Figure 3A illustrates the large variation in measurements that is observed when local units are used.

Expressing the data as a ratio in relation to the local cutoff did not improve the variation within and between methods (Figure 3B). The correlation coefficient between these ratios was 0.288 (p=0.11) for nephelometry and ELISA, 0.443 (p=0.011) for nephelometry and turbidimetry and 0.302 (p=0.093) for ELISA and turbidimetry.

To investigate whether expression of RF level in relation to a standard reference serum would increase the reproducibility of results between laboratories and between methods, the absolute levels of the two patient sera were divided by the RF levels obtained for the standard serum (RELARES). Although the variance within the methods decreased, the variability between meth-

ods was still considerable (Figure 3C). Here, the correlation coefficients were 0.469 (p=0.008) between nephelometry and ELISA, 0.452 (p=0.012) between nephelometry and turbidimetry, and 0.537 (p=0.002) between ELISA and turbidimetry. As is shown, this effort did not lead to harmonization and reflects the difficulty with using standard sera to homogenize RF level measurements.

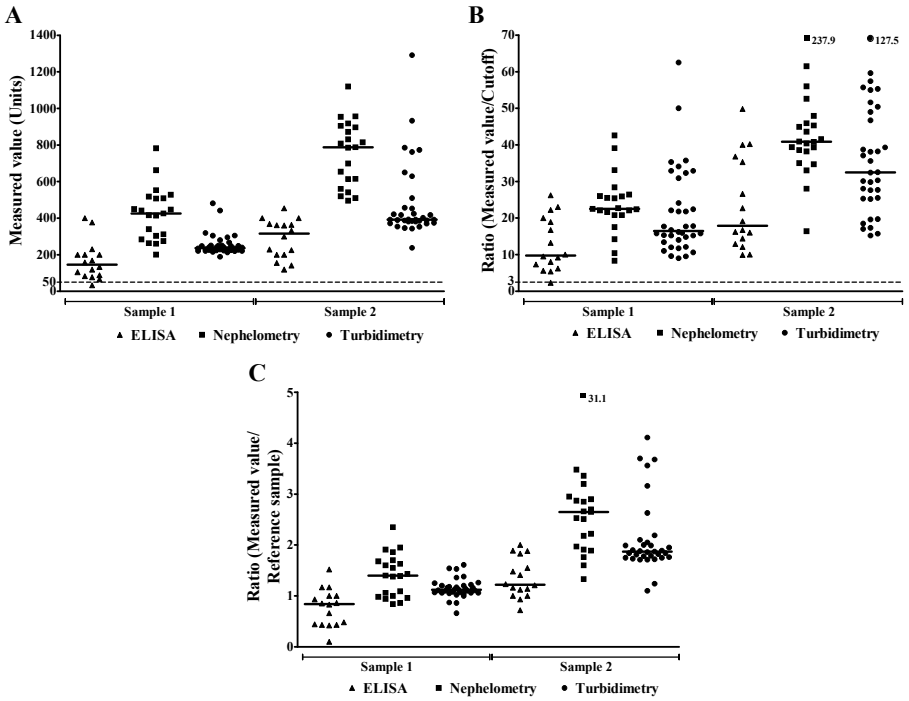

Figure 3. Comparison of RF measurements between different detection methods and different test facilities in samples positive for RF. Each dot represents a single measurement for a sample observed in a separate test facility. Horizontal bars reflect the median. (A) Units were measured in U/ml for ELISA, in kU/l for nephelometry and in IU/l for turbidimetry. The dashed line at 50 units represents the cutoff value of RF_{50}, the definition of a high RF level that is used in literature.[5,6] (B) The number of units determined by each method of measurement divided by the corresponding cutoff value. The dashed line at a ratio of 3 represents three times the reference cutoff value, the definition of a high RF level that is used in the 2010 ACR/EULAR criteria.[4] (C) The number of units determined for each method of measurement divided by the level obtained for the standard serum (RELARES) in the corresponding test facility

DISCUSSION

Detailed knowledge of the individual characteristics of the 2010 ACR/EULAR criteria is necessary to optimally use these criteria in daily clinical practice. The characteristics of the "low-positive RF" versus "high-positive RF" seem to hamper uniform application of the 2010 ACR/EULAR criteria.

The test characteristics and prognostic ability of high RF levels and the presence of ACPA were compared in early UA patients. The data, originating from three cohorts, revealed that the balance between LR+ and LR- as well as between PPV and NPV was more favorable for ACPA positivity than for high level RF. This finding was made with regards to diagnosing RA and having persistent arthritis. The same observations were done when the severity of the course of RA was studied, which substantiated the findings.

The main outcome measure used in the current study was the development of RA by fulfilling the 1987 ACR criteria for RA. An advantage of these criteria is that they could be uniformly applied in the different cohorts in Germany, Norway and the Netherlands. In the light of the new 2010 ACR/EULAR criteria however, this outcome measure may seem an outdated definition of RA. Obviously, the 2010 ACR/EULAR criteria can not be used for the purpose of the present study because of circularity; both the presence of ACPA and RF level are part of these criteria. Usage of MTX treatment as outcome measure, such as done when deriving the 2010 ACR criteria for RA, has limitations as well. UA patients in the Leiden cohort were included since 1993 and at that time DMARDs were infrequently prescribed in early UA. Hence differences in MTX prescription are dependent on the inclusion year, impairing fair comparisons. In addition, when prescribed, MTX is used for other diagnoses as well, for example psoriatic arthritis. An alternative outcome is the expert's opinion on the presence of RA. However, the expert opinion is likely not independent of the 1987 ACR-criteria for RA. Having worked with the 1987-ACR criteria for about twenty years, clinicians may, consciously or unconsciously, refer to these criteria in their judgements. In the present study, comparable observations were done when using RA development, arthritis persistency or RA severity as outcome, suggesting that the findings are not depending on one outcome measure.

Two definitions of high RF levels were studied in three cohorts. The definitions were RF_{50}, the RF level that in previous publications was labeled as high level, and three times the reference value, the definition of high RF included in the 2010 classification criteria for RA. It was observed that the post test probabilities (PPV, NPV) varied between the cohorts. For example the NPV was the highest in the NOR-VEAC and the lowest in the Berlin EAC. These values are influenced by different percentages of UA patients that developed RA during the observation period (the pretest probability). On the other hand, despite this difference, the same tendency in the level data with high RF compared to ACPA positivity was seen, strengthening the findings. The sensitivities and specificities for high RF levels differed between the cohorts as well. This may partly be due to the different cutoff levels used to define RF positivity. Subsequently, RF_{50} may be a twofold increase compared to the cutoff in some cohorts (as was the case in the Berlin EAC and the NOR-VEAC) but it may present a tenfold increase when other methods are applied (as was the case in the Leiden EAC). Although this argument may apply to a lesser extend to the three times the reference value definition for high level RF, also here the stringency with which the reference value was chosen (manufacturer instructions or according to in house reference

groups) affect the test characteristics of this variable. The differences in test characteristics of the presence of ACPA were smaller than for RF level.

Another factor that may contribute to differences in measured RF levels and differences in resulting test characteristics are the different techniques that can be used to measure RF. Here in all cohorts ELISA's were used. Generally for each technique, several variants are prevalent, among which both in house and commercially available kits. The manufacturers of these commercially available tests have not provided a 100% standardization of these kits to a reference kit with regards to detection and quantization of RF. Previously International Units/ml have been established but this method only yields standardized results in case the Boehringer nephelometer is used. The prevalent methods also differ with regard to the origin of the antibodies that are directed against RF (human or rabbit) and the isotypes of the antibodies that are tested. Nephelometry usually measure complexes of IgM-, IgG- and IgA-RF, whereas ELISAs are specifically directed against one isotype, for instance IgM-RF.

Appropriate and uniform application of the RF level criterion of the 2010 criteria for RA requires harmonization of all available RF tests. Efforts to harmonize RF determinations have been done by Dutch and European task forces. In the Netherlands this was done by the development of a standard serum consisting of pooled serum of RF-positive patients (RELARES). However, as reported, this did not result in better reproducibility between laboratories. Considerable variability was still observed, not only between various methods - ELISA, nephelometry and turbidimetry - for determining RF, but also within each method for different laboratories. Considering the present difficulties it is not feasible that worldwide harmonization will be achieved in a short term for measuring RF. This study did not address the possibility to harmonize anti-CCP level measurements. In our experience, harmonizing ACPA measurements may be less complicated (data not shown). Therefore, supposed that a modification of the 2010 ACR/EULAR criteria will arise in time, we propose to omit the RF level and use only ACPA, with different weighed scores for positivity and level.

In conclusion, defining a high level of RF is intricate due to the variation in RF levels obtained when different methods are applied. This problem hampers uniform application of the 2010 ACR/EULAR criteria for RA. The data of the present study revealed that the overall prognostic ability of ACPA positivity outweighs that of high level RF in UA patients. For this reason we suggest that in a future modification of the classification criteria for RA the RF level is not incorporated in contrast to ACPA determination.

REFERENCES

1. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988; 31(3):315-24.
2. Funovits J, Aletaha D, Bykerk V, Combe B, Dougados M, Emery P et al. The 2010 American College of Rheumatology/European League Against Rheumatism classification criteria for rheumatoid arthritis: methodological report phase I. Ann Rheum Dis 2010; 69(9):1589-95.
3. Neogi T, Aletaha D, Silman AJ, Naden RL, Felson DT, Aggarwal R et al. The 2010 American College of Rheumatology/European League Against Rheumatism classification criteria for rheumatoid arthritis: Phase 2 methodological report. Arthritis Rheum 2010; 62(9):2582-91.
4. Aletaha D, Neogi T, Silman A, Funovits J, Felson D, Bingham III CO et al. 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010; 62(9):2569-81.
5. Jansen AL, van der Horst-Bruinsma I, van Schaardenburg D, van de Stadt RJ, de Koning MH, Dijkmans BA. Rheumatoid factor and antibodies to cyclic citrullinated Peptide differentiate rheumatoid arthritis from undifferentiated polyarthritis in patients with early arthritis. J Rheumatol 2002; 29(10):2074-6.
6. Nell VP, Machold KP, Stamm TA, Eberl G, Heinzl H, Uffmann M et al. Autoantibody profiling as early diagnostic and prognostic tool for rheumatoid arthritis. Ann Rheum Dis 2005; 64(12):1731-6.
7. van der Linden MP, van der Woude D, Ioan-Facsinay A, Levarht EW, Stoeken-Rijsbergen G, Huizinga TW et al. Value of anti-modified citrullinated vimentin and third-generation anti-cyclic citrullinated peptide compared with second-generation anti-cyclic citrullinated peptide and rheumatoid factor in predicting disease outcome in undifferentiated arthritis and rheumatoid arthritis. Arthritis Rheum 2009; 60(8):2232-41.
8. Stone R, Coppock JS, Dawes PT, Bacon PA, Scott DL. Clinical value of ELISA assays for IgM and IgG rheumatoid factors. J Clin Pathol 1987; 40(1):107-11.
9. van der Helm-van Mil AH, le Cessie S, van Dongen H, Breedveld FC, Toes RE, Huizinga TW. A prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: how to guide individual treatment decisions. Arthritis Rheum 2007; 56(2):433-40.
10. van Aken J, van Bilsen JH, Allaart CF, Huizinga TW, Breedveld FC. The Leiden Early Arthritis Clinic. Clin Exp Rheumatol 2003; 21(5 Suppl 31):S100-S105.
11. Thabet MM, Huizinga TW, van der Heijde DM, van der Helm-van Mil AH. The prognostic value of baseline erosions in undifferentiated arthritis. Arthritis Res Ther 2009; 11(5):R155.
12. Detert J, Bastian H, Burmester GR. [Update of early arthritis and early rheumatoid arthritis]. Dtsch Med Wochenschr 2005; 130(33):1891-6.
13. Mjaavatten MD, Haugen AJ, Helgetveit K, Nygaard H, Sidenvall G, Uhlig T et al. Pattern of joint involvement and other disease characteristics in 634 patients with arthritis of less than 16 weeks' duration. J Rheumatol 2009; 36(7):1401-6.
14. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 1999; 26(3):743-5.
15. 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 (Oxford) 2010.
16. van der Woude D, Young A, Jayakumar K, Mertens BJ, Toes RE, van der Heijde D et al. Prevalence of and predictive factors for sustained disease-modifying antirheumatic drug-free remission in rheumatoid arthritis: results from two large early arthritis cohorts. Arthritis Rheum 2009; 60(8):2262-71.
17. Otten HG, Daha MR, de Rooij HH, Breedveld FC. Quantitative detection of class-specific rheumatoid factors using mouse monoclonal antibodies and the biotin/streptavidin enhancement system. Br J Rheumatol 1989; 28(4):310-6.
18. Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek. Nederlands Referentiepreparaat voor de bepaling van reumafactoren en de anitperinucleaire factor. www.skml.nl 2010 Available from: URL: http://www.skml.nl/referentiematerialen/rheuma-factor-en-anti-perinucleaire-factor2
19. Klein F, Janssens MB. Standardisation of serological tests for rheumatoid factor measurement. Ann Rheum Dis 1987; 46(9):674-80.

20. van der Linden MP, Feitsma AL, le Cessie S, Kern M, Olsson LM, Raychaudhuri S et al. Association of a single-nucleotide polymorphism in CD40 with the rate of joint destruction in rheumatoid arthritis. Arthritis Rheum 2009; 60(8):2242-7.

Supplementary Table 1. Comparison of different high level cutoffs for RF and the reference ACPA for predicting arthritis persistency in UA patients included in the Leiden EAC

Patient Cohort	Autoantibody test (cutoff value)	No. of UA patients with a positive test result (%)	PPV	NPV	Likelihood Ratio		Sensitivity	Specificity
			% (95% CI)	% (95% CI)	Pos. (95% CI)	Neg. (95% CI)	% (95% CI)	% (95% CI)
Leiden EAC n=569	RF (5.0)[§]	139 (24.4)	82.0 (75.6-88.4)	45.3 (40.6-50.1)	2.87 (1.93-4.28)	0.76 (0.70-0.83)	32.7 (27.7-37.6)	88.6 (84.4-92.8)
	RF (15.0)[‡]	86 (15.1)	86.0 (78.7-93.4)	43.1(38.6-47.5)	3.89 (2.16-6.99)	0.83 (0.78-0.89)	21.2 (16.9-25.5)	94.5 (91.5-97.5)
	RF (50.0)[¥]	35 (6.2)	85.7 (74.1-97.3)	40.3 (36.1-44.4)	3.78 (1.49-9.60)	0.94 (0.90-0.97)	8.6 (5.7-11.5)	97.7 (95.8-99.7)
	ACPA[†]	132 (23.2)	88.6 (83.2-94.1)	46.9 (42.2-51.6)	4.92 (2.95-8.19)	0.71 (0.66-0.78)	33.5 (28.6-38.5)	93.2 (89.9-96.5)

This table is based on 569 UA patients included before March 2005 to provide at least 5 years of followup for establishing persistent disease. 61.3% of these patients had persistent arthritis after 5 years. ACPA: anti-citrullinated-peptide-antibodies; RF: rheumatoid factor; UA: undifferentiated arthritis; RA: rheumatoid arthritis; PPV: positive predictive value; NPV: negative predictive value; 95% CI: 95% confidence interval; [§]Reference cutoff for RF-positivity; [‡]High RF cutoff level: three times reference level; [¥]High RF cutoff level: an absolute level of 50 U/ml; [†]Reference cutoff for ACPA-positivity