



OPEN ACCESS

EXTENDED REPORT

Application of the 2010 ACR/EULAR classification criteria in patients with very early inflammatory arthritis: analysis of sensitivity, specificity and predictive values in the SAVE study cohort

Iuliia Biliavska,^{1,2} Tanja A Stamm,² Jose Martinez-Avila,² Thomas W J Huizinga,³ Robert B M Landewé,⁴ Günter Steiner,² Daniel Aletaha,² Josef S Smolen,² Klaus P Machold²

Handling editor Tore K Kvien

¹Department of non-coronary Heart Disease and Clinical Rheumatology, NSC Institute of Cardiology, Kiev, Ukraine

²Department of Rheumatology, Medical University of Vienna, Vienna, Austria

³Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands

⁴Department of Clinical Immunology & Rheumatology, Academic Medical Center/ University of Amsterdam and Atrium Medical Center Heerlen, Amsterdam, The Netherlands

Correspondence to

Professor Klaus Peter Machold, Department of Rheumatology, Internal Medicine III, Medical University of Vienna, Währinger Gürtel 18-20, Vienna A-1090, Austria; Klaus.machold@meduniwien.ac.at

Accepted 7 August 2012
Published Online First
15 September 2012



Open Access
Scan to access more
free content

ABSTRACT

Objective Performance of the 2010 American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) rheumatoid arthritis (RA) criteria was analysed in an internationally recruited early arthritis cohort (≤ 16 weeks symptom duration) enrolled in the 'Stop-Arthritis-Very-Early' trial. This sample includes patients with a variety of diseases diagnosed during follow-up.

Methods Two endpoints were defined: Investigators' diagnosis and disease-modifying antirheumatic drug (DMARD) treatment start during the 12-month follow-up. The 2010 criteria were applied to score Patients' baseline data. Sensitivity, specificity, predictive values and areas under the receiver operating curves of this scoring with respect to both endpoints were calculated and compared to the 1987 criteria. The optimum level of agreement between the endpoints and the 2010 classification score ways estimated by Cohen's κ coefficients.

Results 303 patients had 12-months follow-up. Positive predictive values of the 2010 criteria were 0.68 and 0.71 for RA-diagnosis and DMARD-start, respectively. Sensitivity for RA-diagnosis was 0.85, for DMARD-start 0.8, whereas the 1987 criteria's sensitivities were 0.65 and 0.55. The areas under the receiver operating curves of the 2010 criteria for RA-diagnosis and DMARD-start were 0.83 and 0.78. Analysis of inter-rater-agreement using Cohen's κ demonstrated the highest κ values (0.5 for RA-diagnosis and 0.43 for DMARD-start) for the score of 6.

Conclusions In this international very early arthritis cohort predictive and discriminative abilities of the 2010 ACR/EULAR classification criteria were satisfactory and substantially superior to the 'old' 1987 classification criteria. This easier classification of RA in early stages will allow targeting truly early disease stages with appropriate therapy.

INTRODUCTION

Early diagnosis and adequate therapeutic intervention with disease-modifying antirheumatic drugs (DMARDs) has become a major goal in the management of rheumatoid arthritis (RA), because it significantly improves clinical outcomes and reduces the level of joint damage and disability. However, similar to other chronic diseases, there

is no «gold» standard for the diagnosis of RA.¹⁻⁴ The 1987 American College of Rheumatology (ACR) RA classification criteria, while very specific in patients with established RA, are limited in their sensitivity to identify early disease stages.⁵ In 2010 the ACR and the European League Against Rheumatism (EULAR) released new classification criteria for RA.^{6,7} These aim to identify and classify, among patients with inflammatory arthritis, those with early RA and in particular those with the highest risk of persistent and/or erosive disease. These individuals most likely benefit from early start of therapy with DMARDs. To develop these criteria, data from nine early arthritis cohorts were analysed to identify the factors (and their relative weights) associated with a clinical decision to start methotrexate (MTX) within the first 12 months.^{8,9} Recently, several studies have assessed the diagnostic accuracy of the 2010 ACR/EULAR classification criteria in comparison to other diagnostic or classification criteria (ie, the 1987 criteria) in several cohorts of patients;¹⁰⁻¹⁵ in addition, the 1987 and 2010 criteria have been compared to each other in one of the cited studies¹⁴ and three other cohorts.¹⁶⁻¹⁸ Overall, the 2010 classification criteria performed well as compared with the 1987 criteria especially regarding the diagnosis of early RA, but some patients were missed, especially when being seronegative and having arthritis of < 10 joints.¹⁴ Moreover, these cohort analyses included (with one exception) patients with longer disease duration and all were done in the framework of a single centre/region

With increasing levels of public information on the importance of early diagnosis of RA and referral recommendations, the most important challenge today relates to patients with very early arthritis. The 'Stop Arthritis Very Early' (SAVE) trial¹⁹ included between 2004 (first patient's first visit: 1 March 2004) and 2007 (last patient's last visit: 31 August 2007) exactly such a population, namely patients with any type of arthritis of ≤ 4 months of symptoms and thus reflected the whole clinical spectrum of individuals with very early joint disease, spanning from those experiencing spontaneous remission to those truly developing RA. Moreover, SAVE derived its large number

of patients not from one centre or centres of a single city or country, but comprised almost 29 centres from many countries throughout Europe, Central America and Central Asia.

The aim of the present study was to assess the performance (sensitivity, specificity, positive and negative predictive values) of the 2010 ACR/EULAR classification criteria in this multi-centric cohort of patients with very early inflammatory arthritis (disease duration ≤ 16 weeks).

METHODS

Study population

The SAVE study has been described in detail elsewhere.¹⁹ Patients with inflammatory arthritis of ≥ 1 joint, symptom duration ≤ 16 weeks and no prior DMARD treatment were enrolled into this double blind placebo controlled trial. Patients with arthritis due to trauma, suspected septic arthritis, gout or only distal interphalangeal joint involvement were excluded. All 389 patients received a single intramuscular injection of 120 mg of methylprednisolone or placebo. Primary outcomes and results of this double blind placebo controlled trial have been published.¹⁹ The study was conducted according to the Declaration of Helsinki. Ethical Committee approval was obtained at every participating centre and informed consent for participation was signed by every participant (Current Controlled Trials Nr.: ISRCTN 86668322).

Baseline and follow-up assessments

Demographic data and symptom duration on the day of inclusion into the SAVE study were recorded. For each patient a 66/68 joint count was performed. In addition, C-reactive protein, erythrocyte sedimentation rate, rheumatoid factor (RF) and anti-citrullinated peptide antibodies (ACPA) were measured. The assessments were carried out at baseline and weeks 2, 12 and 52. During the course of the SAVE study, the investigators caring for the patients were asked at week 12 and 52 to provide their diagnosis if possible. Data on the drug therapy during the 52 week follow-up period were also collected. During follow-up, 80 patients were lost at various timepoints between visits, six additional patients were excluded because of protocol violations. For the present analysis, only data for the patients with 12 months follow-up were used. Details on drop-outs can be found in ref. 19.

Definition of endpoints

For analysis of criteria performance, patients' baseline data were evaluated according to the recently proposed ACR/EULAR scoring system.^{6, 7} For sensitivity, specificity and predictive value analyses, the cutpoint (score ≥ 6) developed in the ACR/EULAR criteria initiative was used. In order to evaluate the diagnostic test performance of the 2010 ACR/EULAR classification criteria, two 'gold standards' were defined: Firstly, diagnosis according to the investigator at any time (both at planned study visits or in between) during the trial (before week 52) was defined as endpoint. Treatment with DMARDs at 52 weeks was selected as an alternative endpoint. The trial protocol did not specify any rules or guidance for DMARD start, so this decision reflects solely the clinical judgement of the caring clinician. MTX use, which is now considered first standard therapy for RA according to the recent recommendations,^{4, 20} was assessed separately, but, at the time of the study, first-line strategies differed among rheumatologists and start of any DMARD therapy (not only MTX) was considered appropriate for initial treatment in RA. First DMARDs used in the present patient cohort were MTX, Sulfasalazine, (Hydroxy-

Chloroquine, Leflunomide, Etanercept and Infliximab; glucocorticoid use, which may also be considered 'disease modifying' was not considered 'DMARD' for the purpose of the present analysis. Likewise, formulations of diagnoses were not standardised and left to the discretion of the treating investigators/clinicians. This was done in order to reflect the clinical situation, in which formal 'classification' criteria frequently are inappropriate and, especially in very early arthritis, may be misleading. In addition, RA patients may refuse DMARD start for personal reasons and DMARD may be deemed necessary for diagnoses other than RA. Therefore, a third analysis was done using 'Diagnosis: RA and/or DMARD start' as endpoint.

Statistical analysis

Statistical tests were performed using R: programming environment for data analysis and graphics (V2.13.1). Quantitative variables are described as median \pm IQR or range (for demographic variables). Qualitative variables are given as number (percentage). To compare the distributions of results between groups of patients, analysis of variance, Kruskal-Wallis rank sum tests and 2-sample tests for equality of proportions with continuity correction were used, where appropriate. Levels of significance were set at $p < 0.05$, with Bonferroni's correction for multiple comparisons when necessary (see text/tables).

For each of the described endpoints sensitivity was plotted against 1-specificity to obtain the receiver operating characteristics (ROC) curve. Positive and negative predictive value (PPV and NPV) at the proposed cut-off values, as well as corresponding areas under the curve (AUC) were calculated. To define agreement between the criteria and the endpoint, Cohen's κ coefficient was calculated, a robust statistical measure which reflects not only a simple percent agreement, but also takes into account the agreement occurring by chance.²¹

RESULTS

For the present analysis, data from the 303 patients (out of 389 included into the SAVE study) who completed the 12 months follow up were used. The distribution of diagnoses (given by the investigators based on their expert judgement) as well as the proportion of DMARD treated individuals in each diagnostic category are shown in figure 1. Demographic data or disease activity at baseline did not differ between patients who did and those who did not complete 12 months follow up (data not shown).

Table 1 depicts the baseline characteristics of all patients as well as their distribution within the diagnostic categories. Joint counts and disease duration differed significantly between patients diagnosed as RA and those with 'undifferentiated arthritis' (UA) or other diagnoses. Likewise, ACPA and RF were significantly more frequent in RA patients than in other individuals. In contrast, acute phase reactant levels were not significantly different among diagnostic groups, although they were somewhat higher numerically in RA patients.

The sensitivity, specificity, PPV and NPV of the 2010 ACR/EULAR classification criteria at the proposed cut-off (≥ 6 points) for the two endpoints are shown in table 2. The 2010 ACR/EULAR classification criteria specifically call for exclusion of patients who have 'other better explanations' for their joint swelling than RA,^{6, 7} such patients were deliberately not excluded in the SAVE trial which looked at arthritis of any kind with the exception of known gout, septic arthritis and osteoarthritis. To account for this requirement of the 2010 classification criteria, an additional analysis was performed in

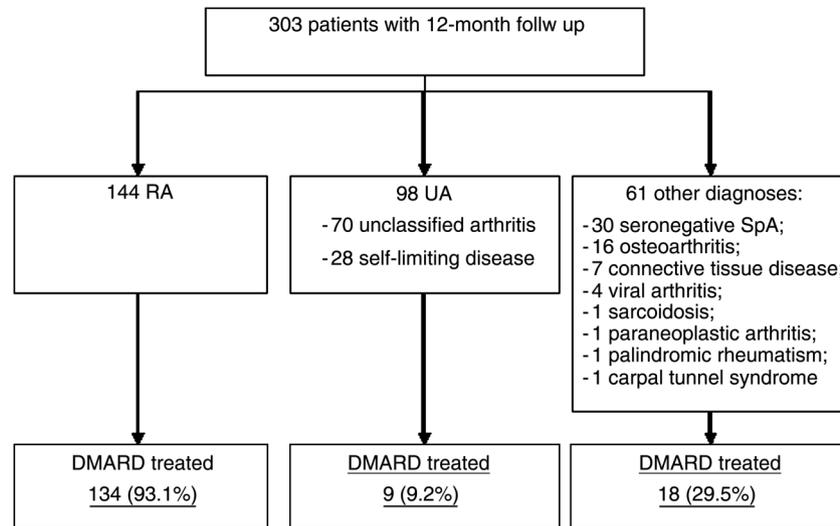


Figure 1 Distributions of patients with inflammatory synovitis grouped according to clinical diagnoses. DMARD, disease-modifying antirheumatic drug; RA, rheumatoid arthritis; SpA, spondyloarthritis; UA, undifferentiated arthritis.

which patients whose treating physicians were able to assign a distinct diagnosis other than RA or UA at or before week 12 (n=35) were excluded, resulting in similar sensitivity and specificity (table 2).

Because the 1987 criteria require the presence of symptoms for ≥ 6 weeks and the SAVE study also included patients with shorter disease duration, sensitivity and specificity excluding

the 110 patients with symptom duration of < 6 weeks were calculated: 0.58 and 0.72 for ‘DMARD treatment’, 0.67 and 0.78 for ‘diagnosis of RA’, and 0.59 and 0.78 for the combined endpoint.

36 (11.9%) patients had missing values for both ACPA and RF due to missing baseline samples (24 in the RA, eight in the UA and four in the ‘other diagnosis’ groups). In one additional ACPA-negative RA patient, RF could not be determined due to technical reasons. According to the 2010 ACR/EULAR application guidelines the value of RF/ACPA for these patients was assumed as ‘negative/normal’^{6 7} in the primary analysis described above. When MTX use (n=120) rather than use of any DMARD (n=161) was accounted for, the respective values were similar, namely 0.84/0.57 (PPV/NPV were 0.53/0.85 respectively). After exclusion of 35 patients with ‘other better explanations’, sensitivity/specificity of the 2010 ACR/EULAR for MTX-use were 0.84/0.54 (PPV/NPV were 0.57/0.82) Setting all missing values to ‘low positive’ (score: 2) would have changed sensitivity/specificity for ‘diagnosis of RA’ to 0.89/0.62 and for ‘DMARD treatment’ to 0.83/0.58. If all missing values had been ‘high positive’ (score: 3) the respective values would be 0.90/0.61 and 0.84/0.67.

Table 3 shows the distribution/frequencies of the components of the 2010 ACR/EULAR criteria within the different diagnostic groups (excluding again the subgroup of 35 patients with ‘other’ diagnoses before week 12). Among RA patients the proportion of individuals with involvement of > 10 joints and with ‘high titre’ ACPA or RF was significantly higher compared to the other diagnostic categories. Likewise, among RA patients the number of individuals with > 6 weeks duration was significantly higher. Among RA patients, 75% had a score of ≥ 6 , whereas 75% of the UA patients had a score < 6 , underscoring the validity of this cut-off in distinguishing RA from undifferentiated (poly- or oligo-) arthritis.

Using diagnoses (RA vs non-RA) and DMARD start as endpoints, two ROC curves were plotted (figure 2). The AUC of the 2010 ACR/EULAR criteria for diagnoses and DMARD start were 0.83 and 0.78, respectively. In comparison, the 1987 ACR classification criteria ROC-curves in this cohort had an AUC of 0.72 for diagnosis (p=0.00451), and 0.65 for DMARD start (p=0.00145), demonstrating substantially improved performance of the 2010 criteria in early arthritis patients.

Table 1 Clinical baseline characteristics of and differences between patients diagnosed as rheumatoid arthritis (RA), ‘undifferentiated arthritis’ (UA) or ‘other’

Characteristics	Patients included in analysis			All (n=303)
	RA (n=144)	UA (n=98)	Other diagnoses (n=61)	
Gender				
Female, n (%)	109 (75.7)	80 (81.6)	39 (63.9)	228 (75.2)
Age, years, mean \pm SD	50.53 \pm 14.86	41.59 \pm 16.2**	50.92 \pm 18.0	47.72 \pm 16.47
Disease duration, days, median (IQR)	70 (38.5;97.0)	43(18.0;78.0)**	49 (24;75)*	57 (30;88)
SJC (66/68), median (IQR)	8 (4;13)	3 (2;8)**	4 (2;8)**	5 (2;11)
TJC(66/68), median (IQR)	15 (8;22)	5 (2;8)**	7.5 (2;16)**	10 (3;19)
CRP, mg/dl, median (IQR)	1.2 (0.32;3.59)	0.75 (0.3;2.65)	0.82 (0.31;2.6)	1.0 (0.31;2.83)
ESR, mm/h, median (IQR)	28.5 (14;44.0)	18 (8;44.0)	30.0 (12;50)	25 (12.0;45.5)
RF positive, n (%)	65 (45.1)	7 (7.1)**	6 (9.8)**	78 (25.7)
ACPA positive, n (%)	61 (42.4)	4 (4.1)**	1 (1.6)**	66 (21.8)
DAS28, mean \pm SD	5.53 \pm 1.27	4.30 \pm 1.29**	4.60 \pm 1.37**	4.95 \pm 0.61

Significance level was set at 0.006 to correct for multiple comparisons.

*p<0.001 compared to patients with RA.

**p<0.0001 compared to patients with RA.

***p<0.006 compared to patients with other diagnoses.

Intergroup differences were analysed by analysis of variance (ANOVA) (multiple groups) and Kruskal-Wallis-test (pairwise comparisons in case of significant results in ANOVA).

ACPA, anti-citrullinated peptide antibodies; CRP, C-reactive protein; DAS28, Disease activity score (28 joints); ESR, erythrocyte sedimentation rate; RF, Rheumatoid factor; SJC, swollen joint count; TJC, tender joint count.

Table 2 Sensitivity, specificity, PPV and NPV (95% CIs in parentheses) of 2010 ACR/EULAR and 1987 ACR classification criteria in patients with very early inflammatory arthritis.

	2010 ACR/EULAR classification criteria			2010 ACR/EULAR classification criteria*			1987 ACR classification criteria		
	DMARD start	Diagnosis: RA	Diagnosis: RA and/or DMARD Start†	DMARD start	Diagnosis: RA	Diagnosis: RA and/or DMARD Start‡	DMARD start	Diagnosis: RA	Diagnosis: RA and/or DMARD Start†
Sensitivity	0.80 (0.73–0.86)	0.85 (0.78–0.90)	0.79 (0.73–0.85)	0.80 (0.74–0.86)	0.85 (0.78–0.90)	0.80 (0.73–0.85)	0.55 (0.47–0.63)	0.65 (0.56–0.73)	0.57 (0.50–0.64)
Specificity	0.63 (0.55–0.71)	0.64 (0.56–0.71)	0.66 (0.58–0.74)	0.61 (0.53–0.70)	0.64 (0.55–0.73)	0.64 (0.56–0.73)	0.76 (0.67–0.83)	0.80 (0.72–0.86)	0.80 (0.72–0.87)
PPV	0.71 (0.65–0.78)	0.68 (0.62–0.77)	0.75 (0.69–0.82)	0.82 (0.76–0.87)	0.74 (0.64–0.78)	0.78 (0.69–0.82)	0.71 (0.62–0.79)	0.73 (0.64–0.80)	0.78 (0.70–0.85)
NPV	0.73 (0.66–0.81)	0.82 (0.76–0.87)	0.70 (0.63–0.76)	0.71 (0.66–0.81)	0.78 (0.72–0.87)	0.68 (0.63–0.76)	0.60 (0.52–0.68)	0.73 (0.65–0.80)	0.59 (0.53–0.66)

*35 patients excluded because of 'other better explanations', that is, diagnosis of specific non-RA-disease within 3 months.

†this category comprises all RA patients (including 10 patients who declined DMARD and/or had contra-indications) and 27 DMARD treated patients with other diagnoses or UA

‡(n = 17 after exclusion of 'other better explanations').

ACR, American College of Rheumatology; DMARD, disease-modifying antirheumatic drug; EULAR, European League Against Rheumatism; NPV, negative predictive value;

PPV, positive predictive value; RA, rheumatoid arthritis.

In order to define the level of agreement between the endpoints (clinical diagnosis and DMARD start) and the 2010 ACR/EULAR classification score Cohen's κ coefficients were calculated. The highest levels of agreement (in the 'moderate' range: 0.5 for diagnosis and 0.43 for DMARD start), were found at the proposed cutpoint of the classification score in both calculations (figure 3). For the 1987 ACR classification criteria, levels of agreement were again lower: for diagnosis, κ was 0.45 and for DMARD start, 0.30.

Finally, characteristics of the misclassified patients (ie, those patients without a 'diagnosis RA' who had a score ≥ 6 or with a 'diagnosis RA' scoring < 6) are presented in table 4. Among 'false positive' (non-RA with a score ≥ 6) patients, 91% had polyarticular disease with either abnormal levels of acute phase reactants (68.2%) or longer symptom duration. The majority of these individuals had chronic UA (n=26, 59%), six patients

(14%) experienced permanent remission ('self-limiting disease'). Diagnoses of the others were osteoarthritis (n=5, 11.4%), seronegative spondyloarthritis (n=3, 6.8%), connective tissue disease (n=3, 6.8%) and viral arthritis (n=1, 2.3%); these diagnoses had not been made at 12 weeks and, therefore, these patients had no 'other better explanation' precluding application of the 2010 ACR/EULAR criteria. 22 'false negative' RA patients (Score < 6) were mostly ACPA/RF seronegative (81.8%) and had less than 10 involved joints (86.4%).

DISCUSSION

The SAVE study exclusively incorporated a population of patients in the earliest symptom stages (≤ 16 weeks), who were recruited into an international multi-centre study and followed prospectively over 12 months. This group of patients mirrors closely very early arthritis patients in a 'real life' setting and across many countries on three continents. The 2010 ACR/EULAR classification criteria for RA were developed to facilitate early recognition of RA, to guide therapeutic intervention and also to form homogeneous early RA patient groups for clinical trials. Furthermore, the criteria aimed to achieve increased sensitivity and specificity compared to the 1987 ACR classification criteria among patients with early disease.^{6 7 22} Thus, the present analysis is the first study validating the new RA criteria on the background of international practices in patients with very early arthritis. Importantly, data from these more than 300 patients were not included in the derivation and formulation of the 2010 ACR/EULAR criteria.

The endpoints used in this study (diagnosis of RA and DMARD start) were chosen because they reflect the two most important clinical decisions when counselling patients with early arthritis: (i) a diagnosis of RA in a given patient usually implies a worse prognosis than most other diagnoses; and (ii) starting DMARD treatment is associated with some risks concerning potential toxicities or other limitations for the patients, for example, regarding pregnancy or alcohol consumption.

Application of the ACR/EULAR 2010 criteria in this international very early arthritis cohort demonstrated substantially increased sensitivity and somewhat lesser specificity compared the 'old' 1987 classification criteria. Likewise, positive predictive values were virtually identical between the 2010 and the 1987 criteria, whereas negative predictive values appeared substantially improved with the 2010 criteria. This easier classification of RA in its early stages will allow to target truly early disease stages with appropriate therapy.

Of note, the SAVE trial was performed in the same time frame as the early arthritis cohorts used for the derivation of

Table 3 N of patients in each category (percentage in parentheses) of the 2010 classification criteria components in patients with RA and other diagnoses

Symptoms	RA (n=144)	UA (n=98)	Other diagnosis (after week 12, n=26)
Joint involvement			
1 large joint (%)	0 (0)	4 (4.0)	1 (3.9)
2–10 large joints (%)	2 (1.3)	8 (8.1)	5 (19.3)**
1–3 small joint (%)	15 (10.5)	30 (30.3)*	5 (19.3)***
4–10 small joints (%)	23 (16.1)	23 (23.2)	3 (11.5)
>10 joints (%)	104 (72.2)	33 (33.7)**	12 (63.2)*
Serology			
Normal RF or ACPA (%)	72 (50.0)	89 (90.8)**	21 (80.8)*
Low positive RF or ACPA (%)	12 (8.4)	4 (4.0)	2 (7.7)
High positive RF or ACPA	60 (41.96)	5 (5.1)**	3 (11.5)*
Acute Phase reactants			
Abnormal CRP or ESR (%)	106 (73.6)	55 (56.1)*	19 (73.1)
Symptom duration			
<6 weeks	35 (24.5)	48 (48.5)*	7 (26.9)
Score, mean (IQR)	7 (6;9)	4 (3;6)**	6 (3.25;7)**

Intergroup difference was estimated with proportion test: 2-sample test for equality of proportions with continuity correction

*p<0.05 compared to patients with RA.

**p<0.001 compared to patients with RA.

***—p<0.001 compared to patients with UA.

ACPA, anti-citrullinated peptide antibodies; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; RA, rheumatoid arthritis; RF, rheumatoid factor; UA, undifferentiated arthritis.

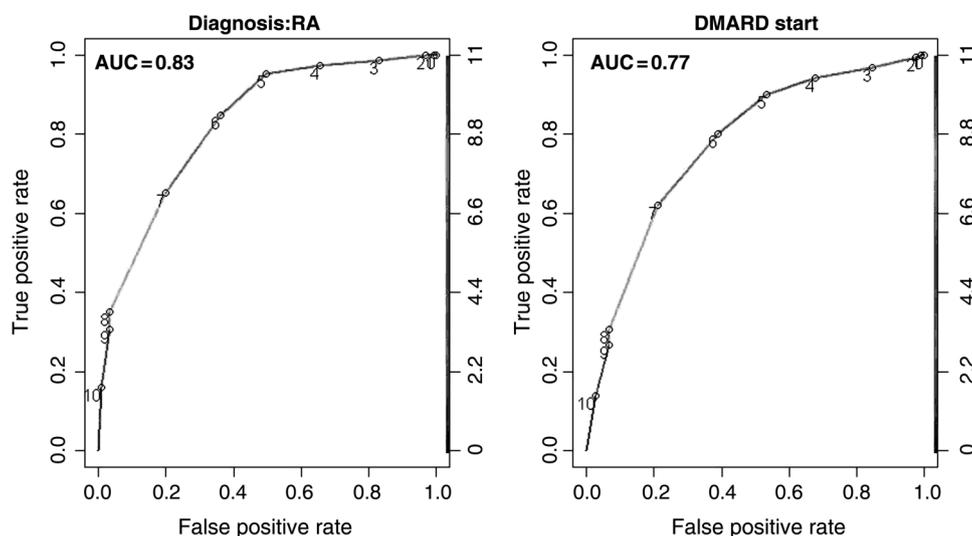


Figure 2 ROC-curves for the two endpoints (“Diagnosis of RA” and disease-modifying antirheumatic drug-treatment) for the ACR/EULAR 2010 scores (numbers in the curve). AUC, area under the curve; ROC, receiver operating characteristics.

the 2010 criteria, namely from 2000 onwards. Nevertheless, we decided not to use MTX treatment as the gold standard as done in the 2010 classification criteria, but the broader category ‘DMARD treatment’ to encompass the totality of potential therapies for RA and increase the sample size given the relative small number of patients studied compared with the patient numbers available for the derivation of the new classification criteria. Importantly, however, when we performed a sensitivity analysis using MTX as gold standard, it yielded virtually identical results.

It is further noteworthy that remarkably similar results were obtained in the present investigation compared with published

analyses: sensitivity and specificity of the 2010 criteria were around 80 and 60 percent, respectively, for both ‘diagnosis of RA’ and ‘DMARD-treatment’ in all published studies.

In patients with very early arthritis, when compared with those with sometimes substantially longer duration of disease as evaluated in most other validation studies, the risk of failing the classification criteria might be higher because of a higher propensity for spontaneous remission;²³ indeed, the spontaneous remission rate in the SAVE trial was in the order of 18%.¹⁹ Nevertheless, the data obtained are in line with those of the mentioned publications (refs. 10–18), revealing a sensitivity of at least 85% and a specificity of at least 64% for the clinical diagnosis of RA according to the investigators, depending on the type of analysis performed. Similar data were obtained when DMARD or MTX start was used as an anchor.

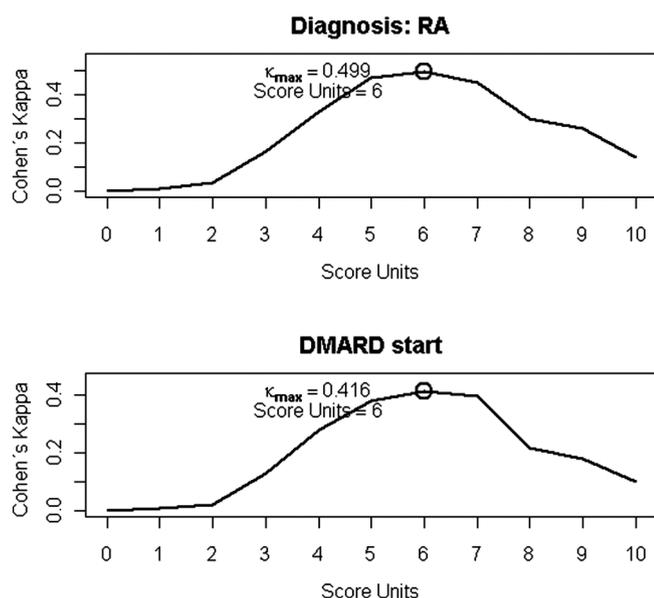


Figure 3 Analysis of agreement between the 2010 ACR/EULAR classification criteria and the two endpoints (diagnosis of RA according to expert opinion and start of disease-modifying therapy): Cohen’s κ coefficients were highest at the proposed cutpoint. The strength of agreement was moderate (κ coefficient 0.80–1.00=very good; 0.60–0.80=good; 0.40–0.60=moderate; 0.20–0.40=fair; less than 0.20=poor).

Table 4 Distribution of the components of the 2010 ACR/EULAR score set among misclassified patients (percentage in parentheses)

Category	Misclassified	
	False positive n=44	False negative n=22
Joint involvement		
1 large joint (%)	0 (0)	0 (0)
2–10 large joints (%)	0 (0)	2 (9.1)
1–3 small joints (%)	1 (2.3)	7 (31.8)
4–10 small joints (%)	3 (6.8)	10 (45.5)
>10 joints (%)	40 (90.9)	3 (13.6)
Serology		
Normal RF or ACPA (%)	35 (79.6)	18 (81.8)
Low positive RF or ACPA (%)	1 (2.3)	1 (4.6)
High positive RF or ACPA	8 (18.2)	3 (13.6)
Acute-phase reactants		
Normal CRP or ESR (%)	14 (31.8)	9 (40.9)
Abnormal CRP or ESR (%)	30 (68.2)	13 (59.1)
Symptoms duration		
<6 weeks	14 (31.8)	10 (45.5)
≥6 weeks	30 (68.2)	12 (54.6)
Score, median (IQR)	7 (6;7)	5 (5;7.5)

ACPA, anti-citrullinated peptide antibodies; CRP, C-reactive protein; ESR-erythrocyte sedimentation rate; RF, rheumatoid factor.

Interestingly, the highest agreement with the diagnosis of RA by the investigators was found at the level of 6 of the 10 points of the classification criteria, validating the cutpoint developed by the ACR-EULAR task force in an independent patient population.

In light of these data, a considerable number of patients were 'over-classified'; however, almost 70% of these patients had indeed a symptom duration of more than 6 weeks (most labelled as UA), which in itself is considered to be indicative of 'chronicity'. Thus, arguably, these individuals were in need of 'disease modifying' treatment despite not meriting, in the opinion of their treating physicians, a diagnosis of 'rheumatoid arthritis'. DMARD treatment in these patients would thus generally not be considered overtreatment. In fact, the 2010 criteria were also developed having the risk of persistence in mind.^{6,7} In any case, having a few patients 'overtreated' with DMARDs may be less problematic than, as with the 1987 criteria, missing to classify (and treat) patients timely and thus risking progression of joint damage.²⁴ The single most frequent reason for 'over-classification', however, was polyarticular disease affecting >10 joints. These data can be interpreted in a way that non-RA with polyarticular onset has the greatest risk of misclassification (only one additional point needed over the 5), while for the—in 80% seronegative—RA patients missing this category is a risk factor for 'under-classification'.

On the other hand, individuals who have seronegative oligoarticular arthritis may be missed for classification as RA due to the relative weight of the autoantibodies among the criteria. In addition, the presence or absence of RF or ACPA induces some circularity for the clinician: antibody positive patients are more likely to receive a diagnosis of RA than the rest. However, since the disease course in seronegative patients is usually more benign, delaying or even withholding DMARD treatment in this group of patients may not be too problematic. Moreover, the 2010 criteria have been developed for classification, while diagnosis of RA can still be made in an individual patient even when failing the classification criteria and vice versa. While a number of patients would have been classified by the 2010 criteria falsely, their score was only one point off in most of the cases: median score of false positives being 7, and of false negatives being 5.

The study also highlights a common problem of diagnostic/classification criteria: In the classification setting a higher sensitivity will allow more patients with true RA to go into early studies, accepting the fact that some will be included falsely. For clinical practice, the clinical diagnosis may be informed by the classification result, and a more sensitive tool may be better suited to screen the suspected RA patients. The primary use for classification criteria, however, is for clinical studies.

One hallmark of the 2010 ACR/EULAR criteria is the prerequisite of excluding patients with 'other better explanations' for their arthritis. In the present analysis, applying the 'other explanation' criterion excluded 57% of the patients with 'other diseases' during the first 3 months and yielded somewhat higher predictive values without substantially affecting sensitivity or specificity. In our analysis, especially patients with longer disease duration (>6 weeks) and polyarticular involvement at the initial visit were prone to develop such 'other diagnoses' over the course of 1 year. Because the 'other better explanation' was evaluated after 12 weeks in the present analysis, this bears an important implication for clinical practice: some of the 'very early arthritis patients' clearly classifiable as RA are 'at risk' to be diagnosed as another disease shortly after presentation and thus the classification algorithm should only

be applied once all results of investigations allowing differential diagnosis are available. Alternatively, the initial 'classification' may need to be re-evaluated after a short period of time.

The major limitation of this study relates to the fact that only few follow-up examinations were made (three study visits over the period of 1 year) and that clinical algorithms for diagnoses and/or DMARD start may vary widely in different institutions, the latter introducing an element of inconsistency. However, the strength of the study lies in its multicentre and international nature such that the results can be seen as representative of very early arthritis cohorts anywhere. Moreover, none of the patients had symptoms of longer than 16 weeks and the whole spectrum of early arthritis patients was included.

Another limitation with regard to the chosen endpoints is the fact that investigators were aware of the laboratory results such as erythrocyte sedimentation rate and C-reactive protein and potentially also of ACPA and RF tests (if done locally as well). The results of these tests may have influenced their clinical diagnosis to a varying degree, according to clinical knowledge and experience of the physicians at that time, reflecting the contemporary clinical real-life situation. For the present analysis, testing for ACPA and RF was done centrally on the initial blood samples and these results were used for this validation.

Finally, clinical trials often tend to have a selection of patients. This could be caused by various mechanisms, but both the patients and the rheumatologists often play an important role in causing such selection bias. However, the multicentre and multinational nature of this study may have helped to reduce such bias.

In summary, in this real-life multicentre and multinational study, the 2010 ACR/EULAR classification criteria for RA showed a substantially higher sensitivity than the 1987 criteria, while the specificity was somewhat lower. Thus, it is easier to classify RA in its early stages now than using the old criteria, allowing to truly target early disease stages with appropriate therapy, both 'standard' or experimental.

Acknowledgements We thank Love Amoyo and Carl Walter Steiner for handling, storing and retrieving the patients' samples.

Contributors IB: data extraction, data analysis, manuscript writing; TAS: database management, data extraction, data analysis; JM-A: statistics, data structuring; TWJH, RBML: manuscript writing; GS: serology analyses, data analysis; DA, JSS: data analysis, manuscript writing; KPM: supervision of data extraction, analysis and statistics, manuscript writing.

Funding This research was conducted while Iulia Biliavska was an ARTICULUM Fellow. Additional support by funding from the European Community's Seventh Framework Programme FP7 under grant agreement number HEALTH-F2-2008-223404 ('Masterswitch') and by the Innovative Medicines Initiative Joint Undertaking under grant agreement number 115142 ('BTCure'), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme and EFPIA companies' in kind contribution.

Competing interests None.

Ethics approval Ethics Committee of the Medical University of Vienna.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

1. **Combe B**, Landewe R, Lukas C, *et al*. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCSIT). *Ann Rheum Dis* 2007;**66**:34–45.

2. **Emery P**, Breedveld FC, Dougados M, *et al*. Early referral recommendation for newly diagnosed rheumatoid arthritis: evidence based development of a clinical guide. *Ann Rheum Dis* 2002;**61**:290–7.
3. **Huizinga TW**, Machold KP, Breedveld FC, *et al*. Criteria for early rheumatoid arthritis: from Bayes' law revisited to new thoughts on pathogenesis. *Arthritis Rheum* 2002;**46**:1155–9.
4. **Smolen JS**, Landewe R, Breedveld FC, *et al*. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. *Ann Rheum Dis* 2010;**69**:964–75.
5. **Banal F**, Dougados M, Combescure C, *et al*. Sensitivity and specificity of the American College of Rheumatology 1987 criteria for the diagnosis of rheumatoid arthritis according to disease duration: a systematic literature review and meta-analysis. *Ann Rheum Dis* 2009;**68**:1184–91.
6. **Aletaha D**, Neogi T, Silman AJ, *et al*. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Arthritis Rheum* 2010;**62**:2569–81.
7. **Aletaha D**, Neogi T, Silman AJ, *et al*. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Ann Rheum Dis* 2010;**69**:1580–8.
8. **Funovits J**, Aletaha D, Bykerk V, *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**:1589–95.
9. **Neogi T**, Aletaha D, Silman AJ, *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**:2582–91.
10. **van der Linden MP**, Knevel R, Huizinga TW, *et al*. Classification of rheumatoid arthritis: comparison of the 1987 American College of Rheumatology criteria and the 2010 American College of Rheumatology/European League Against Rheumatism criteria. *Arthritis Rheum* 2011;**63**:37–42.
11. **Britsemmer K**, Ursum J, Gerritsen M, *et al*. Validation of the 2010 ACR/EULAR classification criteria for rheumatoid arthritis: slight improvement over the 1987 ACR criteria. *Ann Rheum Dis* 2011;**70**:1468–70.
12. **Alves C**, Luime JJ, van ZD, *et al*. Diagnostic performance of the ACR/EULAR 2010 criteria for rheumatoid arthritis and two diagnostic algorithms in an early arthritis clinic (REACH). *Ann Rheum Dis* 2011;**70**:1645–7.
13. **Cader MZ**, Filer A, Hazlehurst J, *et al*. Performance of the 2010 ACR/EULAR criteria for rheumatoid arthritis: comparison with 1987 ACR criteria in a very early synovitis cohort. *Ann Rheum Dis* 2011;**70**:949–55.
14. **Kaneko Y**, Kuvvana M, Kameda H, *et al*. Sensitivity and specificity of 2010 rheumatoid arthritis classification criteria. *Rheumatology (Oxford)* 2011;**50**:1268–74.
15. **Varache S**, Cornec D, Morvan J, *et al*. Diagnostic accuracy of ACR/EULAR 2010 criteria for rheumatoid arthritis in a 2-year cohort. *J Rheumatol* 2011;**38**:1250–7.
16. **Krabben A**, Huizinga TW, van der Helm-van Mil AH. Undifferentiated arthritis characteristics and outcomes when applying the 2010 and 1987 criteria for rheumatoid arthritis. *Ann Rheum Dis* 2012;**71**:238–41.
17. **de Hair MJ**, Lehmann KA, van de Sande MG, *et al*. The clinical picture of rheumatoid arthritis according to the 2010 ACR/EULAR criteria: Is this still the same disease? *Arthritis Rheum* 2012;**64**:389–93.
18. **Fautrel B**, Combe B, Rinciveal N, *et al*. Level of agreement of the 1987 ACR and 2010 ACR/EULAR rheumatoid arthritis classification criteria: an analysis based on ESPOIR cohort data. *Ann Rheum Dis* 2012;**71**:386–9.
19. **Machold KP**, Landewe R, Smolen JS, *et al*. The Stop Arthritis Very Early (SAVE) trial, an international multicentre, randomised, double-blind, placebo-controlled trial on glucocorticoids in very early arthritis. *Ann Rheum Dis* 2010;**69**:495–502.
20. **Saag KG**, Teng GG, Patkar NM, *et al*. American College of Rheumatology 2008 recommendations for the use of nonbiologic and biologic disease-modifying antirheumatic drugs in rheumatoid arthritis. *Arthritis Rheum* 2008;**59**:762–84.
21. **Armitage P**, Berry G, Matthews J. Kappa measure of agreement. In: *Statistical methods in Medical Research*. 4th edition. Malden, Oxford, Carlton, Berlin: Blackwell Science Ltd, 2002:698–704.
22. **Arnett FC**, Edworthy SM, Bloch DA, *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;**31**:315–24.
23. **Quinn MA**, Green MJ, Marzo-Ortega H, *et al*. Prognostic factors in a large cohort of patients with early undifferentiated inflammatory arthritis after application of a structured management protocol. *Arthritis Rheum* 2003;**48**:3039–45.
24. **Nell VP**, Machold KP, Eberl G, *et al*. Benefit of very early referral and very early therapy with disease-modifying anti-rheumatic drugs in patients with early rheumatoid arthritis. *Rheumatology (Oxford)* 2004;**43**:906–14.