RHEUMATOLOGY

Original article

Sensitivity and specificity of 2010 rheumatoid arthritis classification criteria

Yuko Kaneko[1], Masataka Kuwana[1], Hideto Kameda[1] and Tsutomu Takeuchi[1]

Abstract

Objective. To validate the sensitivity and specificity of the 2010 RA classification criteria.

Methods. A total of 313 undiagnosed subjects, who first visited Keio University Hospital with joint symptoms, including arthralgia, joint swelling and morning stiffness, without any previous treatment except for NSAIDs, were included in the present study. A clinical diagnosis of RA was made by rheumatologists, and the gold standard diagnosis of RA was defined as an indication for instituting DMARDs for RA.

Results. Seventy-six subjects were diagnosed as gold standard RA. Among these, 8 did not have any swollen joints, 50 were classified as definite RA under the 2010 criteria and the other 18 as not having RA. Eighty-two subjects were eligible for the 2010 criteria, and the sensitivity and specificity under the 2010 criteria were 73.5 and 71.4%, respectively, compared with 47.1 and 92.9% under the 1987 criteria. But the sensitivity of the 2010 criteria decreased to 15.8% when both RF and anti-CCP were negative. According to the result of a receiver-operated characteristic (ROC) curve of the scoring system, if swollen joints and differential diagnosis are not accurately detected, it would be better to use a score of 5 as the cut-off level to detect RA.

Conclusion. The 2010 classification criteria have a high sensitivity and have been verified to be useful for distinguishing RA at an early stage.

Key words: Rheumatoid arthritis, Classification criteria, Scoring system, Sensitivity, Specificity.

Introduction

RA is a chronic inflammatory disease characterized by progressive destructive arthritis with pain and disability [1]. Recent progress in its treatment, such as MTX and biological DMARDs, has given remarkable benefits to RA patients [2–6]. To manage RA patients appropriately, a diagnosis and a treatment strategy are needed as early as possible [7]. However, at present, an RA diagnosis is usually made under the 1987 ACR classification criteria [8], which are considered to be unsuitable for an early diagnosis [9–10]. Since 2007, the European League against Rheumatism (EULAR) and the ACR have been cooperatively dealing with a revision of the classification criteria, which was finally published in August 2010 [11–13]. The new criteria consist of a classification scoring system, which noticeably puts a great deal of emphasis on small joint involvement and seropositivity of RF or ACPAs. In detail, classification as definite RA is based on the presence of synovitis in at least one joint, the absence of an alternative diagnosis better explaining the synovitis and a total score from individual scores in four domains (the number and site of involved joints, serological abnormality, elevated acute-phase response and symptom duration).

It has been described that the focus of the new classification criteria was not on developing diagnostic criteria or reference tools for primary care physicians, but on facilitating the study of persons with earlier stages of RA. However, since hereafter we are mainly going to use the 2010 classification criteria as an aid in the diagnosis of RA in the clinical field, we should be well acquainted with their strengths and limitations. The aim of this study is to validate the sensitivity and specificity of the 2010 criteria, and to find certain

[1]Department of Internal Medicine, Division of Rheumatology, Keio University School of Medicine, Tokyo, Japan.

Correspondence to: Yuko Kaneko, Department of Internal Medicine, Division of Rheumatology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo, Japan. E-mail: ykaneko@z6.keio.jp

CLINICAL SCIENCE

characteristics of patients with RA who are not classified as RA and vice versa.

Subjects and methods

This study was a retrospective single centre observational study. In order to optimize the quality and reproducibility of this validation study, the work was designed to comply with the criteria of the Standards for Reporting of Diagnostic Accuracy initiative [14].

Subjects

The subjects, all of whom first visited the out-patient clinic of the Division of Rheumatology, Department of Internal Medicine at the Keio University School of Medicine in January 2009 through March 2010, were reviewed retrospectively. Three hundred and fourteen undiagnosed subjects exhibited joint symptoms, including arthralgia, joint swelling and morning stiffness, without any previous treatment with the exception of NSAIDs. Among these, one patient was excluded from the study because of insufficient laboratory data to comply with the new criteria. Ultimately, 313 subjects were included in the present study. Medical ethics committee approval was waived because the study was a retrospective cohort study using anonymized information.

Diagnoses of RA and other diseases

Diagnoses of RA were made by at least one of six rheumatologists in our institution from a comprehensive standpoint, using clinical histories including when and how symptoms started, physical findings including the site and extent of involved joints and extra-articular lesions, blood tests including RF, ACPA, acute-phase reactants and MMP and X-rays. MRI of symptomatic joints was also used when diagnosis was not able to be settled, and synovitis with bone erosion or osteitis was considered as the presence of RA. Because the absolute gold standard diagnosis of RA does not exist, in the present study, the gold standard for a diagnosis of RA was defined as an indication for instituting DMARDs for RA, including salazosulphapyridine, bucilamine, tacrolimus, MTX, infliximab, etanercept, adalimumab and tocilizumab. The six above-mentioned rheumatologists are all specialists in rheumatology, each with >10 years of clinical experience. Diagnoses of other diseases were also made through a similar process. Subjects regarded as not being affected by particular diseases were termed no appreciable disease (NAD). Subjects observed having modest arthritis but where diagnosis of a particular disease was not sure enough for treatment despite repeated examinations, were termed undifferentiated peripheral inflammatory arthritis (UPIA).

Assessment of clinical manifestations and laboratory findings

Demographics and clinical manifestations, including sex, age, duration of symptoms, the number of tender joints and the number of swollen joints, were evaluated. Blood samples were examined in our hospital laboratory. The upper limits of CRP, measured by dry chemistry (Mitsubishi Chemical Medicine, Tokyo, Japan), the ESR, measured by the Westergren test, IgM-RF, measured by a latex-enhanced immunonephelometric assay (Eiken Chemical, Tochigi, Japan) and anti-CCP, measured by an ELISA (Medical & Biological Laboratories, Nagano, Japan) were 0.35 mg/dl, 10 mm/h for men and 15 mm/h for women, 20 IU/l and 4.5 U/ml, respectively.

Statistical analysis

Subject characteristics were summarized using medians and ranges, and the values of CRP and ESR, as well as the number of involved joints, were summarized using mean (S.D.). Comparisons of frequency between the two groups were performed using the Pearson chi-squared test. Comparisons of mean value were performed by Student's *t*-test. Sensitivity *vs* the false positive frequency (one-specificity) for the scoring system was analysed by a receiver-operated characteristic (ROC) curve. All reported *P*-values are two-sided. $P < 0.05$ was considered to be statistically significant. Data were analysed with SPSS version 18.0 (SPSS, Inc., Chicago, IL, USA).

Results

### Characteristics and diagnoses of 313 subjects

The subject characteristics were as follows: age, median (range) years, 54 (14–86); sex, *n* (%), female, 79; duration of symptoms, median (range) weeks, 18 (1–1040); interval between the first visit and the time of diagnosis, median (range) weeks, 2 (1–40). Diagnoses of subjects at the last visit were 76 with RA, 4 UPIA, 68 NAD and 165 other diseases. All subjects were observed until they were diagnosed or for >3 months if they could not be accurately diagnosed (i.e. UPIA).

At the point when the first laboratory and radiographic findings were available, mostly within 3 weeks from the first visit, the patients were assessed and subjected to the 2010 classification criteria. A flow diagram is shown in Fig. 1. Of 313 subjects, 124 had at least one swollen joint and, among these, 82 were eligible to be subjected to the classification scoring system. Fifty-four subjects achieved a total score of ⩾6, and their clinical diagnoses were 50 RA, 1 UPIA and 3 NAD. Twenty-eight subjects showed a score of <6, and their diagnoses were 18 RA, 2 UPIA and 8 NAD.

Among 76 RA patients, RF and anti-CCP were positive in 50 (66%) and 46 (61%) patients, respectively. Regarding the length of time between the first visit to our hospital and the time of diagnosis of RA, 71 (93%) subjects were diagnosed within 12 weeks, 3 (4%) within 24 weeks and 2 (3%) after >24 weeks.

Diagnoses of another 165 subjects included OA (*n* = 74), post-menopausal syndrome (PMS; *n* = 14), tendonitis (*n* = 13), SS (*n* = 12), SLE (*n* = 6), PM/DM (*n* = 4), PsA (*n* = 4), viral infection (*n* = 4), PMR (*n* = 4), palindromic rheumatism (*n* = 3), adult onset Still's disease (*n* = 3), post-injury (n = 3), AS (*n* = 2), shoulder periarthritis (*n* = 2),

FIG. 1 Flow chart of result. Of 314 subjects who visited our institute with joint symptoms without any treatment, 313 subjects were included in this study. One hundred and twenty-four subjects had at least one swollen joint, and among these, 42 were diagnosed with other diseases and 82 were submitted to the scoring system, resulting in 54 subjects with a total score of ⩾6. The clinical diagnoses of subjects at the last visit are shown in the dotted square.
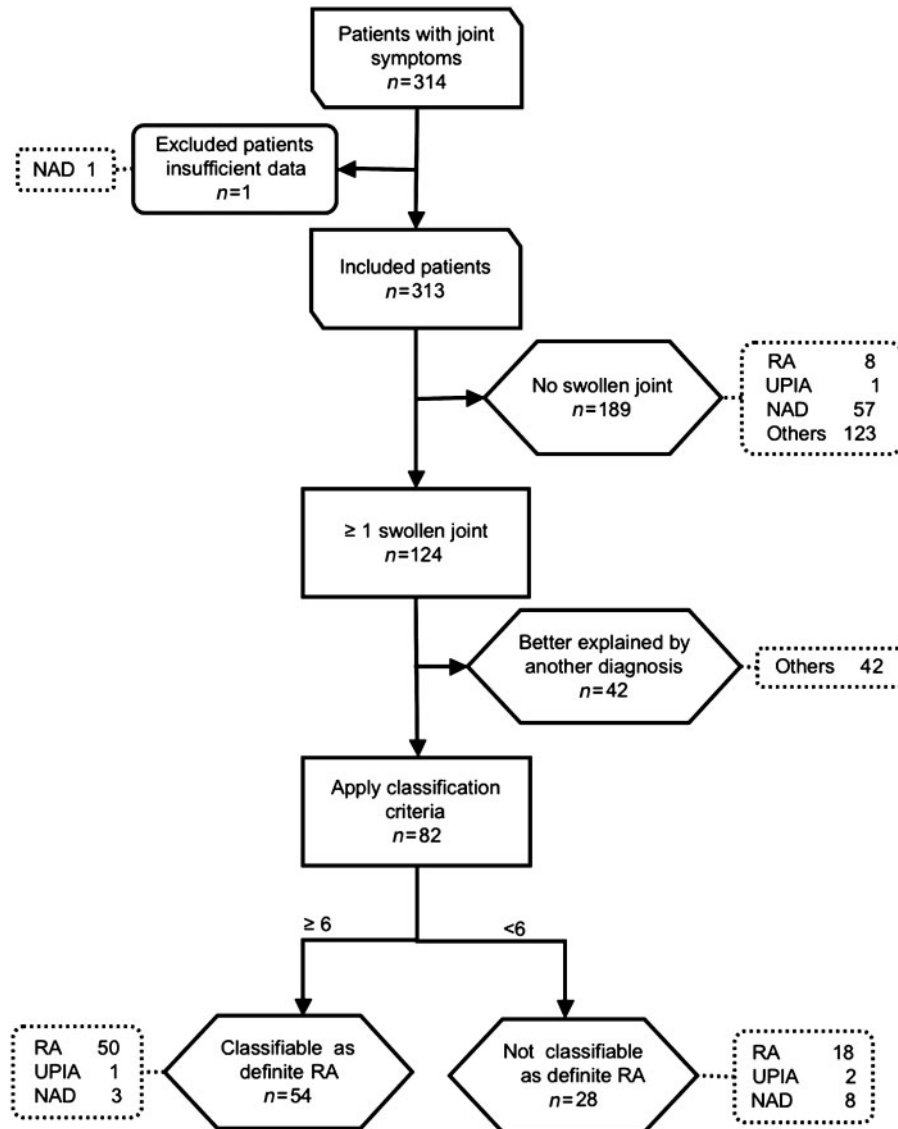
pseudogout ($n=2$), steroid withdrawal syndrome ($n=2$), FM ($n=2$), SSc ($n=2$), remitting seronegative symmetrical synovitis with pitting oedema ($n=1$), humeral epicondylitis ($n=1$), diffuse fasciitis ($n=1$), sarcoidosis ($n=1$), infectious endocarditis ($n=1$), acute respiratory distress syndrome ($n=1$), amyloid arthropathy ($n=1$), SAPHO (synovitis, acne, pustulosis, hyperostosis, osteitis) syndrome ($n=1$) and HScP ($n=1$).

Comparison of the 2010 and 1987 criteria

Table 1 presents a comparison of the 2010 and 1987 criteria. For 82 subjects who had at least one swollen joint not better explained by other diseases, the sensitivity of the 2010 criteria was much better than that of the 1987

criteria, but the specificity was worse (73.5 vs 47.1% and 71.4 vs 92.9%, respectively). The positive predictive values (PPVs) were comparable, and the negative predictive values (NPVs) and the positive likelihood ratios were better in the 1987 criteria (92.5 vs 97.0%, 35.7 vs 26.5% and 2.6 vs 6.6%, respectively).

Features of RA patients with or without classification as RA under the new criteria

Features of 68 RA patients with or without classification as RA are shown in Table 2. The positivity of RF and/or anti-CCP and the swollen/tender small joint counts were significantly higher in patients who were classifiable as definite RA under the 2010 criteria than in those who

TABLE 1 Comparison of sensitivity, specificity and accuracy between the 1987 and 2010 criteria

Outcome measure	1987 criteria ($n = 82$)	2010 criteria				
		Subjects with swollen joints not better explained by other diseases			All subjects with swollen joints	All subjects
		($n = 82$)	Seropositive[a] ($n = 54$)	Seronegative[b] ($n = 28$)	($n = 124$)	($n = 313$)
Sensitivity, %	47.1	73.5	95.9	15.8	73.5	72.4
Specificity, %	92.9	71.4	20.0	100	80.4	89.9
PPV, %	97.0	92.5	90.4	100	82.0	69.6
Negative prediction value, %	26.5	35.7	33.3	36.0	71.4	91.0
Positive likelihood ratio	6.6	2.6	1.2	NA	3.8	7.2

[a]RF and/or anti-CCP was positive. [b]Both RF and anti-CCP were negative. NA: not available.

TABLE 2 Comparison of RA patients who were classifiable as RA with those who were not

Characteristic	Classifiable ($n = 50$)	Not classifiable ($n = 18$)	P-value
RF positivity	43 (86)	1 (6)	*<0.001*
Low titre[a]	17	0	
High titre[a]	26	1	
Anti-CCP positivity	40 (80)	0 (0)	*<0.001*
Low titre[a]	3	0	
High titre[a]	37	0	
CRP positivity	32 (64)	14 (78)	0.38
CRP level, mean (S.D.), mg/dl	2.2 (3.1)	2.4 (3.4)	0.83
ESR positivity	45 (90)	15 (83)	0.43
ESR level, mean (S.D.), mm/h	55 (39)	54 (44)	0.44
Swollen small joint count[b], mean (S.D.)	5.1 (4.9)	2.5 (2.2)	*0.01*
Swollen large joint count[b], mean (S.D.)	1.6 (1.8)	0.6 (0.9)	0.09
Tender small joint count[b], mean (S.D.)	3.7 (3.3)	1.9 (1.9)	*0.006*
Tender large joint count[b], mean (S.D.)	1.6 (2.2)	0.8 (1.3)	0.76

Values are *n* (%) unless otherwise indicated. *P* values < 0.05, given in italics, were considered to be statistically significant. [a]High titre was defined as a value that was more than three times the upper limit in our own institute, following the new criteria. [b]Small/large joint was determined in accordance with the 2010 criteria.

were not. We divided 82 subjects into two groups according to the presence or absence of RF and/or anti-CCP, and the sensitivity and specificity were re-evaluated, as shown in Table 1. In the group of patients in whom RF and anti-CCP were both negative, sensitivity decreased remarkably to 15.8%. If we could include all subjects who had at least one swollen joint ($n = 124$), or all subjects who were recruited in the present study ($n = 313$), in the 2010 criteria, the specificity would increase (Table 1).

RA patients who were not classifiable as RA under the new criteria

Eighteen RA patients were not classifiable as RA under the new criteria at the point when the first laboratory and radiographic findings became available. Six patients exhibited small erosions on X-rays, but these were not significant and it was not obvious whether their histories were compatible with the 2010 criteria. Five patients were diagnosed with RA by reference to MRI findings. In

addition, eight patients who had not had any swollen joints on the first visit and had not been subjected to the scoring system were later found to have swollen joints and were diagnosed with RA. Seven of a total of 26 patients had come to satisfy the 2010 criteria as definite RA during the period of 3–33 weeks from the first visit, whereas the others were treated with DMARDs before being able to fulfil the new criteria.

Cases of patients with other diagnoses who achieved a total score of ⩾6 under the 2010 criteria

If the 2010 criteria were applied to all subjects who were recruited in the present study, 11 subjects given other diagnoses achieved a total score of ⩾6. The features of these patients are shown in Table 3. Their diagnoses included three NAD, one UPIA, one PsA, two OA, one PMS, one SS, one SLE and one DM. Except for cases with an arthritis similar to RA, NAD and OA subjects with a minor count of swollen joints, high-titre RF positivity and

TABLE 3 Features of non-RA patients who achieved total score of ⩾6

Physical and experimental findings on the first visit										
Sex	Age	TJC	SJC	ESR	CRP	RF	Anti-CCP	Duration, weeks	Diagnosis	Treatment
F	31	1	3	28	0.9	41	0	21	UPIA	–
F	34	1	1	8	0.01	76	0	265	NAD	–
F	59	0	2	10	0.03	72	0	18	NAD	–
F	41	1	1	16	0.11	0	31	28	NAD	
M	46	3	2	58	10.22	153	100	12	PsA	MTX
M	43	10	11	26	0.06	0	0	9	DM	PSL
F	50	4	5	14	0.06	28	9.8	14	SLE	NSAIDs
F	55	2	1	18	0.02	64	0	104	PMS	–
F	53	1	1	18	0.1	79	0	52	OA	–
F	62	3	2	13	0.1	0	23	520	OA	–
F	57	14	6	4	0.01	0	0	11	SS	–

TJC: tender joint count (both small and large); SJC: swollen joint count (both small and large); PSL: prednisolone.

mildly elevated ESR were apt to be misclassified as having RA under the 2010 criteria.

Availability of scoring system and difficulties in detection of swollen joints and differential diagnoses

In the present study, 82 patients were subjected to the scoring system. A ROC curve depicted to decide the best cut-off score showed that the best was 6, as was the same with the definition of 2010 criteria (data not shown). However, it is not always easy to detect swollen joints and accurately make diagnoses of other diseases. Supposing a doctor had difficulty in assessing swollen joints and differential diagnoses, we tried to assign all 313 subjects to the scoring system. The results are shown in Fig. 2A. The median score was 7 in RA and 2 in non-RA subjects. A ROC curve in this setting was depicted (Fig. 2B) and the ROC plot that was the closest to the upper left corner was a score of 5 in this setting.

Discussion

Over the past decade, the clinical setting of RA has changed considerably. Destructive joint damage was shown to begin at an early stage [15, 16], and an early diagnosis with aggressive therapy may alter or modify the natural history of this destructive and dreadful disease [17]. The 1987 ACR classification criteria used widely to diagnose RA have been criticized for their low discriminative ability in recent onset arthritis [9, 10]. The main cause of this was that the 1987 criteria were created using data from established RA patients with a mean disease duration of 7.7 years [8]. Harrison et al. [9] reported that the Norfolk Arthritis Register data showed that only 38% of new cases of inflammatory polyarthritis could be classified as RA using the 1987 criteria when first seen. Moreover, only 50% of RA patients satisfied the 1987 criteria at 6 months and only 80% even at 2 years after enrolment [9]. Thus, the 2010 classification criteria were developed in order to

distinguish RA earlier and start effective treatment as soon as possible to prevent or minimize joint destruction [7, 11–13].

At the time the 1987 criteria were declared, sensitivity and specificity were reported to be 91–94 and 89%, respectively [8]. In our study, sensitivity and specificity were 47.1 and 92.9%, respectively, using the 1987 criteria, while those using the 2010 criteria were 73.5 and 71.4%. van der Linden et al. [18] reported that both the sensitivity and specificity of the 2010 criteria were 74% when using DMARD-initiation within the first year as RA outcome in the Leiden Early Arthritis clinic. Our data were quite similar to their results. The sensitivity was better under the 2010 criteria, although the specificity, NPV and the likelihood ratio were better under the 1987 criteria. These results demonstrate that the 2010 criteria are superior to the 1987 criteria for the detection of RA in early stages, rather than for diagnoses. However, sensitivity under the new criteria decreased to 15.8% when both RF and anti-CCP were negative, which is considered to be a limitation of the new criteria. For example, a seronegative patient with 10 swollen/tender joints and elevated CRP and ESR for >6 weeks, who was strongly suspected to have a persistent and destructive disease (i.e. RA), could not achieve a total score of 6.

Eighteen RA patients and an additional eight patients without any swollen joints when first seen, were not classifiable as RA under the 2010 criteria at the point when the first laboratory and radiographic findings became available. Among these, while 19 patients had been treated with DMARDs before being subjected to the new criteria and could not be considered assessable because of improvement, the other seven patients who were just observed with or without NSAIDs came to be classifiable as having RA within 33 weeks (six within 12 weeks and one at 33 weeks). When we subjected the patients to the new criteria cumulatively over 12 weeks, the sensitivity increases up to at least 81.6%. It can be said that these criteria are useful to diagnose RA within 12 weeks, even

Fig. 2 (A) Scores of 313 subjects under the scoring system. The distribution of the scores at the point when first laboratory and radiographic findings were available is shown. The median scores were 7 and 2 in RA patients and non-RA subjects, respectively. (**B**) A ROC curve. A ROC curve was depicted to decide the best cut-off score. The ROC plot that was the closest to the upper left corner was a score of 5.
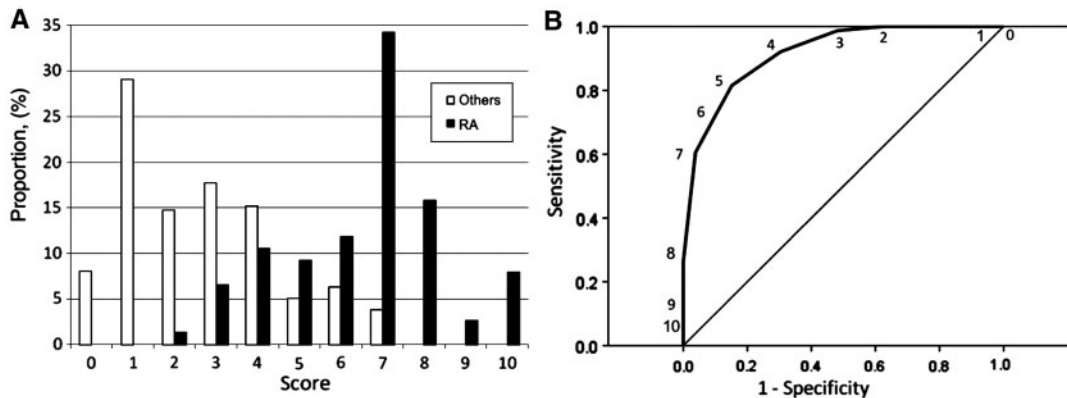
if we could not classify patients as having RA when first seen.

Meanwhile, except for cases with an arthritis similar to RA, NAD and OA subjects with one or two small swollen joints, non-specific high-titre RF positivity and mildly elevated ESR tended to be misclassified as having RA. If we were to classify such subjects as RA and start treatment with DMARDs, we might overtreat them. So we should be careful with this point when using the 2010 criteria.

The utility of the scoring system in various situations was also verified. Even if swollen joints and other diseases could not be accurately assessed, that is, if all patients with joint symptoms were to be submitted to the 2010 criteria, the sensitivity would be comparable and the specificity would be raised to 89.9%. Considering the result from the ROC curve, we could make presumptions about whether subjects with joint symptoms might be affected with RA if they were to attain a cut-off score of 5. Young *et al.* [19] reported that there has been little change in referral time from onset of symptoms to a rheumatologist over 25 years in a large RA inception register in the UK. It is important to avoid delay in consultation to rheumatologists as well as to make an early diagnosis of RA. If primary care physicians were to use the 2010 criteria, they might better refer patients to a rheumatologist or at least monitor them carefully under the UPIA recommendation [20] with a score of 5, so as not to miss RA patients.

There are some limitations to this study. One of these was the definition of the gold standard for RA. This definition contained risk of misdiagnosis. And the data used by rheumatologists in our institution to diagnose RA were partly corresponding to items of the new criteria, so the sensitivity might be highly overestimated. However, since the six rheumatologists who diagnosed the subjects in this study were all specialists in rheumatology, each with >10 years of clinical experience, almost all of the diagnoses were believed to be correct. Moreover, we determined the institution of not only MTX but also other DMARDs to be the gold standard. Since in our country, MTX is permitted for use by the Health, Labour and Welfare Ministry only after other DMARDs fail, only 44 (57%) of 76 patients had MTX initiated as their first treatment. Another limitation was that this study was a hospital-based study. Since our hospital is a major academic medical institute, there is a possibility that many of our subjects were more likely to have RA, and the PPV might be estimated as higher than it really is.

In conclusion, the present study showed that the 2010 classification criteria have high sensitivity and are useful for distinguishing early RA. However, it should be cautioned that the sensitivity decreased remarkably when both RF and anti-CCP were negative and that subjects with a small number of swollen joints, non-specific high-titre RF positivity and mildly elevated ESR were apt to be misclassified as having RA. If general physicians use the 2010 criteria to distinguish RA, a cut-off score of 5 would be better in order not to miss RA patients. Further studies with a larger cohort may be needed to optimize these criteria in the practical field.

Rheumatology key messages

- The 2010 classification criteria have high sensitivity and are useful for distinguishing early RA.
- The sensitivity of the 2010 classification criteria decreased remarkably when both RF and anti-CCP were negative.
- A cut-off score of 5 might be better in the practical field.

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

References

1 Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. Lancet 2009;373:659–72.

2 Breedveld FC, Weisman MH, Kavanaugh AF et al. The PREMIER study: a multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. Arthritis Rheum 2006;54:26–37.

3 Lipsky PE, van der Heijde DM, St Clair EW et al. for the Anti-tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. Infliximab and methotrexate in the treatment of rheumatoid arthritis. N Engl J Med 2000;343:1594–602.

4 Klareskog L, van der Heijde D, de Jager JP et al. TEMPO (trial of etanercept and methotrexate with radiographic patient outcomes) study investigators. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. Lancet 2004;363:675–81.

5 Yokota S, Imagawa T, Mori M et al. Efficacy and safety of tocilizumab in patients with systemic-onset juvenile idiopathic arthritis: a randomised, double-blind, placebo-controlled, withdrawal phase III trial. Lancet 2008;371:998–1006.

6 Nam JL, Winthrop KL, van Vollenhoven RF et al. Current evidence for the management of rheumatoid arthritis with biological disease-modifying antirheumatic drugs: a systematic literature review informing the EULAR recommendations for the management of RA. Ann Rheum Dis 2010;69:976–86.

7 Smolen JS, Landewé 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.

8 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.

9 Harrison BJ, Symmons DP, Barrett EM et al. The performance of the 1987 ARA classification criteria for rheumatoid arthritis in a population based cohort of patients with early inflammatory polyarthritis. American Rheumatism Association. J Rheumatol 1998;25:2324–30.

10 Harrison B, Symmons D. Early inflammatory polyarthritis: results from the Norfolk Arthritis Register with a review of the literature. II. Outcome at three years. Rheumatology 2000;39:939–49.

11 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.

12 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.

13 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.

14 Bossuyt PM, Reitsma JB, Bruns DE et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem 2003;49:1–6.

15 Wolfe F, Sharp JT. Radiographic outcome of recent-onset rheumatoid arthritis: a 19-year study of radiographic progression. Arthritis Rheum 1998;41:1571–82.

16 van der Heijde DM, van Leeuwen MA, van Riel PL, van de Putte LB. Radiographic progression on radiographs of hands and feet during the first 3 years of rheumatoid arthritis measured according to Sharp's method (van der Heijde modification). J Rheumatol 2005;22: 1792–6.

17 Verstappen SM, Jacobs JW, Bijlsma JW et al. Utrecht Arthritis Cohort Study Group. Five-year followup of rheumatoid arthritis patients after early treatment with disease-modifying antirheumatic drugs versus treatment according to the pyramid approach in the first year. Arthritis Rheum 2003;48:1797–807.

18 van der Linden MP, Knevel R, Huizinga TW et al. Classification of rheumatoid arthritis - comparison of the 1987 ACR and 2010 ACR/EULAR criteria. Arthritis Rheum. Advance Access published October 21, 2010, doi:10.1002/art.30100.

19 Young A, Dixey J, Williams P et al. An evaluation of the strengths and weaknesses of a register of newly diagnosed rheumatoid arthritis, 1986–2010. Rheumatology. Advance Access published October 7, 2010.

20 Machado P, Castrejon I, Katchamart W et al. Multinational evidence-based recommendations on how to investigate and follow-up undifferentiated peripheral inflammatory arthritis: integrating systematic literature research and expert opinion of a broad international panel of rheumatologists in the 3E Initiative. Ann Rheum Dis. Advance Access published August 19, 2010, doi: 10.1136/ard.2010.130625.