Sensitivity and specificity of 2010 rheumatoid arthritis classification criteria

Yuko Kaneko¹, Masataka Kuwana¹, Hideto Kameda¹ and Tsutomu Takeuchi¹

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...
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 in 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, bucillamine, 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 showed a 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), tendinitis (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),
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...
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).

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

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>1987 criteria</th>
<th>2010 criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>((n = 82))</td>
<td>((n = 82))</td>
</tr>
<tr>
<td>Sensitivity, %</td>
<td>47.1</td>
<td>73.5</td>
</tr>
<tr>
<td>Specificity, %</td>
<td>92.9</td>
<td>71.4</td>
</tr>
<tr>
<td>PPV, %</td>
<td>97.0</td>
<td>92.5</td>
</tr>
<tr>
<td>Negative prediction value, %</td>
<td>26.5</td>
<td>35.7</td>
</tr>
<tr>
<td>Positive likelihood ratio</td>
<td>6.6</td>
<td>2.6</td>
</tr>
</tbody>
</table>

\(a\)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

- RF positivity
- Low titre\(a\)
- High titre\(a\)
- Anti-CCP positivity
- Low titre\(a\)
- High titre\(a\)
- CRP positivity
- CRP level, mean (s.d.), mg/dl
- ESR positivity
- ESR level, mean (s.d.), mm/h
- Swollen small joint count\(b\), mean (s.d.)
- Swollen large joint count\(b\), mean (s.d.)
- Tender small joint count\(b\), mean (s.d.)
- Tender large joint count\(b\), mean (s.d.)

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

Values are \(n (%)\) unless otherwise indicated. \(\text{P-values } < 0.05, \text{ given in italics, were considered to be statistically significant.}\)

\(\text{High titre was defined as a value that was more than three times the upper limit in our own institute, following the new criteria.}\)

**Table 3** Cases of patients with other diagnoses who achieved a total score of \(\geq 6\) 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 \(\geq 6\). 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...
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 new 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

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

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


