Diagnostic performance of the ACR/EULAR 2010 criteria for rheumatoid arthritis and two diagnostic algorithms in an early arthritis clinic (REACH)

Celina Alves, Jolanda Jacoba Luime, Derkjen van Zeven, Anne-Margriet Huisman, Angelique Elisabeth Adriana Maria Weel, Pieternella Johanna Barendregt, Johanna Maria Wilhelmina Hazes

ABSTRACT

Introduction An ACR/EULAR task force released new criteria to classify rheumatoid arthritis at an early stage. This study evaluates the diagnostic performance of these criteria and algorithms by van der Helm and Visser in REACH.

Methods Patients with symptoms ≤12 months from REACH were used. Algorithms were tested on discrimination, calibration and diagnostic accuracy of proposed cut-points. Two patient sets were defined to test robustness; undifferentiated arthritis (UA) (n=231) and all patients including those without synovitis (n=513). The outcomes evaluated were methotrexate use and persistent disease at 12 months.

Results In UA patients all algorithms had good areas under the curve 0.79, 95% CI 0.73 to 0.83 for the ACR/EULAR criteria, 0.80, 95% CI 0.74 to 0.87 for van der Helm and 0.83, 95% CI 0.77 to 0.88 for Visser. All calibrated well. Sensitivity and specificity were 0.74 and 0.66 for the ACR/EULAR criteria, 0.1 and 1.0 for van der Helm and 0.59 and 0.93 for Visser. Similar results were found in all patients indicating robustness.

Conclusion The ACR/EULAR 2010 criteria showed good diagnostic properties in an early arthritis cohort reflecting daily practice, as did the van der Helm and Visser algorithms. All were robust. To promote uniformity and comparability the ACR/EULAR 2010 criteria should be used in future diagnostic studies.
citrullinated peptide (Elia CCP on immunoCAP 250; Phadia Freiburg, Germany), C-reactive protein (local standards) and ESR (local standards). X-rays of hands and feet were assessed for bony erosions at baseline. For a detailed description of REACH, see Geuskens et al.14

Statistical analyses
To assess overall performance the prediction algorithms were tested on discrimination and calibration.15 Discrimination is the ability of an algorithm to differentiate correctly between patients with and without the disease. Calibration reveals the ability to estimate the probability of the diagnosis for individuals correctly by comparing the probability predicted by the algorithm and the observed probability. To evaluate discrimination receiver operating characteristic curves, including corresponding areas under the curve (AUC), were calculated. Calibration was evaluated using calibration plots and the Hosmer–Lemeshow test.15 The latter indicates good calibration if a non-significant result appears. To assess diagnostic performance of the algorithms sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV) were estimated at cut-points proposed for treatment initiation among patients at risk of RA. For the ACR/EULAR 2010 criteria and Visser algorithm a score of 6 or more16 was used and for van der Helm a score of 8 or more15 was used. To test robustness this analysis was repeated among all patients included in REACH. This group had a lower previous disease probability by a case-mix of synovitis and inflammatory joint complaints without synovitis. Synovitis was defined as joint swelling.

As a classifier for correct diagnosis two outcomes were evaluated at 1 year: the use of methotrexate and persistent disease, defined as synovitis present at physical examination after 1 year, or the use of disease-modifying antirheumatic drugs (DMARD) including biological agents. Patients with a definite alternative diagnosis such as gout were not classified as persistent disease. A complete case analysis was done.

RESULTS
Validation cohort
Up to 31 October 2008, 875 patients were referred to REACH and had 1-year follow-up. One hundred and 13 patients did not fulfill the inclusion criteria and 31 patients were lost to follow-up at baseline (see supplementary figure S1, available online only). Patients used in the development of the ACR/EULAR 2010 criteria were excluded (n=216).1 Table 1 reports baseline characteristics of all patients (n=513). Patients had a mean age of 50 years, 73% were women and the median symptom duration was 106 days (range 1–366 days). At baseline 48% (n=246) were presented with synovitis. After 1 year, 148 of 513 used methotrexate, of whom 22 did not have synovitis at baseline, and 231 of 513 patients had persistent disease, of whom 59 did not have synovitis at baseline.

Discrimination
Table 2 shows AUC of each diagnostic algorithm for both outcomes. In undifferentiated arthritis (UA) patients (n=231) the AUC for methotrexate use were comparable, with overlapping 95% CI, 0.79 (95% CI 0.73 to 0.83) for the ACR/EULAR 2010 criteria, 0.80 (95% CI 0.74 to 0.87) for the van der Helm algorithm and 0.85 (95% CI 0.77 to 0.88) for the Visser algorithm. For persistent disease the AUC were 0.77 (95% CI 0.71 to 0.85) for the ACR/EULAR 2010 criteria, 0.78 (95% CI 0.71 to 0.85) for the van der Helm algorithm and 0.77 (95% CI 0.71 to 0.85) for the Visser algorithm. In all patients (n=513) the AUC were comparable for both outcomes, with slightly better performance of the van der Helm algorithm; 0.88 (95% CI 0.84 to 0.91) and 0.83 (95% CI 0.79 to 0.87).

Calibration
Calibration plots of all diagnostic algorithms are shown in figure S2 (see supplementary figure S2, available online only). In UA patients (n=513) calibration was worse than in all patients, although the Hosmer–Lemeshow test was not significant for any of the calibration plots. All algorithms calibrated well in all patients (n=513), confirmed by the Hosmer–Lemeshow test.

Evaluating diagnostic performance using proposed cut-points
To identify patients in need of treatment proposed cut-points were tested in UA patients. The ACR/EULAR criteria showed a sensitivity of 0.74 (95% CI 0.65 to 0.82) and a specificity of 0.66 (95% CI 0.54 to 0.76), with the cut-point of 6 or higher using methotrexate as a classifier for correct diagnosis (table 3). The Visser algorithm and the van der Helm algorithm had a lower sensitivity, 0.47 and 0.59 for the Visser algorithm for both outcomes and 0.08 and 0.10 for the van der Helm algorithm. Specificity was higher: 0.93 for the Visser algorithm and 1.0 for the van der Helm algorithm.

The PPV is the probability that a patient has the disease if the test is positive. The van der Helm algorithm had the highest PPV; 1.0. The NPV is the opposite probability and was highest for the ACR/EULAR criteria with 0.65 for methotrexate use and 0.46 for persistent disease, slightly higher than the Visser algorithm.

DISCUSSION
The results of our study show that the new ACR/EULAR 2010 criteria could aid diagnostics in early arthritis patients. They had good overall performance, with a sufficiently high AUC and good performance of the proposed cut-point of 6 for persistent disease, which could be considered RA. The other algorithms performed well when tested for discriminatory properties.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Patient characteristics for each patient set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UA (n=231)</td>
</tr>
<tr>
<td>Women (%)</td>
<td>68 (14)</td>
</tr>
<tr>
<td>Age, years (mean, SD)</td>
<td>53 (14)</td>
</tr>
<tr>
<td>SJC (median, range)</td>
<td>4 (1–38)</td>
</tr>
<tr>
<td>TJC (median, range)</td>
<td>7 (0–42)</td>
</tr>
<tr>
<td>RF positive (%)</td>
<td>35%</td>
</tr>
<tr>
<td>Anti-CCP positive (%)</td>
<td>26% mv=6</td>
</tr>
<tr>
<td>ESR, mm/h (median, range)</td>
<td>18 (1–103) mv=7</td>
</tr>
<tr>
<td>CRP, mg/l (median, range)</td>
<td>6 (1–180) mv=16</td>
</tr>
<tr>
<td>Erosions (%)</td>
<td>9% mv=4</td>
</tr>
<tr>
<td>RA, according to 1987 ACR criteria</td>
<td>29% mv=3</td>
</tr>
<tr>
<td>RA, according to 2010 ACR/EULAR criteria</td>
<td>45% mv=12</td>
</tr>
<tr>
<td>Persistent arthritis at 1 year</td>
<td>45% mv=9</td>
</tr>
</tbody>
</table>

ACR, American College of Rheumatology; CCP, cyclic citrullinated protein; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; EULAR, European League Against Rheumatism; mv, missing values; RA, rheumatoid arthritis; RF, rheumatoid factor; SJC, swollen joint count; TJC, tender joint count; UA, undifferentiated arthritis.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Area under the receiver operating characteristic curves with 95% CI for each algorithm and each patient set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACR/EULAR 2010</td>
</tr>
<tr>
<td>Outcome methotrexate use</td>
<td>0.79 (0.73 to 0.85)</td>
</tr>
<tr>
<td>All patients</td>
<td>0.79 (0.75 to 0.83)</td>
</tr>
<tr>
<td>Outcome persistent disease</td>
<td>0.77 (0.71 to 0.85)</td>
</tr>
<tr>
<td>All patients</td>
<td>0.74 (0.70 to 0.78)</td>
</tr>
</tbody>
</table>

UA, undifferentiated arthritis.
We defined two outcomes; methotrexate use similar to the definition of the ACR/EULAR 2010 and persistent disease (either synovitis or DMARD use at 12 months). This may have led to misclassification in two ways. Patients could be classified as true positive because they were still using methotrexate or other DMARD at 12 months, whereas in fact some patients may not need treatment. Likewise, patients may have had episodes of arthritis with no episode or DMARD use at 12 months, while later on they developed persistent arthritis.

In conclusion, the new ACR/EULAR 2010 criteria showed good diagnostic properties in an early arthritis cohort reflecting daily clinical practice, as did the van der Helm and Visser algorithms. All were robust. To promote uniformity and comparability we would suggest using the ACR/EULAR 2010 criteria in future diagnostic studies.

Acknowledgements The authors would like to thank all patients and general practitioners who participated as well as the research nurses, trainee rheumatologists and rheumatologists (other than the authors) who contributed to the REACH cohort.

Funding Data management was partly funded by the Dutch Arthritis Association.

Competing interests None.

Ethics approval This study was conducted with the approval of the medical ethics committee at Erasmus MC.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

Diagnostic performance of the ACR/EULAR 2010 criteria for rheumatoid arthritis and two diagnostic algorithms in an early arthritis clinic (REACH)

Celina Alves, Jolanda Jacoba Luime, Derkjen van Zeven, Anne-Margriet Huisman, Angelique Elisabeth Adriana Maria Weel, Pieternella Johanna Barendregt and Johanna Maria Wilhelmina Hazes

Ann Rheum Dis published online May 27, 2011

Updated information and services can be found at:
http://ard.bmj.com/content/early/2011/05/27/ard.2010.142299

These include:

**Supplementary Material**
Supplementary material can be found at:
http://ard.bmj.com/content/suppl/2011/05/07/ard.2010.142299.DC1

**References**
This article cites 14 articles, 3 of which you can access for free at:
http://ard.bmj.com/content/early/2011/05/27/ard.2010.142299#BIBL

**Open Access**
This paper is freely available online under the BMJ Journals unlocked scheme, see http://ard.bmj.com/info/unlocked.dtl

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Topic Collections**
Articles on similar topics can be found in the following collections:

- Open access (593)
- Degenerative joint disease (4641)
- Musculoskeletal syndromes (4951)
- Connective tissue disease (4253)
- Immunology (including allergy) (5144)
- Rheumatoid arthritis (3258)

**Notes**

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/