The performance of anti-cyclic citrullinated peptide assays in diagnosing rheumatoid arthritis: a systematic review and meta-analysis

L. Mathsson Alm1,3, D.L. Fountain2, K.K. Cadwell2, A.M. Madrigal2, G. Gallo3, M. Poorafshar3

ABSTRACT

Objective. We assessed the ability of anti-cyclic citrullinated peptide (CCP) tests to diagnose rheumatoid arthritis (RA), comparing the effect of manufacturer assay type, study design (single- and two-gated) and duration of disease (early vs. established).

Methods. We searched seven databases for relevant diagnostic studies containing data on CCP tests in known or suspected RA patients. We used a bi-variate model to produce summary estimates for test sensitivity, specificity, and positive and negative likelihood ratios. Summary Receiver Operating Characteristic (sROC) curves were derived to compare early versus established RA.

Results. 83 studies were identified and included. For individual manufacturer tests there was considerable variation in both pooled sensitivity (range 67–83%) and specificity (range 90–96%) estimates. This heterogeneity was also observed when grouping studies into two-gated and single-gated designs. Study design and disease duration impacted on sensitivity, with single-gated study designs and early RA patients resulting in lower estimates than two-gated and established disease, respectively.

Conclusion. This review highlights the large number of CCP tests that are now commercially available and the considerably variation in their diagnostic performance. This variation, although partly influenced in this analysis by the study design (single-gated vs. two-gated), seems to have the most consistent diagnostic performance overall.

Introduction

Rheumatoid arthritis (RA) is a common, chronic autoimmune systemic inflammatory disease affecting between 0.5% and 1% of the adult population worldwide (1, 2). The prevalence of RA increases with age and is more common in women than men. Although the exact cause of RA remains unknown it is thought that it results from environmental exposure in genetically susceptible individuals (3). Specifically, a recent review has highlighted the importance of genetic loci and polymorphisms that impact upon the expression of human leukocyte antigen (HLA) molecules, lymphocyte activation, and protein citrullination pathways (4).

RA is characterised by chronic inflammation of the synovial joints and typically presents as stiffness, pain and swelling in one or more joints, usually the small joints of the hands and feet. If left untreated the chronic inflammation will induce the formation of pannus tissue, leading to joint, cartilage, and bone destruction. Patients with RA have significantly reduced health related quality of life caused by the pain, fatigue and loss of bodily function associated with the disease progression (5), increased risk of lung and cardiovascular disease (4) and premature mortality (6). It is vital that RA is diagnosed early in order to ensure the disease is effectively treated; health outcomes have been shown to improve as patients receive early aggressive disease-modifying anti-rheumatic drugs (DMARDs) (e.g. methotrexate) (5). The American College of Rheumatology (ACR) classification criteria is con-
sidered the gold standard and accepted method of diagnosing RA. In 2010 the ACR and European League against Rheumatism (EULAR) developed a new classification criteria for RA to facilitate the identification of patients at an earlier stage of the disease process (7). The 2010 ACR classification criteria for RA are based on the confirmed presence of synovitis in at least one joint, absence of an alternative diagnosis that better explains the synovitis, and achievement of a total score of 6 or greater (out of a possible 10) using an algorithm with four categories: joint involvement, serology, acute phase and duration of symptoms (8).

The ACR criteria utilise the presence or absence of rheumatoid factor (RF), a high-affinity autoantibody directed against the Fc portion of immunoglobulin (9), as well as antibodies against peptides and proteins containing citrulline, a modified form of the amino acid arginine (10). In recent years studies have indicated that anti-citrullinated peptide antibodies (ACPA) tests, including anti-cyclic citrullinated peptide (CCP) tests, are more specific than RF for the diagnosis of RA (11, 12). Furthermore, anti-citrullinated peptide antibodies are thought to anticipate the clinical manifestation of RA by many years and could be useful biomarkers for predicting RA development (7, 13).

Four systematic reviews have previously been published on anti-CCP tests for diagnosing rheumatoid arthritis (11, 12, 14, 15). However, there has been no recent comprehensive review of the literature, in all relevant databases, to assess the performance of the wide range of anti-CCP tests available on the market to diagnose RA. Therefore here, we performed a new systematic literature review of the diagnostic accuracy of anti-cyclic citrullinated peptide (CCP) tests for RA diagnosis. The primary analysis of this study focussed on the variation in the performance of CCP tests currently available. In light of recently published data we also performed subgroup meta-analyses to evaluate how manufacturer variability in diagnostic performance was impacted by study design (single vs. two gated), and the ability of CCP tests to diagnose early versus established RA.

Methods
We prepared a review protocol prior to the start of the study, which was based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) guidelines for systematic review and meta-analysis (16).

Search methods for identification of studies
We conducted an electronic literature search from 2004 to March 2015 in Medline, Medline In-Process, Embase, PubMed, Cochrane Central Register of Controlled Trials, NIH Clinicaltrials.gov, and the WHO International Clinical Trials Registry Platform. The electronic search was supplemented by hand searching of reference lists of studies identified as eligible for inclusion, as well as several recently published systematic reviews. We applied no country limit. Language was limited to English. The search strategies focused on terms associated with rheumatoid arthritis, autoantibodies and anti-citrullinated protein antibodies.

Study selection criteria
Studies were selected for inclusion based on pre-defined eligibility criteria. Diagnostic case-control (two-gate), cross-sectional and cohort studies (single-gate) using CCP diagnostic tests (index test) attributed to a named manufacturer were eligible. Study participants had suspected or known RA, and were diagnosed according to the 1987 or 2010 ACR criteria for diagnosing RA. Control participants could be healthy or affected by rheumatological diseases other than RA. Studies were required to report sensitivity and specificity using the manufacturers recommended cut-off threshold value. Reviews, discussion papers, non-research letters, editorials, qualitative studies and case studies were not eligible.

Two reviewers independently screened titles and abstracts and disagreements were resolved through participation of early and established; (iv) status of control participants (healthy, diseased or mixed); (v) index test characteristics (including test name, manufacturer, assay type, generation, and cut-off); and (vi) diagnostic performance of the index test (sensitivity and specificity, positive and negative likelihood ratios, positive and negative predictive values, and odds ratios).

We extracted data for all reported index tests. Where reported, we extracted data separately by RA type (early, established or mixed) and control type (healthy, diseased or mixed).

Assessment of methodological quality
The methodological quality of each study was assessed using the QUADAS-2 tool (17), as recommended by The Cochrane Collaboration. QUADAS-2 is designed to assess the quality of primary diagnostic accuracy studies. The tool has four domains: (i) patient selection; (ii) index test; (iii) reference standard; and (iv) patient flow. Two reviewers independently assessed the methodological quality of each study. Disagreements were resolved through discussion and consensus, with a third reviewer used as arbiter where necessary.

Sub-analyses
Using sensitivity, specificity and likelihood ratio plots we visually compared the ability of single and two-gated study designs to accurately diagnose RA by calculating pooled estimates.
Diagnostic performance of CCP tests in RA / L. Mathsson Alm et al.

Table I. Summary of the study characteristics.

<table>
<thead>
<tr>
<th>Number of gates</th>
<th>RA type</th>
<th>Control type</th>
<th>Number of studies</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-gated</td>
<td>Early</td>
<td>Non-RA</td>
<td>14</td>
<td>(22-35)</td>
</tr>
<tr>
<td>Single-gated</td>
<td>Mixed</td>
<td>Non-RA</td>
<td>2</td>
<td>(36, 37)</td>
</tr>
<tr>
<td>Single-gated</td>
<td>NR</td>
<td>Non-RA</td>
<td>6</td>
<td>(38-43)</td>
</tr>
<tr>
<td>Two-gated</td>
<td>Early</td>
<td>Healthy</td>
<td>4</td>
<td>(45-48)</td>
</tr>
<tr>
<td>Two-gated</td>
<td>Early</td>
<td>Mixed</td>
<td>3</td>
<td>(49-51)</td>
</tr>
<tr>
<td>Two-gated</td>
<td>Established</td>
<td>Disease</td>
<td>3</td>
<td>(52-54)</td>
</tr>
<tr>
<td>Two-gated</td>
<td>Established</td>
<td>Healthy</td>
<td>3</td>
<td>(46, 53, 55)</td>
</tr>
<tr>
<td>Two-gated</td>
<td>Established</td>
<td>Mixed</td>
<td>4</td>
<td>(51, 56-58)</td>
</tr>
<tr>
<td>Two-gated</td>
<td>Mixed</td>
<td>Disease</td>
<td>8</td>
<td>(59-66)</td>
</tr>
<tr>
<td>Two-gated</td>
<td>Mixed</td>
<td>Healthy</td>
<td>4</td>
<td>(60, 67-69)</td>
</tr>
<tr>
<td>Two-gated</td>
<td>Mixed</td>
<td>Mixed</td>
<td>6</td>
<td>(58, 70-74)</td>
</tr>
<tr>
<td>Two-gated</td>
<td>NR</td>
<td>Disease</td>
<td>17</td>
<td>(75-91)</td>
</tr>
<tr>
<td>Two-gated</td>
<td>NR</td>
<td>Healthy</td>
<td>1</td>
<td>(92)</td>
</tr>
<tr>
<td>Two-gated</td>
<td>NR</td>
<td>Mixed</td>
<td>13</td>
<td>(44, 52, 86, 93-104)</td>
</tr>
<tr>
<td>Two-gated</td>
<td>NR</td>
<td>Non-RA</td>
<td>1</td>
<td>(104)</td>
</tr>
</tbody>
</table>

RA: rheumatoid arthritis; NR: not reported.

for each CCP test manufacturer. To investigate the ability of CCP tests to accurately diagnose early RA, we performed a second sub-analysis, pooling results from studies reporting the sensitivity and specificity of CCP in early and established RA patient subgroups. Where studies reported data on more than one index test for the same patient population, we selected a single set of sensitivity and specificity data for the meta-analyses using the following manufacturer hierarchy: Euro-diagnostica > Axis-Shield > Inova Diagnostics > Thermo Fisher Scientific > Euroimmun > Abbott > Roche Diagnostics. Additionally, we prioritised CCP2 tests over CCP3 tests for inclusion in the analysis. This selection method was applied to the analysis to avoid pseudo-replication. Where studies reported sub-group analyses for early and established patients using the same patient control group, both data-sets were included in the respective analyses.

**Statistical analysis and meta-analysis**

To produce summary estimates for test sensitivity and specificity, we employed a random effects bivariate model (18). CCP test threshold was set to the manufacturer’s recommended cut-off. We constructed a 2x2 table for each study and populated it with values back-calculated from extracted sensitivity and specificity values (true-positive, false-positive, true-negative, false-negative). Where more than one study reported a CCP test, we calculated pooled estimates of sensitivity and specificity with 95% confidence intervals (CIs) and displayed them in forest plots. We derived summary positive and negative likelihood ratios from the summary estimates of sensitivity and specificity; likelihood ratios can be interpreted as follows: <0.1 often conclusive decrease in the likelihood of disease, =1 implies no change in the likelihood of disease, and >10 often conclusive increase in the likelihood of disease. For all analyses we used R software (www.R-project.org) and the package “mada” (meta-analysis of diagnostic accuracy), (cran.r-project.org/web/packages/mada/mada.pdf). For the sub-analyses of disease duration, we used the bivariate model to estimate summary sensitivity and specificity estimates, with 95% confidence and prediction regions, and derived summary receiver operating characteristic (ROC) curves. This approach allowed for assessment of between-study heterogeneity in sensitivity and specificity and for visualising the trade-off (negative correlation) between sensitivity and specificity commonly seen in diagnostic meta-analyses. The summary ROC curves were used to assess the heterogeneity between studies. The ROC graph plots sensitivity against the false-positive rate. The shape of the curve between these two fixed points depends on the discriminatory ability of the test. The area under the curve (AUC) equals 1 for a perfect test and 0.5 for a completely uninformative test. The AUC is the probability that a diseased individual will have a higher test result than a non-diseased individual, provided that all individuals are selected at random.

**Results**

**Search results and characteristics of the studies**

The flow of studies through the review process is shown in Supplementary Fig. S1. Our electronic search identified 3100 articles, of which 83 met the eligibility criteria and were included in the review. A single study reported data from a first generation CCP assay, 60 studies used second generation CCP2 assays, five studies used third generation CCP3 assays and 17 studies used both CCP2 and CCP3 assays. Twenty-two studies employed a single-gated study design in which a single group of patients suspected of RA were analysed. The remaining 61 studies adopted a two-gated (case-control) design. Amongst the single-gated studies, the majority focussed on early RA patients (14/22). Amongst the two-gated case control studies, eight had evaluable data-sets for early arthritis patients, ten for established arthritis patients, and 18 for mixed arthritis. Thirty-two two-gated studies did not report RA status or disease duration. A summary of study characteristics is shown in Table I.

**Diagnostic performance of CCP tests**

The pooled estimates for sensitivity, specificity, and positive and negative likelihood ratios for specific CCP tests are shown in Fig 1. We analysed 125 data-sets that reported diagnostic performance using 11 different index tests with a named manufacturer. We included data sets from 82 of the 83 studies: 77 CCP2 and 22 CCP3 data sets. One study was not included because it had assessed a first generation CCP test. There was considerable variation between the different tests for the pooled sensitivity estimates, which ranged
from 67% (Axis-Shield Diastat and EuroDiagnostica Immunoscan) to 83% (Abbott Architect). Despite this variation, with one exception, the observed differences in sensitivity between manufacturer tests were not found to be statistically significant. Only the Abbott Architect test was found to have a statistically superior pooled sensitivity estimate compared with another CCP test, namely the Inova Diagnostics Quanta Lite (CCP2) test. However, there was a corresponding trade-off in specificity as the Abbott Architect test was also the test with the lowest pooled specificity estimate. The pooled specificity estimates for all tests varied between 90% (Abbott Architect) and 96% (Euroimmun Anti-CCP ELISA and Thermo Fisher Scientific EliA). Both the Thermo Fisher Scientific EliA and Euroimmun Anti-CCP ELISA tests were found to have statistically higher specificity estimates than the Inova Diagnostics Quanta Lite 3.1 test. However, no other statistically significant differences in specificity were observed between the manufacturer tests. The pooled estimates for the positive likelihood ratios ranged from 7.24 (Abbott AxSYM) to 18.63 (Thermo Fisher Scientific and the negative likelihood ratios ranged from 0.19 (Abbott Architect) to 0.36 (Axis-Shield Diastat).

Sub-analysis: Diagnostic performance of manufacturer tests split by single and two-gated study design

Forest plots of pooled estimates for sensitivity and specificity for each manufacturer grouped by single-gated and two-gated study designs are shown in Fig 2. Forest plots of pooled estimates for positive and negative likelihood ratios grouped by single-gated and two-gated study designs are shown in Fig 3. For studies adopting a single-gated study design the highest pooled sensitivity estimate was observed for the Abbott AxSYM (84%) although there was a corresponding loss of specificity as a result compared to other manufacturers”, with the test displaying the lowest specificity estimate of 91%. In the single-gated analysis the Thermo Fisher Scientific EliA and Inova Diagnostics Quanta Lite (CCP2) tests gave the high-

A.

<table>
<thead>
<tr>
<th>Test Manufacturer</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbott Architect (CCP2)</td>
<td>0.83 [0.72, 0.90]</td>
<td>0.90 [0.83, 0.94]</td>
</tr>
<tr>
<td>Abbott AxSYM (CCP2)</td>
<td>0.70 [0.60, 0.87]</td>
<td>0.91 [0.84, 0.95]</td>
</tr>
<tr>
<td>Axis-Shield Diastat (CCP2)</td>
<td>0.67 [0.61, 0.72]</td>
<td>0.90 [0.83, 0.97]</td>
</tr>
<tr>
<td>Euro-diagnostica (CCP2)</td>
<td>0.73 [0.63, 0.80]</td>
<td>0.94 [0.88, 0.98]</td>
</tr>
<tr>
<td>Euroimmun Anti-CCP ELISA (CCP2)</td>
<td>0.67 [0.62, 0.73]</td>
<td>0.94 [0.89, 0.98]</td>
</tr>
<tr>
<td>Euroimmun Anti-CCP ELISA (CCP2)</td>
<td>0.70 [0.65, 0.77]</td>
<td>0.96 [0.90, 0.97]</td>
</tr>
<tr>
<td>Inova Diagnostics Quanta Lite (CCP2)</td>
<td>0.68 [0.64, 0.71]</td>
<td>0.95 [0.90, 0.96]</td>
</tr>
<tr>
<td>Inova Diagnostics Quanta Lite (CCP2)</td>
<td>0.73 [0.65, 0.79]</td>
<td>0.94 [0.88, 0.97]</td>
</tr>
<tr>
<td>Inova Diagnostics Quanta Lite 3.1 (CCP2)</td>
<td>0.70 [0.64, 0.78]</td>
<td>0.92 [0.86, 0.95]</td>
</tr>
<tr>
<td>Roche Diagnostics Eclipse (CCP2)</td>
<td>0.74 [0.63, 0.81]</td>
<td>0.94 [0.89, 0.97]</td>
</tr>
<tr>
<td>Thermo Fisher Scientific EliA (CCP2)</td>
<td>0.74 [0.71, 0.77]</td>
<td>0.96 [0.93, 0.98]</td>
</tr>
</tbody>
</table>

B.

<table>
<thead>
<tr>
<th>Test Manufacturer</th>
<th>Positive LR (95% CI)</th>
<th>Negative LR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbott Architect (CCP2)</td>
<td>7.98 [5.20, 12.28]</td>
<td>0.19 [0.12, 0.32]</td>
</tr>
<tr>
<td>Abbott AxSYM (CCP2)</td>
<td>7.24 [4.14, 12.40]</td>
<td>0.25 [0.16, 0.43]</td>
</tr>
<tr>
<td>Axis-Shield Diastat (CCP2)</td>
<td>13.27 [6.90, 21.70]</td>
<td>0.30 [0.22, 0.41]</td>
</tr>
<tr>
<td>Euro-diagnostica (CCP2)</td>
<td>11.89 [7.35, 17.00]</td>
<td>0.31 [0.24, 0.40]</td>
</tr>
<tr>
<td>Euroimmun Anti-CCP ELISA (CCP2)</td>
<td>11.42 [7.05, 15.02]</td>
<td>0.25 [0.18, 0.34]</td>
</tr>
<tr>
<td>Euroimmun Anti-CCP ELISA (CCP2)</td>
<td>17.85 [12.96, 24.07]</td>
<td>0.30 [0.25, 0.35]</td>
</tr>
<tr>
<td>Inova Diagnostics Quanta Lite (CCP2)</td>
<td>13.65 [8.19, 22.83]</td>
<td>0.30 [0.24, 0.37]</td>
</tr>
<tr>
<td>Inova Diagnostics Quanta Lite (CCP2)</td>
<td>11.83 [7.45, 18.70]</td>
<td>0.30 [0.24, 0.37]</td>
</tr>
<tr>
<td>Inova Diagnostics Quanta Lite 3.1 (CCP2)</td>
<td>0.44 [0.17, 1.07]</td>
<td>0.21 [0.10, 0.46]</td>
</tr>
<tr>
<td>Roche Diagnostics Eclipse (CCP2)</td>
<td>11.17 [6.20, 20.72]</td>
<td>0.28 [0.20, 0.45]</td>
</tr>
<tr>
<td>Thermo Fisher Scientific EliA (CCP2)</td>
<td>18.00 [12.01, 28.56]</td>
<td>0.27 [0.25, 0.30]</td>
</tr>
</tbody>
</table>

Fig. 1. The diagnostic accuracy of CCP2 and CCP3 index tests grouped by manufacturer. Forest plots showing pooled estimates and 95% CI for (A) sensitivity and specificity, and (B) positive and negative likelihood ratios for each test manufacturer.

Fig. 2. The pooled sensitivity and specificity estimates of manufacturer tests used in single and two-gated study designs. Forest plots showing pooled sensitivity and specificity estimates and 95% CIs for (A) single-gated study designs and (B) two-gated study designs for each test manufacturer.
est specificities (both estimated at 97%). The single-gated studies produced both the lowest sensitivity estimates and the largest range of estimates (52%–84%) in comparison with the two-gated designs (68%–73%). Moreover, with regard to Thermo Fisher Scientific EliA test displayed the highest positive likelihood ratio of 21.8 (95% CI, range 2.5–74.2), while the other manufacturers’ ranged between 7.2–19.0.

Sub-analysis: Diagnostic performance of CCP tests in early and established RA patients
Twenty-eight of the 83 studies included in the systematic literature review reported sensitivity and specificity data of CCP tests (all generations) in early RA, and 12 studies reported the sensitivity and specificity of CCP tests in established RA patients. Studies that reported on the diagnostic performance of CCP tests in mixed RA populations were not included in the analysis unless a sub-group analysis of either early or established patients was available. In early RA patients, the pooled CCP test sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio were 58%, 95%, 1.00, and 0.28, respectively (Fig. 4). In established RA patients, the pooled CCP test sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio were 73%, 96%, 2.4%, and 0.3%, respectively. The sROC analysis revealed non-overlapping 95% CIs between early and established RA patients, showing a trend that CCP testing has superior diagnostic performance in established RA patients compared to early RA patients (Fig. 5).

Study quality
The proportion of studies that fulfil each QUADAS-2 criterion is summarised in Supplementary Fig. S2. Of the 83 studies, only three satisfy all of the criteria in the QUADAS-2 check-list. In the patient selection domain, 71% of studies were classified at a high risk of bias for employing a case-control design. In the index test domain, 75% of studies were classified at a high risk of bias because CCP test results had been interpreted with knowledge of participant’s RA status. In the reference standard domain, only 16% of studies were at a high risk of bias because, for the majority of studies, the diagnosis of RA by the ACR criteria (the reference standard) had pre-dated the CCP test. In the flow and timing domain, none of the studies were considered to be at high risk of bias. In all studies the same reference standard was applied (diagnosis of RA by the ACR criteria) as specified in the systematic review inclusion criteria. In terms of applicability no studies were considered to be at risk of bias for either patient selection, the reference standard or index test domains.

Discussion
In the last 10 years, five systematic literature reviews investigating the accuracy of ACPA and RF tests in the diagnosis of RA have been conducted (11, 12, 14, 15, 19). In 2007, the meta-analysis performed by Nishimura (10)
led to the conclusion that anti-CCP antibody positivity is more specific than IgM RF positivity for diagnosing RA and early RA. The increase in evidence supporting the value of ACPA testing for the diagnosis of RA led to the incorporation of ACPA testing in the updated ACR guidelines in 2010 (7). Moreover, the combination of anti-CCP positivity and RF positivity has been revealed to increase the probability of a true-positive result relative to either of the antibody tests in isolation (19). Recent systematic reviews investigating the effect of CCP generations by Whiting et al. (2010) (12) and Zhang et al. (2014) (15) found no significant differences between CCP2 and CCP3; therefore, we did not explore CCP generation. Our current systematic review is the first to perform a meta-analysis to investigate variation between manufacturer assays and, additionally, included analyses of study design (single- and two-gated), and duration of disease (early compared to established RA).

Overall, in this analysis using manufacturers recommended cut-offs we have demonstrated that the current commercially available CCP tests have high specificity and reasonable sensitivity for the diagnosis of RA. Despite no single test proving to be statistically superior to another in diagnostic performance, the analysis did reveal some differences in the pooled estimates of sensitivity and specificity between the commercially available CCP assays. Additionally, considerable within-assay variability was revealed, in particular for single-gated studies. Further studies are needed to establish what the clinical impact of such heterogeneity in diagnostic accuracy might be.

Among variables analysed, study design has been recognised as a potentially important source of bias in diagnostic accuracy studies (17). To accurately reflect diagnosing RA in clinical practice, a diagnostic study should adopt a single-gated consecutive or random patient flow with all patients suspected of RA, and ideally, showing early symptom duration. In the absence of single-gated design, the preferred design is two-gated (i.e. case-control studies). It is preferable that two-gated case-control studies use controls with non-RA rheumatology diseases, rather than a healthy control group because comparing confirmed diseased groups to healthy groups will likely exaggerate diagnostic accuracy (17). To prevent over-estimation of both sensitivity and specificity studies should ideally include difficult-to-diagnose patients and difficult-to-rule-out controls. The visual comparison of sensitivity estimates between single- and two-gate studies performed in this study does suggest that the performance of CCP test is influenced by study design as, on the whole, the single-gated designs result in lower sensitivities and higher within-assay variability.

A comparison of manufacturers in the single-gate analysis, broadly representative of the scenario in which the test would be used in practice, revealed that based on the pooled positive likelihood ratios, the Thermo Fisher Scientific – EliA and the Inova Diagnostics Quanta Lite (CCP2) outperformed the other manufacturers (21.8 and 19, respectively vs. 7.2–13.6 for other manufacturers). The Thermo Fisher Scientific – EliA and the Inova Diagnostics Quanta Lite (CCP2) showed reliability for diagnosing suspected RA and implied that a positive result with these assays is often conclusive of the increase in the likelihood of disease. Overall, these tests showed lower between-study variations compared to other manufacturers, as demonstrated by the small confidence intervals on the pooled sensitivity estimates, thus showing reliability and consistency for ruling-in suspected RA. With regard to disease duration, our RA status sub-analysis suggests that patients with established RA are easier to detect using CCP testing than those with early disease. Nevertheless, there was a higher percentage of early RA studies that used a single-gated design (64% single-gated / 36% two-gated) compared to established RA studies (100% two-gated), which could explain the lower sensitivity observed in early RA diagnoses could be influenced by study design, in addition to physiological factors such as auto-antibody titres. It is well known that RA is difficult to diagnose at early onset, and although ACPAs add diagnostic value to the ACR criteria, the results of this study continue to demonstrate that there is still a need for novel biomarkers to further improve the diagnosis of RA. It is important to note that, as well as supporting the diagnosis of RA, CCP tests are also recognised as playing an important prognostic role since they often precede the onset of RA symptoms by years and are associated with significant and rapid radiographic progression (20, 21). Therefore, CCP tests can be used in cohort studies with long-term follow-up to identify patients who are at higher risk of developing RA. Future studies need to evaluate antibody tests specifically in early and difficult-to-diagnose RA patients, utilising appropriate study designs and recruiting both patient and control groups that are relevant to a clinical setting.
Acknowledgements
We would like to thank Barbara Masicciano for her advice throughout this project.

References
40. KIM S, KIM JH, LEE JH, KIM HS: Evaluation of three automated enzyme immunoassays for detection of anti-cyclic citrullinated peptide antibodies in qualitative and quan-


63. MAKSYMOWYCH WP, NAIDES SJ, BYKERK V et al.: Serum 14-3-3 is a novel marker that complements current serological measurements to enhance detection of patients with rheumatoid arthritis. J Rheumatol 2014; 41: 2104-13.

Diagnostic performance of CCP tests in RA / L. Mathsson Alm et al.