ORIGINAL RESEARCH

Diagnostic issues in difficult-to-treat rheumatoid arthritis: a systematic literature review informing the EULAR recommendations for the management of difficult-to-treat rheumatoid arthritis

Nadia M T Roodenrijs, Melinda Kedves, Atila Hamar, György Nagy, Jacob M van Laar, Désirée van der Heijde, Paco M J Welsing

ABSTRACT

Objectives To summarise the evidence on diagnostic issues in difficult-to-treat rheumatoid arthritis (D2T RA) informing the EULAR recommendations for the management of D2T RA.

Methods A systematic literature review (SLR) was performed regarding the optimal confirmation of a diagnosis of rheumatoid arthritis (RA) and of mimicking diseases and the assessment of inflammatory disease activity. PubMed and Embase databases were searched up to December 2019. Relevant papers were selected and appraised.

Results Eighty-two papers were selected for detailed assessment. The identified evidence had several limitations: (1) no studies were found including D2T RA patients specifically, and only the minority of studies included RA patients in whom there was explicit doubt about the diagnosis of RA or presence of inflammatory activity; (2) mostly only correlations were reported, not directly useful to evaluate the accuracy of detecting inflammatory activity in clinical practice; (3) heterogeneous, and often suboptimal reference standards were used and (4) only very few studies had a low risk of bias.

To ascertain a diagnosis of RA or relevant mimicking disease, no diagnostic test with sufficient validity and accuracy was identified. To ascertain inflammatory activity in patients with RA in general and in those with obesity and fibromyalgia, ultrasonography (US) was studied most extensively and was found to be the most promising diagnostic test.

Conclusions This SLR highlights the scarcity of high-quality studies regarding diagnostic issues in D2T RA. No diagnostic tests with sufficient validity and accuracy were found to confirm nor exclude the diagnosis of RA nor its mimicking diseases in D2T RA patients. Despite the lack of high-quality direct evidence, US may have an additional value to assess the presence of inflammatory activity in D2T RA patients, including those with concomitant obesity or fibromyalgia.

Key messages
► Ascertain the diagnosis of rheumatoid arthritis (RA) and the inflammatory origin of the complaints are important in the management of difficult-to-treat (D2T) RA.
► This systematic literature review, conducted to inform the EULAR recommendations for the management of D2T RA, provides an extensive overview of the current literature regarding diagnostic issues in D2T RA.
► The identified evidence had several limitations: (1) study population could not be considered as having D2T RA; (2) typically no appropriate diagnostic association measures were reported; (3) heterogeneous and suboptimal reference standards were used and (4) most studies (thus) had a moderate to high risk of bias.
► No diagnostic tests with sufficient validity and accuracy were found to confirm nor exclude the diagnosis of RA nor its mimicking diseases in D2T RA patients.
► Despite the lack of high-quality direct evidence, ultrasonography may have an additional value to traditional clinical assessment to assess the presence of inflammatory activity in D2T RA patients, including those with concomitant obesity or fibromyalgia.

INTRODUCTION

Treatment options for rheumatoid arthritis (RA) have largely expanded and treatment strategies have improved over the past decades. Nowadays, many patients reach remission or low disease activity when following the current EULAR recommendations and/or American College of Rheumatology (ACR) guideline for the management of RA. However, there is still a substantial proportion of RA patients that remains symptomatic even though they have been treated according to these


Received 10 November 2020
Revised 21 December 2020
Accepted 9 January 2021

Check for updates

© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to Dr Nadia M T Roodenrijs; n.m.t.roodenrijs@umcutrecht.nl

recommendations. This patient group is referred to as having ‘difficult-to-treat (D2T) RA’. This disease state is expected to affect 5%–20% of all patients with RA, depending on the specific definition used.3-5 D2T RA has recently been defined as patients who failed at least two biological/targeted synthetic disease-modifying antirheumatic drugs (b/tsDMARDs) with different mechanisms of action after failing conventional synthetic (cs)DMARD therapy. Additionally, patients should have signs and/or symptoms suggestive of active disease, which is perceived as problematic by the patient and/or rheumatologist.6 The unmet need for these patients was previously underlined by an international survey that was conducted among rheumatologists.7 Consequently, the importance has been acknowledged by EULAR with the approval of a Task Force on the development of management recommendations for D2T RA.

In D2T RA patients, DMARD therapy is frequently changed in routine daily practice in case of signs and/or symptoms suggestive of active disease.4 However, D2T RA is a heterogeneous disease state and various factors could contribute to the persistence of these signs and/or symptoms: factors related to inflammation (eg, having underlying immunological disease mechanisms driving ‘true’ refractory disease or treatment non-adherence), factors of non-inflammatory origin (eg, concomitant fibromyalgia) or both.4-7 All these contributing factors may require different pharmacological and non-pharmacological therapeutic strategies,1 which are reviewed in a separate systematic literature review (SLR).9

Importantly, intensification or other changes in DMARD therapy to reduce inflammation may only be appropriate in patients with insufficient response to therapy due to inflammatory RA activity.4 Symptoms of other diseases, for example, psoriatic arthritis and polyarticular gouty arthritis, may mimic RA possibly leading to misdiagnosis of the disease.4-8 Additionally, coexistence of certain circumstances, for example, obesity, pain syndromes and osteoarthritis, may hamper proper grading of disease activity by influencing diagnostic measures.18 Therefore, in D2T RA, it will be important to ascertain the diagnosis of RA and the presence of inflammatory RA activity before adjusting therapeutic strategies.

The aim of this SLR was first to explore and summarise how to optimally confirm the diagnosis of RA in a D2T RA patient and how to optimally diagnose and rule out alternative or coexisting mimicking diseases. In addition, this SLR focused on the assessment of the presence of inflammatory activity in D2T RA patients and in those with comorbidities that may influence this assessment. This SLR, together with the other SLR focusing on therapeutic strategies in D2T RA,9 was conducted to inform the EULAR recommendations for the management of D2T RA.

METHODS
Research questions
This SLR was conducted following the EULAR standardised operating procedures.11 Three clinical questions on diagnostic issues in D2T RA patients were proposed by the fellow (NMTR), comethodologist (PMJW) and post-doctoral fellow (AH) and then approved by the steering committee (GN (convenor), JMvL (coconvenor), DvdH (methodologist) and MK (fellow)). At the first Task Force meeting, which was held in August 2018, the questions were discussed, amended and then approved by the whole Task Force.

The clinical questions were focused on diagnostic techniques for (1) the confirmation of the diagnosis of RA or a relevant differential diagnoses (either as alternative or coexisting mimicking disease), (2a) the assessment of inflammatory activity in RA patients and (2b) the assessment of inflammatory activity in patients with RA with comorbidities that might influence the assessment of inflammatory activity. Mimicking diseases deemed of interest were gouty arthritis, calcium pyrophosphate deposition disease, psoriatic arthritis, spondyloarthritis, polynyalgia rheumatica, systemic lupus erythematosus, reactive arthritis, paraneoplastic syndromes, osteoarthritis and fibromyalgia. Comorbidities of interest that might influence the assessment of inflammatory activity were infections, malignancies, obesity, pain syndromes (including fibromyalgia), osteoarthritis, subluxations and joint dislocations. The clinical questions were transformed into epidemiological questions using the ‘Patients, Indicator test, Comparison test (ie, reference standard), Outcome format’ (online supplemental file).12

Search strategy
The databases of PubMed and Embase were searched for papers in English until December 2018 for search 1 and December 2019 for search 2. Additionally, the conference abstracts of EULAR and ACR were screened, from 2017 to 2018 for search 1 and from 2017 until 2019 for search 2. Advice regarding the setup of the search strategy was provided by two experienced librarians of Utrecht University (FPW and PHW).

The first search focused on the diagnosis of RA and relevant differential diagnoses. In addition to terms for RA and terms related to diagnostic studies, terms for misdiagnosis and common alternative and coexisting mimicking diseases were included (online supplemental file for search details). During the Task Force meeting, it was agreed to perform a limited search on recent literature on this topic as not much research on (mis)diagnosis relevant to our project was expected to be present and to have a more focused approach given the many clinical questions on D2T RA that were defined by the Task Force. Therefore, a search limit was set to the last ten years and reference screening was not performed.

The second search focused on the assessment of inflammatory activity. In addition to terms for RA and
terms related to diagnostic studies, terms for D2T RA and comorbidities that might influence assessment of inflammatory activity, general terms for inflammation and specific tests to assess inflammatory activity were included (search details in online supplemental file). A search limit was set to the last ten years. In addition, the reference lists of selected papers were manually screened. References published in the year 2000 and later were eligible for inclusion. This cut-off was chosen because of the introduction of bDMARDs around this time and, herewith, the beginning of a new diagnostic and therapeutic landscape regarding acceptable disease activity in the field of RA.

Selection of studies
First, titles and abstracts were screened in duplicate by the fellows (NMTR and MK) according to a set list of selection criteria (online supplemental file) until the percentage of conflicts was below 5%. In case of conflicts or when in doubt, eligibility was discussed with the comethodologist (PMJW). Second, all full text versions of the selected papers were screened in duplicate by the fellows (NMTR and MK). Disagreements were discussed with the comethodologist (PMJW) until consensus was reached.

As specific evidence on D2T RA patients was expected to be scarce, we decided not to focus on D2T RA patients only, but on a broader population of RA patients. Regarding the diagnosis of RA and relevant differential diagnoses, papers were eligible when focusing on patients clinically diagnosed with RA and suspected of a mimicking disease, patients suspected of RA according to the classification criteria in whom a new diagnostic test was evaluated, or patients suspected of RA, but not satisfying classification criteria. Regarding the assessment of inflammatory activity, we decided to only exclude papers when the study population included treatment naïve RA patients. Additionally, in this population, diagnostic tests beyond currently used reference standards had to be evaluated.

Data extraction and quality assessment
Information on study design, patient characteristics, index test, reference standard and diagnostic outcomes were extracted from the included papers using a predetermined format (online supplemental file).

Risk of bias (RoB) and applicability of the included original papers were assessed using the Quality Assessment of Diagnostic Accuracy Studies tool V.2 13 and highest RoB as found among categories was reported here (low, moderate, high). For SLRs, RoB was assessed using ‘A Measurement Tool to Assess systematic Reviews’ V.2 and overall RoB was reported according to its scoring system (low, moderate, high, critically high). 14

One important item included in RoB assessment is the reference standard used. For the diagnosis of RA and relevant differential diagnoses, we deemed a clinical diagnosis according to a rheumatologist as the appropriate reference standard. For the assessment of the presence of inflammatory activity, the preferred reference standard differed between study populations. For the general established RA population, we considered validated Disease Activity Score, Composite Disease Activity Index (eg, DAS28 or CDAI) as appropriate to assess the presence of inflammatory activity at patient level, and the clinical assessment of swelling in the joint at joint level (ie, in a specific joint). In patients in whom there is explicit doubt about the presence of inflammatory activity (including patients with mimicking diseases), the traditional measures are not trustworthy. Therefore, in studies assessing this population we considered (scores based on) established imaging measures as a more appropriate reference standard.

Data extraction and quality assessment were performed in duplicate by the fellows (NMTR and AH) until the number of conflicts was below 5%. Disagreements and remaining doubts were discussed with the comethodologist (PMJW) until consensus was reached.

Statistical analyses
Extracted data were summarised descriptively regarding study and patient characteristics and reported diagnostic association measures. Preferably, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), likelihood ratios (LRs) and ORs were reported. If these were not available, other association measures (typically (Pearson or Spearman) correlation coefficients) were reported, although these measures do not well reflect diagnostic accuracy of measures and thus provide lower quality of evidence. 15 Pooling of results was considered based on clinical and statistical homogeneity.

RESULTS
Study characteristics
The first search regarding the diagnosis of RA and relevant differential diagnoses yielded 2111 unique papers. Title and abstract screening resulted in 337 papers, which were fully reviewed. Four of them fulfilled the selection criteria and were included for data extraction (figure 1A). One additional paper was included via the search on the assessment of inflammatory activity as this paper focused on the diagnosis of RA. Of these five papers, one paper regarded the optimal confirmation of a diagnosis of RA 16 and four papers the confirmation of a coexisting mimicking disease in patients with RA. 17–20

The second search on the assessment of inflammatory disease activity resulted in 3858 unique papers. After title and abstract screening, 237 papers were selected for full-text screening and 45 papers were selected for inclusion. Additionally, 32 papers were selected via reference screening (figure 1B). Seventy of 77 papers were selected for the assessment of inflammatory activity in general 21–96 the seven remaining papers studied the assessment of inflammatory activity in patients with RA with specific comorbidities. 91–97
Heterogeneity in diagnostic tests, diagnostic association measures and reference standards used prohibited pooling the data in an appropriate way. The majority of studies regarding the assessment of inflammatory activity reported correlations only, instead of the preferred diagnostic association measures (ie, sensitivity, specificity, PPV, NPV, LRs or ORs). All quantitative information on diagnostic tests is summarised in online supplemental tables 1–5.

Most studies were found to have a moderate or high RoB. For confirmation of the diagnosis of RA, predominantly because the cut-off for the optimal sensitivity and specificity was selected by using a receiver operating characteristic (ROC) curve analysis of the data of the same patient cohorts (ie, no predefined cut-off). For assessment of inflammatory activity, predominantly because the reference standard used was not optimal. The patient flow and timing of index test vs reference standard were generally described clearly and were appropriate in most studies.

Overall concerns about applicability for the majority of studies were moderate or high, mainly since the patient populations of included studies did not contain (D2T) RA patients in whom there was explicit doubt about the diagnosis. RA, rheumatoid arthritis; SLR, systematic literature review.
diagnosis of RA or the presence of inflammatory activity (RoB assessment and concerns regarding applicability per paper in online supplemental tables 1–5).

**Optimal confirmation of the diagnosis of RA**

One study (low RoB) was found assessing the confirmation of the diagnosis of RA, in patients with a self-reported diagnosis (table 1, online supplemental table 1). Index tests that were assessed in this study were the ACR 1987 classification criteria and adapted versions of these criteria by including synovitis on ultrasonography (US), erosions on US or X-rays and rheumatoid factor (RF) and anticitrullinated protein antibodies (ACPA) positively in various combinations. Additionally, the RA MRI scoring system (RAMRIS) scale was assessed as an alternative for the ACR 1987 classification criteria. The reference standard in this study was the clinical diagnosis made by one rheumatologist after a retrospectively conducted review of all available relevant evidence (except for outcomes of MRI, US and ACPA). The ACR 1987 criteria were found to have a sensitivity of 44% and specificity of 94%. Using the adapted ACR 1987 criteria including Grey scale (GS) synovitis on US, erosions on US and RF, the sensitivity increased to 72% at the expense of a minor decrease in specificity to 91%. Using the RAMRIS scale for the assessment of synovitis in metacarpophalangeal (MCP) joints 2–5, resulted in a sensitivity of 69% and a specificity of 100% (table 1).

**Diagnosis of alternative or coexisting mimicking diseases in patients with RA**

Four papers were found on coexisting mimicking diseases in RA, papers on alternative mimicking diagnoses in patients with RA were not found. Three of four papers reported on fibromyalgia as a coexisting mimicking disease (table 2, online supplemental table 2). In all three papers, the cut-off for the optimal sensitivity and specificity was selected by using an ROC analysis of the data of the same patient cohorts, resulting in a high RoB. The first study assessed the Fibromyalgia Rapid Screening Tool as a diagnostic test for fibromyalgia in consecutive patients suspected of RA or a mimicking disease. Using the clinical diagnosis of fibromyalgia according to a rheumatologist as reference standard, a sensitivity of 83% and specificity of 88% were found. In the second study, a score, derived using the individual components of the DAS28 and rearranging the formula, was used to assess a diagnosis of fibromyalgia. Using the diagnosis of fibromyalgia according to the 2010 criteria, a sensitivity of 81% and specificity of 80% were found. The third study used a case–control design, in which microRNA let-7a, −21–5 p, −143 and −103a-3p were assessed to diagnose concomitant fibromyalgia in established patients with RA. MicroRNA-143 was found to be downregulated in patients with concomitant fibromyalgia. A sensitivity and specificity of 90% and 70% were found.

The only study with a low RoB, was a cross-sectional study reporting on bacterial infections as a mimicking disease in patients with RA who presented with a flare (table 2, online supplemental table 2). The reference standard for the presence of bacterial infections was the agreed diagnosis by physicians based on symptoms, bacterial culture tests, imaging and response to antibiotic therapy. Erythrocyte sedimentation rate (ESR) >15 mm/hour was found to have the highest sensitivity with 98%. Procalcitonin >0.5 ng/mL was found to have the highest specificity with 98%.

**Assessment of inflammatory activity**

Seventy papers evaluated the assessment of inflammatory activity in RA patients at patient and/or joint level. Fifty-eight different diagnostic tests were analysed: 51 biomarkers, 6 imaging measures and one used histology (online supplemental tables 3 and 4). Different reference
<table>
<thead>
<tr>
<th>Paper</th>
<th>Design</th>
<th>Study population</th>
<th>Diagnostic test for mimicking disease</th>
<th>Reference standard for mimicking disease*</th>
<th>Time interval</th>
<th>Sens, % (95% CI)†</th>
<th>Spec, % (95% CI)†</th>
<th>RoB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fan, 2016</td>
<td>CS</td>
<td>Established RA (n=279), with final diagnosis:</td>
<td>Fibromyalgia Rapid Screening Tool ≥5 (/6)</td>
<td>FM according to:</td>
<td>Concurrent</td>
<td>71 (42 to 92)</td>
<td>83 (51 to 95)</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RA with concomitant diagnosis of FM according to ACR criteria (n=52) or to rheumatologist (n=93)</td>
<td></td>
<td>1990 ACR criteria119</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RA without FM (according to ACR criteria: n=227; or to rheumatologist: n=186)</td>
<td></td>
<td>Rheumatologist’s opinion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghib, 2018</td>
<td>CC</td>
<td>Established RA (n=20):</td>
<td>MicroRNA-143 &gt;1.0</td>
<td>RM: NR</td>
<td></td>
<td>90</td>
<td>70</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RA with FM (n=10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RA without FM but with VAS-pain ≥50 (n=10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salaffi, 2018</td>
<td>CS</td>
<td>Established RA (n=292), with final diagnosis:</td>
<td>Derived DAS28 patient-reported components (DAS28-P)* &gt;0.6312</td>
<td>RM according to ACR 2010 criteria119</td>
<td></td>
<td>81 (67 to 92)</td>
<td>80 (75 to 85)</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RA with FM (n=43)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RA without FM (n=249)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sato, 2012</td>
<td>CS</td>
<td>Established RA (n=118), with final diagnosis:</td>
<td>Procalcitonin &gt;0.5 ng/mL</td>
<td>Bacterial Infection:</td>
<td>Within period</td>
<td>26</td>
<td>98</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Flare of RA (n=18)</td>
<td></td>
<td>symptoms, bacterial culture of hospital admission</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-infectious complications (n=23)</td>
<td></td>
<td>tests, imaging studies, and response to antibiotic therapy</td>
<td></td>
<td>34</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-bacterial infection (n=15)</td>
<td></td>
<td></td>
<td></td>
<td>66</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bacterial infection (n=62)</td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

* A clinical diagnosis according to a rheumatologist was deemed as the appropriate reference standard.
†95% CI, if reported.
‡Abstract.

ACR, American College of Rheumatology; CC, case control; CRP, C reactive protein; CS, cross-sectional; DAS28, Disease Activity Score Assessing 28 joints; ESR, erythrocyte sedimentation rate; FM, fibromyalgia; GH, global health; H, high (red); L, low (green); NR, not reported; RA, rheumatoid arthritis; RoB, risk of bias; sens, sensitivity; spec, specificity; TJC, tender joint count; VAS, Visual Analogue Scale; WCC, white cell count.
standards were used for inflammatory activity: composite indices (DAS28, CDAD, Simplified Disease Activity Index), clinical assessment (swollen joint count (SJC (28/32/66)), tender joint count (TJC (28/32/66))) and imaging measures (US, MRI, folate scan). No studies with a low or moderate RoB were found that evaluated a diagnostic test in a population in whom there was explicit doubt about the presence of inflammatory activity and that also reported appropriate diagnostic association measures.

**Patient level**

In papers at patient level, 57 different diagnostic tests were assessed (online supplemental table 3). Seventeen biomarkers and two imaging measures (US sum scores and optical spectral transmission (OST) measures) were assessed in more than one study using the same reference standard used per diagnostic test (table 3). The majority of papers at patient level reported correlation measures only.

Only one study (moderate RoB) explicitly evaluated patients in whom there was doubt about the presence of inflammation, although this study did not report appropriate diagnostic association measures. In patients who had symptoms suggestive of inflammatory joint pain, weak or non-statistically significant correlations were found between DAS28 and US sum scores (US sum scores of hands and feet: r=0.14; US sum scores of MTP joints: r=0.03). In established patients with RA in whom there was not explicit doubt about the presence of inflammation, moderate to strong correlations between US sum scores and composite indices were found in eight other papers (range of r: 0.40–0.70, statistically significant (s) in six of eight papers (two low RoB, five moderate RoB, one high RoB)).

One of these papers was an SLR, in which the authors concluded that US can be a valuable tool to globally assess the extent of synovitis, although it is presently difficult to determine a minimal number of joints to be included in an US sum score.

Only four papers reported an appropriate diagnostic association measure and had a low or moderate RoB, although these papers assessed established patients with RA in whom there was not explicit doubt about the presence of inflammation. All four papers had a low or moderate RoB and assessed a different biomarker using the reference standard deemed appropriate to us (ie, MRI) as a reference standard in the US sum score. Additionally, the SLR about the multi-biomarker disease activity (MBDA) score (including 22 studies, moderate RoB) reported that in three of four papers the MBDA score discriminated between low vs moderate/ high disease activity (MBDA≥30). The appropriate diagnostic association measures were not reported in the SLR and could only be calculated in one of these three papers (DAS28-CRP≥2.7 (at the 6 months visit (ie, non-treatment naïve patients): sensitivity 69%, specificity 64%). Furthermore, moderate statistically significant correlations were reported between the MBDA score and DAS28-CRP (r: 0.41 (pooled r, SLR) and 0.52, both moderate RoB).

**Joint level**

At joint level, 15 different diagnostic tests were assessed (table 4, online supplemental table 4). Four diagnostic tests (clinically swollen joints, OST measures, US and MRI) were assessed in more than one study with the same reference standard used per diagnostic test. In none of the studies, there was explicit doubt about the presence of inflammatory activity.

Almost all studies had a high RoB, predominantly because the cut-off for the optimal sensitivity and specificity was selected by using an ROC curve analysis of the data of the same patient cohort or because the reference standard was not appropriate. The only paper with a moderate RoB was an SLR (including 14 studies), which was performed without critical flaws. In this SLR, synovitis of different joints was assessed with US as diagnostic test and MRI as a reference standard. However, the reference standard used in this SLR (ie, MRI) was regarded as inappropriate to assess the presence of inflammatory activity in the general established RA population, which hampers its applicability.

Using the reference standard deemed appropriate to us (ie, clinical diagnosis of swelling of a joint), three papers were found assessing OST (high RoB). Each study used different diagnostic association measures to report the diagnostic value of OST measures in different joints (sensitivity 37%–59% and specificity 86%–95%; PPV 46% and NPV 86%; area under the ROC 0.88).

**Assessment of inflammatory activity in patients with RA with comorbidities**

Studies assessing diagnostic tests for the assessment of inflammatory activity in patients with RA with a specific comorbidity that may influence the assessment were found for obesity and fibromyalgia (table 5, online supplemental table 5).

Inflammatory activity in patients with RA with and without obesity was assessed in four papers (patient level: two moderate RoB, one high RoB; joint level: one moderate RoB). In the first study at patient level with moderate RoB, an US sum score of 28 joints and a DAS28 in which SJC was based on US assessment were compared with traditional SJC28 and DAS28. In patients with a body mass index (BMI) below 25, no significant differences were found between the US-based and
### Table 3
Concise summary of papers on the assessment of inflammatory activity at patient level

#### Composite indices as reference standard

<table>
<thead>
<tr>
<th>Papers/design</th>
<th>Replicated diagnostic tests</th>
<th>Reference standards (*: n)†</th>
<th>Results</th>
<th>RoB:*</th>
</tr>
</thead>
<tbody>
<tr>
<td>54 papers (4 SLRs (*14/22/7/14), 26–30 34 35 37–40 43–50 52–58 61–65 68–80 82–88)</td>
<td>Biomarkers: MBDA score (2: 3936); miR-146a (1 SLR: 638); ACPA (8: 568); Neutrophil lymphocyte ratio (3: 523); Platelet lymphocyte ratio (2: 421); Leptin (8: 404); IL-6 (4: 373); VEGF (5: 344); MMP-3 (2: 173); IL-17 (2: 121+NR); TNF(a) (2: 185); RF (2: 165); Resistin (2: 141); IL-2 (2: 111); IL-4 (2: 111); IL-10 (2: 111)</td>
<td>DAS28 (51: 8656+NR); CDAI (7: 4186); SDAI (6: 4140); Composite score, not further specified (1: 1307)</td>
<td>Following replicated diagnostic tests with DAS28 as reference standard: MDBA score: pooled r=0.41 (s, 1 SLR); r=0.52 (s, 1 CS)</td>
<td>L: 5</td>
</tr>
<tr>
<td>RA patients: General RA population, n=9686+ NR</td>
<td></td>
<td></td>
<td></td>
<td>H: 6</td>
</tr>
<tr>
<td>► Suspected of or diagnosed with RA, n=40</td>
<td></td>
<td></td>
<td>SLR concludes: ‘The MBDA demonstrates moderate convergent validity with DAS28-CRP and DAS28-ESR, but weaker correlations with SDAI, CDAI, and RAPID3.’</td>
<td></td>
</tr>
<tr>
<td>► Established RA, n=46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>► ACPA positive, n=72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>► Active RA, n=40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>► With musculoskeletal symptoms, n=27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>► With at least one painful or swollen joint, n=42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>► With at least one joint amenable to biopsy, n=15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>► Who were referred for examination of hands and feet, n=46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical assessment as reference standard</td>
<td>Biomarkers: IL-6 (2: 205); VEGF (2: 205); ACPA (3: 181); Leptin (2: 87)</td>
<td>Imaging (sum scores): US (5: 299); OST measure (2: 109)</td>
<td>SJCA28/32/66 (19: 1170); JCA29/32/66 (20: 1207)</td>
<td>Following diagnostic tests with SJCA as reference standard: IL-6: r=0.39–0.41 (s, 2); VEGF: r=0.13–0.14 (ns, 2); ACPA: r=0.051 (ns, 3); Leptin: r=0.048 (ns, 1); US: r=0.3270–0.78 (s, 2); OST measure: r=0.30–0.50 (s, 2)</td>
</tr>
<tr>
<td>20 papers (17 CS, 2 CS, 1 CC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA patients: General population, n=1002</td>
<td>Imaging (sum scores): US (5: 299); OST measure (2: 109)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>► General RA population, n=1002</td>
<td>Biomarkers: IL-6 (2: 205); VEGF (2: 205); ACPA (3: 181); Leptin (2: 87); Resistin (2: 111)</td>
<td>SJCA28/32/66 (19: 1170); JCA29/32/66 (20: 1207)</td>
<td>Following diagnostic tests with SJCA as reference standard: IL-6: r=0.39–0.41 (s, 2); VEGF: r=0.13–0.14 (ns, 2); ACPA: r=0.051 (ns, 3); Leptin: r=0.048 (ns, 1); US: r=0.3270–0.78 (s, 2); OST measure: r=0.30–0.50 (s, 2)</td>
<td></td>
</tr>
<tr>
<td>► ACVA positive, n=72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>► Active RA, n=40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>► With musculoskeletal symptoms, n=27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>► With at least one painful or swollen joint, n=42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>► With knee effusion, n=30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>► With at least one joint amenable to biopsy, n=15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Following replicated imaging measures with composite index as reference standard: US (all types): r=0.03, 0.40–0.70 (s in *7; NR in *2)</td>
<td>US (GS): r=0.251–0.70 (s, *2)</td>
<td>US (PD): r=0.5043–0.72 (s in *5; NR in *1)</td>
<td>DAS28 (51: 8656+NR); CDAI (7: 4186); SDAI (6: 4140); Composite score, not further specified (1: 1307)</td>
<td></td>
</tr>
<tr>
<td>Following diagnostic tests with SJCA as reference standard: IL-6: r=0.39–0.41 (s, 2); VEGF: r=0.13–0.14 (ns, 2); ACPA: r=0.051 (ns, 3); Leptin: r=0.048 (ns, 1); US: r=0.3270–0.78 (s, 2); OST measure: r=0.30–0.50 (s, 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continued...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

traditional measures. In patients with a BMI above 25, the US28 sum score was significantly higher than SJC28 (mean difference in patients with BMI 25–30: 1.818, p=0.001; BMI >30: 1.600, p=0.049). While comparing US-DAS28 with DAS28, US-DAS28 was only statistically significantly higher than DAS28 in patients with a BMI between 25 and 30 (table 5). In the other study at patient level with moderate RoB, lower extremity SJC (only joints below the waist) was found to be increased in patients with a BMI above 30, corrected for patient and physician global disease activity, ESR and TJC (OR 1.633, p=0.005).95 This association was less clear for SJC44 (OR 1.765, p=0.090), suggesting that upper extremity assessment is not significantly influenced by obesity. In the study at joint level (moderate RoB), clinical assessment of a joint being swollen was found to be overestimated in patients with obesity.91 The probability of synovitis according to US decreased per higher BMI category (BMI <25, BMI 25–30, BMI >30), corrected for age, gender and clinical assessment of a joint being swollen (OR BMI 0.52 (95% CI 0.30 to 0.93, p=0.03)).

Three papers evaluated the assessment of inflammatory activity in RA patients with and without fibromyalgia at patient level (two moderate RoB, one high RoB).92-94 The first study with moderate RoB assessed the correlation of composite indices with 7-joint US scores.93 Statistically significant correlations were found with 7-joint US scores based on GS in patients with and without fibromyalgia (range of r: 0.36 to 0.43 and 0.39 to 0.57, respectively). Using 7-joint US scores based on power Doppler (PD), a significant correlation was found in patients without fibromyalgia (range of r: 0.35–0.38), although correlations were found not to be statistically significant in patients with fibromyalgia (range of r: 0.01–0.12). In the other study with moderate RoB, statistically significant correlations were found between SJC and 7-point US scores for synovitis and for tenosynovitis based on GS and PD in patients without fibromyalgia (range of r: 0.44–0.57).95 Again, correlations were not statistically significant in patients with fibromyalgia (r: not given).

DISCUSSION

In this SLR, evidence was sought regarding the optimal confirmation of RA and relevant differential diagnoses as well as the assessment of inflammatory activity in D2T RA patients in whom there was doubt about the diagnosis or the presence of inflammatory activity. Several limitations were found in the selected evidence. First, no studies were identified including D2T RA patients specifically and only the minority of studies included RA patients in whom there was explicit doubt about the diagnosis of RA or about the presence of inflammatory activity. Second, a heterogeneous collection of diagnostic tests was evaluated using different association measures, hampering pooling of results. Third, only very few studies with a low RoB were found. Additional limitations were found in the evidence regarding the assessment of inflammatory activity in D2T RA patients. Mostly, only correlation measures were reported, which are not directly appropriate to assess a test for indicating the presence or absence of inflammatory disease activity in clinical practice (although a strong correlation is likely a prerequisite). Furthermore, major heterogeneity was found in reference standards used in these studies, reflecting the lack of a true gold standard to assess inflammatory activity.
## Table 4 Concise summary of papers on the assessment of inflammatory activity at joint level

<table>
<thead>
<tr>
<th>Papers/design</th>
<th>Diagnostic tests (*: n)</th>
<th>Reference standards (*: n)†</th>
<th>Results</th>
<th>RoB:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical assessment as reference standard</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 papers (3 CS, 1 CC)</td>
<td>Imaging: US (1: 165); OST measure (3: 176)</td>
<td></td>
<td></td>
<td>H: 4</td>
</tr>
<tr>
<td>341 RA patients</td>
<td>Clinical evaluation according to physician (2: 132); Clinically swollen joint (2: 109); Clinically tender joint (1: 50)</td>
<td>OST measures as diagnostic test with following types of clinical assessment as reference standard: ▶ Clinically swollen: Sens 37%–42% (*1); spec 86%–93% (*1) ▶ Clinically swollen and/or tender: PPV: 46%–50% (*1); NPV: 78%–85% (*1) ▶ Clinical evaluation according to physician: r=0.63, p&lt;0.0001 (*1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Imaging as reference standard</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 papers (1 SLR (*14), 9 CS, 2 CC, 1 CS)</td>
<td>Clinical assessment: Clinically tender joints (2: 102) Clinically swollen joints (3: 163)</td>
<td>US (8: 526); FolateScan (1: 40); MRI (4: 295)</td>
<td></td>
<td>M: 1</td>
</tr>
<tr>
<td>RA patients: ▶ All patients, n=788 ▶ With shoulder pain, n=73</td>
<td>Biomarkers: IL-2, IL-4, IL-6, IL-10, IL-17, TNF, IFN, VEGF (1: 64)</td>
<td>Other imaging measure than reference standard: OST measures (5: 222); US (3: 456); MRI (1: 19); Contrast-enhanced MRI (1: 43); Fluorescence optical imaging (1: 18)</td>
<td></td>
<td>H: 13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Histology as reference standard</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 paper (CS)</td>
<td>Imaging: US (GS and PD; 1: 15)</td>
<td>Krenn index of cellular inflammation (1: 15); Krenn lining layer score (1: 15); Inflammatory cell infiltrates (1: 15);</td>
<td></td>
<td>H: 1</td>
</tr>
<tr>
<td>RA patients with at least 1 joint amenable to biopsy, n=15</td>
<td>US (GS) as diagnostic test with following histology measures as reference standard: ▶ Krenn index of cellular inflammation (r=0.65, p&lt;0.01) ▶ Krenn lining layer score (r=0.52, p&lt;0.05)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>US (PD) as diagnostic test with following histology measures as reference standard: ▶ Krenn index of cellular inflammation (r=0.34, s NR) ▶ Krenn lining layer score (r=0.48, s NR)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
Taking all the above-mentioned limitations into account, the identified evidence should be regarded as indirect for the population of D2T RA patients and the results should be interpreted carefully.

Limited evidence was found to consider specific diagnostic tests to confirm or rule out the diagnosis of RA or relevant differential diagnoses. None of the diagnostic tests in the studies regarding the diagnosis of RA or relevant differential diagnoses were replicated, limiting the validity of the results. The only study with a low RoB showed that adapted ACR 1987 criteria (including GS synovitis, US erosions, RF) and RAMRIS scale of MCP joints had an additional value above the traditional ACR 1987 criteria to rule out the diagnosis of RA (sensitivity 69% and 72%, respectively, compared with 42%), although probably still too low to rule out RA with sufficient certainty. Moreover, classification criteria, such as the ACR 1987 criteria, should only be applied after a diagnosis is made and are inappropriate to make a diagnosis, making these results not applicable to ascertain the diagnosis of RA in clinical practice.

Furthermore, in the other study with a low RoB in RA patients who presented with a flare, ESR <15 mm/hour was shown to be able to rule out and procalcitonin ≥0.5 ng/mL to confirm bacterial infection as a mimicking disease. Some studies were found assessing the diagnosis of (concomitant) fibromyalgia, although all these studies had a high RoB. As ‘best available direct evidence’ to assess the presence of inflammatory activity in RA patients in whom there was explicit doubt about the presence of inflammatory activity, only one study was identified, having a moderate RoB. In this study, only weak and statistically non-significant correlations were reported between an US sum score and DAS28. In the general population of RA patients who are not treatment naïve, US was studied most extensively among all diagnostic tests in papers with low to moderate RoB. All papers reported moderate to strong correlations between DAS28 and US sum scores, although also here appropriate diagnostic association measures were not reported. These moderate to strong correlations in the general RA population together with the absence of at least a moderate correlation in patients in whom there is explicit doubt about the presence of inflammatory activity (and thus in whom traditional measures may not be trusted), suggest that US may have an additional value in these patients. However, the optimal number of joints to include in an US sum score to assess inflammatory activity at patient level differed per study and is currently unclear. This limitation hampers the current use of an US sum score in clinical practice.

As the ‘best available indirect evidence’ to assess the presence of inflammatory activity in RA patients in whom there was not explicit doubt about the presence of inflammatory activity, MMP-3 and the MBDA score were studied most extensively in studies reporting the appropriate diagnostic association measures with low to moderate RoB. However, for MMP-3, no validated cut-off was found and, for the MBDA score, the cut-off could not be validated in all studies, limiting the applicability for use in daily practice. At joint level, studies with low to moderate RoB assessing US as well as other diagnostic tests with the preferred reference standard at joint level (ie, clinically swollen joints) were not found.

Presence of obesity and fibromyalgia in patients with RA was found to hamper proper grading of disease activity using traditional composite indices. Presence of fibromyalgia led to overestimation of disease activity compared with US and modified composite indices, while the influence of obesity on the assessment of disease activity was conflicting between studies. Two studies reported an overestimation of disease activity using traditional composite indices compared with US, at least in the joints of the lower extremities. On the contrary, the presence of obesity was found to lead to underestimation of inflammatory activity using SJ C compared with a US-based SJC in another study. In obese patients, composite indices may not only be influenced by the SJC, but also by acute phase reactants. Acute phase reactants may be elevated through the production of inflammatory mediators from adipocytes, resulting in increased composite indices in obese patients.

US was studied most extensively to assess the presence of inflammatory activity in patients with concomitant obesity or fibromyalgia, in studies having a moderate RoB. In these patients, correlations between US and composite indices were weaker or not statistically significant anymore compared with patients without these comorbidities. This suggests that US may have an additional value to traditional measures to assess inflammatory activity.

**Table 4 Continued**

| Papers/design | Diagnostic tests (*: n) | Reference standards (*: n)† | Results | RoB: *
|---------------|-------------------------|----------------------------|---------|---------|

*Number of studies.
†For the general established RA population, the clinical assessment of swelling in the joint was deemed as appropriate to assess the presence of inflammatory activity at joint level (*, in a specific joint). In patients in whom there is explicit doubt about the presence of inflammatory activity, the traditional measures are not trustworthy. Therefore, in studies assessing this population we considered established imaging measures as a more appropriate reference standard.

†Abstract.
CC, case control; CS, cross-sectional; GS, Grey scale; H, high; IFN, interferon; IL, interleukin; L, low; MCP, metacarpophalangeal; NPV, negative predictive value; NR, not reported; OST, optical spectral transmission; PD, power Doppler; PIF, proximal interphalangeal; PPV, positive predictive value; RA, rheumatoid arthritis; RoB, risk of bias; s, significant; sens, sensitivity; SLR, systematic literature review; spec, specificity; TNF, tumour necrosis factor; US, ultrasonography; VEGF, vascular endothelial growth factor.
RA activity in patients with these comorbidities. No studies were found regarding other comorbidities that might influence assessment of inflammatory disease activity.

A previous EULAR project has focused on the development of the EULAR recommendations for the use of imaging of the joints in RA and there is some overlap with our SLR (recommendation 3: Ultrasound and MRI are superior to clinical examination in the detection of joint inflammation; these techniques should be considered for a more accurate assessment of inflammation). The statement regarding US is consistent with the findings of our SLR. However, the results of our SLR do not clearly indicate the usefulness of MRI. Most studies on MRI were not included in our SLR, predominantly because they were focused on treatment naïve RA patients, were
published before the year 2000 or assessed the change in inflammatory activity instead of the presence of inflammatory activity as relevant for our present work.\textsuperscript{108–117}

In addition to the limitations in the evidence that was found, this SLR has some limitations itself. Although an extensive literature search has been performed, relevant papers might have been missed. Regarding the diagnosis of RA and relevant differential diagnoses, it was chosen to perform a limited search focusing on the last ten years and not to perform reference screening because not much relevant evidence was presently expected before this time and to enable focusing more on the other clinical questions regarding D2T RA where more relevant literature was expected. After the second Task Force meeting was postponed due to the COVID-19 outbreak, it was decided not to update the search for this specific question because of the same above-mentioned reasons. Regarding the assessment of inflammatory activity, the search focused on the last 10 years, although references of selected papers were also screened and relevant papers published from the year 2000 were selected because of the introduction of bDMARDs around this time point. Additionally, for this search, we focused on non-treatment naïve RA patients resulting in the exclusion of papers focusing on RA patients in the early phase of the disease. However, we felt this was well-justifiable as D2T RA patients are by definition established RA patients and evidence on early RA was deemed too indirect for our present work. Although, as above decisions could be considered limitations, it should be stressed that choices were made by the Task Force, including experienced clinicians, researchers and methodologists and with input from experienced librarians. Therefore, we think the methodological stringency of this SLR and its focus on established RA patients in the present diagnostic and therapeutic era, have resulted in a comprehensive overview of the current literature.

Further guidance on the diagnostic issues in D2T RA, including the clinical implications of the results, will be provided by the EULAR Task Force on D2T RA in their recommendations for the management of D2T RA, which will be published soon.\textsuperscript{118} Additionally, a research agenda will be provided including topics that should be addressed in future studies.

In conclusion, this SLR highlights the scarcity of evidence on the optimal confirmation or ruling out of a diagnosis of RA and relevant differential diagnoses in D2T RA patients. Therefore, textbook knowledge on potential alternative and/or coexisting mimicking diseases remains highly relevant. When currently used clinical measures may not be trusted as in D2T RA patients, US may have some additional value to assess the presence of inflammatory activity in these patients as well as in those with concomitant obesity or fibromyalgia. However, more high-quality studies addressing D2T RA patients in whom there is reasonable doubt about the diagnosis and about the presence of inflammatory activity are required.

\textbf{REFERENCES}


influence the severity and activity in rheumatoid arthritis patients? 


