Original Article

Reliability of Magnetic Resonance Imaging (MRI)-scoring of the Metatarsophalangeal-joints of the Foot According to the Rheumatoid Arthritis-MRI Score (RAMRIS)

Y. J. Dakkak (ORCID ID 0000-0002-6140-0310), X.M.E. Matthijssen (0000-0001-7332-8072), D.M. van der Heijde (0000-0002-5781-158X), M. Reijnierse (0000-0002-6906-1851), A.H.M. van der Helm-van Mil (0000-0001-8572-1437)

From the Department of Rheumatology and the Department of Radiology, Leiden University Medical Centre, Leiden, the Netherlands.

Key Indexing Terms: rheumatoid arthritis, MRI, RAMRIS, reliability, foot

Y.J. Dakkak, MD, Department of Rheumatology, Leiden University Medical Centre; X.M.E. Matthijssen, MD, Department of Rheumatology, Leiden University Medical Centre; D. van der Heijde, MD, PhD, Department of Rheumatology, Leiden University Medical Centre; M. Reijnierse, MD, PhD, Department of Radiology, Leiden University Medical Centre; A.H.M. van der Helm-van Mil, MD, PhD, Department of Rheumatology, Leiden University Medical Centre and Department of Rheumatology, Erasmus Medical Centre, Rotterdam.

Address correspondence to Dr. Y.J. Dakkak, Leiden University Medical Centre, Department of Rheumatology, C1-R, P.O. Box 9600, 2300 RC Leiden, the Netherlands. E-mail: y.j.dakkak@lumc.nl

Key messages:

- Applying the RAMRIS to the metatarsophalangeal-joints is reliable for status and change scores

- The reliability of the RAMRIS applied to the metatarsophalangeal-joints is similar to that of the metacarpophalangeal- and wrists-joints

- The results are encouraging for the use of MRI of the metatarsophalangeal-joints in trials in the early phases of RA.

ABSTRACT

Objective: The Rheumatoid Arthritis Magnetic Resonance Imaging Score (RAMRIS) is validated for hand-MRI. Its reliability applied to metatarsophalangeal (MTP-1-5)-joints is unknown and was studied in early arthritis and clinically suspect arthralgia.

Methods: Patients underwent 1.5T MRI of MTP-, metacarpophalangeal (MCP-2-5)- and wrist-joints. Two paired readers scored bone marrow edema (BME), synovitis, tenosynovitis and erosions. Interreader reliability was assessed of 441 consecutive early arthritis patients at baseline, 215 by two readers, the remaining 226 by two different readers. Two readers scored baseline MRIs of 82 consecutive patients with clinically suspect arthralgia, and 40 randomly selected patients by nine readers. Intrareader reliability was determined on a random set of 15 early arthritis patients, scored twice by two readers. For change-scores, 30 early arthritis patients with baseline and 1-year follow-up MRI were scored by two readers. Intraclass correlation coefficients (ICCs), Bland-Altman (BA)-plots and smallest detectable change (SDC) were determined. MRI-data of MTP-joints were compared to wrist- and MCP-joints

Results: Interreader ICCs and mean scores in early arthritis were: BME ICC 0.91-0.92 (mean 1.5 (standard deviation (SD) 2.6)), synovitis 0.90-0.92 (1.3(1.7)), tenosynovitis 0.80-0.85 (1.1(1.8)) and erosions 0.88-0.89 (0.7(1.0)). In clinically suspect arthralgia-patients ICCs were comparable. Intrareader ICCs for inflammatory MRI-features were 0.84-0.98, for erosions 0.71 (reader 1) and 0.92 (reader 2). Change-score ICCs were ≥0.90, except erosions (0.77). SDCs were ≤1.0. BA-plots showed no systematic bias. Reliability-scores of MTP-joints were similar to MCP- and wrist-joints.

Conclusion: Status and change MRI-scores of BME, synovitis, tenosynovitis and erosions of MTP-joints can be assessed reliably by RAMRIS.

INTRODUCTION

Magnetic resonance imaging (MRI) is increasingly used in scientific research in patients with rheumatoid arthritis (RA), as it is a sensitive modality that can visualise inflammation and destruction.(1) As the complexity and large amount of information that is provided by MRI pose a challenge, the Outcomes in Rheumatology Clinical Trials (OMERACT) MRI in RA working group developed the RA MRI Score (RAMRIS) to standardise MRI-scoring for research purposes and clinical trials in particular.(2)

The RAMRIS has thus far been validated for use in the metacarpophalangeal(MCP)- and wrist-joints, but not for use in metatarsophalangeal(MTP)-joints.(3-5) This is unfortunate as joint inflammation in MTP-joints is just as prevalent as in the MCP-joints.(6, 7) In addition, radiographic studies have shown that erosive change occurs more commonly in the feet than in the hands, and also in earlier phases of disease.(8, 9) Thus there is a paradox between the notion that the feet are so commonly affected in early RA and their absence as an outcome measure in trials. Indeed, the RA MRI working group has recently called for validation of the RAMRIS in the MTP-joints.(10)

An important aspect of validation is the reliability of scoring.(11) Reliability studies have been performed for the hand, but cannot be directly extrapolated to the foot, as different joint areas in the past have been found to have different intraclass correlation coefficients (ICCs).(12) Previously, Baan *et al*. measured the reliability of the RAMRIS of the feet in a small subset of patients with longstanding RA (n=29).(13) However, tenosynovitis which is a common feature in early arthritis, was not included in this study. In addition, as no follow-up MRIs were included only the reliability of status scores was assessed. For change scores, one study has been performed by Ejbjerg *et al*. that assessed MRI-detected erosions only.(14) We therefore aimed to assess the inter- and intra-reliability of status scores and the reliability of change scores applied to the MTP-joints for the following MRI-outcomes: bone marrow edema (BME), synovitis, tenosynovitis and erosions. As the

4

focus in rheumatology is shifting from established erosive RA to early arthritis and even to patients with arthralgia that are suspect to progress to arthritis(1), we performed our study in patients with early arthritis and also in patients with clinically suspect arthralgia without apparent arthritis upon physical examination. We added MRI-data of wrist and MCP-joints as comparison to data of MTP-joints.

MATERIALS AND METHODS

Patients

Early Arthritis Cohort

This longitudinal inception cohort included patients with clinically confirmed arthritis and symptom duration <2 years who were naïve to disease-modifying antirheumatic drugs (DMARDs). At baseline, questionnaires were completed, swollen joint counts were performed and serum samples were obtained. Unilateral 1.5 tesla(T) MRI of the MTP-, MCP- and wrist-joints of the most painful side, or the dominant side in the case of equally severe symptoms on both sides was made of patients who were consecutively included from June 2013 onwards.(15) Before contrast administration, T1-weighted fast spin-echo (FSE) sequences in the coronal plane were acquired for MCP- and wrist joints. After intravenous injection of gadolinium contrast, T1-weighted FSE sequences with frequency selective fat saturation (fatsat) were acquired in coronal and axial planes of the MCP-, wrist- and MTP-joints. Patients were asked to stop nonsteroidal anti-inflammatory drugs (NSAIDs) 24 hours before the scan, and the MRI was made before the start of DMARDs. Additional information on the scan-protocol is provided supplementary.

5

Consecutive patients included between June 2013 and April 2016 were studied for status scores. In the cohort, serial MRIs were made of patients included up until January 2015.

Clinically Suspect Arthralgia

This inception cohort included patients with clinically suspect arthralgia of the small joints with a symptom duration of <1 year that, according to the clinical expertise of the rheumatologist, was suspected to progress to RA over time. Per definition, clinically suspect arthralgia was not present if clinical arthritis was observed at physical examination or if another explanation for the arthralgia was more likely.(16) Patients consecutively included between July 2014 and February 2015 were studied, and underwent MRI according to the same MRI-protocol as patients with early arthritis.

Both cohorts were approved by the local Medical Ethical Committee (*approval numbers Early Arthritis Cohort P10.108 and Clinically Suspect Arthralgia P11.210*). All participants signed informed consent.

Readers

All readers were experienced with the OMERACT RAMRIS system and the method by Haavardsholm *et al*. for scoring tenosynovitis.(2, 17) All readers have scored >400 MRIs according to these systems during a training period of several months prior to evaluating the MRIs that are part of this study.

MRI Scoring

All readers evaluated the images independently and in the following order: first the MTP-joints, next the MCP-joints and finally the wrist. The MRI-images were scored blinded to clinical data. BME, synovitis and erosions of MTP-, MCP- and wrist-joints were scored in line with the OMERACT RAMRIS, with the exception that BME was assessed on a contrast-enhanced T1-weighted fat-suppressed sequence,(18) as its use for depicting BME is recommended by the European Society of

Musculoskeletal Radiology (ESSR) and previous studies have demonstrated that it has a strong correlation with the T2-weighted fat suppressed sequence that is advised by the RAMRIS.(19-22) In MTP- and MCP-joints erosions and BME were scored in the proximal and distal part of the joints. Tenosynovitis was scored as described by Haavardsholm *et al*. applied to the flexor and extensor tendons of MTP-1-5, MCP-2-5 and the wrist.(17) Additional information on the method of scoring is provided supplementary, in addition to an example of a score sheet with illustration of the scored tendons (Supplementary Figure 1).

A flowchart of scored patients and readers is presented in Figure 1. Intrareader reliability was assessed based on 441 consecutive early arthritis patients. The first 215 patients were scored by reader 1 and 2, the remaining 226 MRIs by reader 3 and 4. The MRIs of 82 arthralgia patients were scored by reader 5 and 6. Of these 82 MRIs, 40 were randomly selected and scored by seven additional readers, resulting in a total of 9 readers (readers 2-10). For intrareader reliability, the baseline MRIs of 15 early arthritis patients were randomly selected and rescored by reader 1 and 2 after an interval of respectively 6 and 4 months.

91 early arthritis patients underwent MRI at baseline and at 12-months, and were all scored by reader 9. The reliability of change scores was determined using MRIs of 30 patients, which were in addition also scored by reader 10. These 30 patients were selected as follows: 15 randomly, and 15 based on a high baseline MRI-score by reader 9. The 15 patients with a high baseline score were scored as part of a bigger set of patients that also included patients with lower scores; thus the images were scored by readers that were blinded for the MRI-score. We added patients with high baseline scores as they were most prone to change over time. The MRIs were scored in chronological order by both readers.(23, 24)

Statistical methods

For scores of MTP-, MCP- and wrist-joints separately, ICC-estimates and their 95%-confident intervals (95% CI) were calculated (2-way mixed-effects model, absolute-agreement).(25) For the intra-reader ICC the single measures were used, for interreader and change ICCs the average measures. ICC values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good and excellent reliability, respectively.(25) In addition to calculating ICC-values, Bland-Altman (BA) plots were drawn.(26)

For change scores, in addition to the ICC and Bland-Altman plots, the smallest detectable change (SDC) was calculated. SDC expresses the smallest change between two dependently obtained measures that can be interpreted as 'real', that is a change greater than the measurement error.(27) The SDC of each MRI-feature was calculated as follows:

$$SDC = \frac{1.96 \times SDdiff}{(\sqrt{k} \times \sqrt{2})}$$

Here k = 2 as the SDC on the mean scores of both readers was used.(27)

The proportion of patients that showed change in the RAMRIS score was calculated in three ways: using a cut-off of >0 and of >0.5 (of the mean score of 2 readers), and by using the SDC as a cut-off.

Sub analyses for the interreader reliability were performed within the subgroup of patients with the following diagnoses: RA, unclassified arthritis (UA), psoriatic arthritis or spondyloarthritis (SpA) and inflammatory osteoarthritis (OA).

Data were analysed using SPSS version 23.

RESULTS

Patient characteristics

8

Characteristics of patients with early arthritis and clinically suspect arthralgia are shown in Supplementary Table 1. In both cohorts, patients were predominantly female (61% and 84% respectively) and had a mean age of 55 and 46 years respectively. Characteristics of the 30 patients with follow-up MRIs are also presented in Supplementary Table 1, they had a higher swollen joint count than the overall early arthritis group (6 versus 3). Of these 30 patients, 29 were prescribed DMARDs after the baseline visit during the first year of follow-up, one received NSAIDs only.

Interreader reliability

The interreader ICCs and median and mean MRI-scores for patients with early arthritis and with arthralgia are presented in Table 1. The scores of the individual readers are depicted in Supplementary Table 2. For the MTP-joints in patients with early arthritis the mean scores varied from 0.6 (standard deviation (SD) 0.9) for erosions to 1.5 (SD 2.6) for BME. The corresponding ICCs for BME ranged from 0.91-0.92 (95%CI 0.90-0.93 and 0.90-0.94), for synovitis from 0.90-0.92 (95%CI 0.84-0.94 and 0.88-0.94), for tenosynovitis from 0.80-0.85 (95%CI 0.69-0.86 and 0.78-0.90) and for erosions from 0.88-0.89 (95%CI 0.84-0.91 and 0.86-0.92). In arthralgia patients, the mean and median scores of MRI-features were lower, but ICCs were similar and all >0.87, except for BME that had an ICC of 0.77 (95%CI 0.64-0.85) when read by two readers, and an ICC of 0.95 (95%CI of 0.93-0.97) when there were 9 readers. The BA-plots indicated that systematic bias was low, as in Figure 2 the middle line, depicting the mean, was located around 0. Only for tenosynovitis there was a tendency towards more random variation with higher scores (heteroscedasticity).

The interreader reliability of the MRI-features for MCP- and wrists-joints were similar to the MTP-joints (Table 1).

As a sensitivity analyses we looked at the reliability in the separate diagnoses RA (n=157), UA (n=148), SpA (n=45) and inflammatory OA (n=23). The results of the sensitivity analyses of the

separate diagnoses were similar to the results of the patients combined as presented above (Supplementary Table 4 and 5).

Intrareader reliability

The intrareader ICCs, mean and median scores are presented in Table 2. Mean scores of MRI-features in the MTP-joints varied from 0.4 (SD 0.6 SD) for erosions, to 1.7 (SD 2.9) for tenosynovitis. The ICC scores for BME ranged from 0.96-0.98 (95%CI 0.89-0.99 and 0.95-0.99), for synovitis from 0.90-0.98 (95%CI 0.74-0.97 and 0.94-0.99), for tenosynovitis from 0.84-0.97 (95%CI 0.58-0.94 and 0.91-0.99) and for erosions from 0.71-0.92 (95%CI 0.35-0.89 and 0.78-0.97). BA-plots indicated that systematic bias was low and are presented in Supplementary Figure 2.

The intrareader reliability of the MRI-features for MCP- and wrists-joints were similar to the MTP-joints (Table 2).

Reliability of change scores

The mean, median and ICCs of change scores after 1 year of follow-up of 30 early arthritis patients are presented in Table 3. The scores of the individual readers are depicted in Supplementary Table 3. The change in MRI-scores over time in the MTP-joints was small for erosions (mean 0.4 (0.6 SD)) and larger for the inflammatory MRI-features (≥-1.3). The ICC for change scores were ≥0.90 for BME, synovitis and tenosynovitis, and 0.77 (95%CI 0.52-0.89) for erosions. The SDC was ≤1 for all MRI-features, suggesting a high potential to detect changes. The number of patients with true change by using the SDC as a cut-off was similar to the number of patients where change >0.5 was measured; then BME revealed change in 37% of patients, synovitis in 67%, tenosynovitis in 47% and erosions in 17% (Table 3). BA-plots indicated that systematic bias was low and are presented in Figure 3.

The same analyses were performed for the MCP- and wrist-joints, and are presented in Table 3, these results were similar to the MTP-joints.

DISCUSSION

In RA-research, the scoring of MR-images is performed according to the RAMRIS, of which validation as an outcome measure for trials has thus far focussed on the hands.(2, 3) In this study we investigated the reliability of the RAMRIS when applied to the MTP-joints. Overall, we observed good to excellent intra- and interreader reliability for status and change scores. In particular, ICCs for inflammatory features were generally >0.90.

Previously, the reliability of status scores of BME, synovitis and erosions as well as change score of erosions in MTPs have been published and were found to be excellent.(13, 14) Our study is the first to look at the reliability of scoring tenosynovitis at the MTP-joints and change scores of inflammatory MRI-features as measured by BME, synovitis and tenosynovitis in an early arthritis setting. To further support our findings, we also analysed data from the wrist- and MCP-joints to compare this to data of MTP-joints. The current data showed that scoring of MTP-joints was equally reliable. Finally, our findings obtained on hand joints are in concordance with previous MRI-studies of the hands which supports the validity of the present results.(28-30)

A pitfall of ICCs is that they are sensitive to a lack of variability among sampled subjects.(25) We found the intrareader reliability of erosions to be moderate for reader 1 (0.71, 95%CI 0.35-0.89), but excellent for reader 2 (0.92, 95% CI 0.78-0.97). The mean score of erosions was low (0.4 (SD 0.6)), which corresponds to a lack of variability among these subjects that could have resulted in a moderate ICC. In addition, for change scores the reliability of erosions was lower than the inflammatory MRI-features (0.77, 95% CI 0.52-0.89), but still good.(25) Also here the mean change in the score of MRI-features was lowest for erosions (0.4 (SD 0.6)) compared to ≥-1.3 for the other

features, see Table 3). Besides ICCs, BA-plots, and for change scores SDCs, are important to take into consideration as measures of reliability. BA-plots visualise the data and illustrate that levels of agreement were acceptable in both cases (Figure 2 and 3), and for change scores the low SDC suggests a good reliability.(31)

For change scores we selected early arthritis patients that had high baseline scores and were thus most prone to change over time, specifically a decrease in inflammation and possibly an increase in erosions. The mean change scores were low for all MRI-features, and for erosions in particular (at the MTPs the mean change score was 0.4 (SD 0.6)). This is expected, as 29 of the 30 patients received DMARDs, inhibiting the occurrence or progression of erosions.

The focus in rheumatology is shifting from established erosive RA to early arthritis and even to patients with clinically suspect arthralgia. Therefore, we included patients with clinically suspect arthralgia and found the scoring of status scores to be reliable. In different stages of disease, MRI-detected lesions may be more or less frequent, which may influence the reliability of scoring.(32) MRI-detected inflammation was subclinical by definition, as there was no apparent arthritis upon physical examination. As expected, absolute MRI scores for arthralgia patients were lower than those with early arthritis, but the reliability overall was good. This is encouraging for MRI-studies in the pre-arthritis phase.

Radiographic studies have shown that erosive lesions occur more commonly in the feet than in the hands, and also in earlier phases of disease.(8, 9) In our results, the scores of the MRI-features were higher in the hands than in the MTPs, especially in the wrist. This is in concordance with a recent study performed in patients with undifferentiated arthritis on the development of RA, were adding MRI of the foot did not improve predictive accuracy compared to MRI of the hand alone.(33) This was explained by the finding that inflammation in the foot was indeed an early phenomenon, but it almost never occurred without inflammation in the hands.

A strength of this study is that it included a large number of patients from two different cohorts, and scoring was performed by numerous readers with considerable experience with the RAMRIS. We studied an unselected group of patients with early arthritis, rather than a specific group of patients that met stringent inclusion criteria of trials. Reliability studies in this patient population are infrequent, the present data are important for future studies in early arthritis. As a sensitivity analyses we looked at the reliability of scoring in the following diagnoses separately: RA, Ua, SpA and inflammatory OA. This was done for reader 1/2 and reader 3/4 separately, and thus resulted in small numbers of patients, especially in the SpA and inflammatory OA groups (SpA n=15 and 30, and OA n=12 and 11). For the RA and UA groups the reliability of scoring was overall good, for the latter two diagnoses caution should be given when interpreting the results.

The aspect discrimination of the OMERACT Filter was not addressed in our study, and should be subject for further research.(11) In addition, whether the measured change is clinically relevant needs to be determined in studies evaluating the minimal clinical important difference (MCID). The scorings were not timed and thus unfortunately it was not possible to make a statement concerning feasibility.

We applied the RAMRIS that is developed for the wrist and MCP-joints to the MTP-joints, and for tenosynovitis the commonly used score developed by Haavardsholm *et al*.(2, 17) The RAMRIS was recently updated, and now includes joint space narrowing and a slightly modified tenosynovitis-score published by Glinatsi *et al*.(10, 34) The RAMRIS was not yet updated at the start of this study and was therefore not used here. We applied the tenosynovitis score of Haavardsholm *et al.* to the flexor and extensor tendons of the MCP- and MTP-joints. Although the extensor tendons at the MTP- and MCP-joints seem to lack a synovial sheath, inflammation around the extensor tendons at the MCP-joints have been described in RA.(35) And thus, even though the nature of this inflammation is unclear, we believe it is important to further study and validate the scoring of the inflammation observed around the extensor tendons, which includes assessing its' reliability.

13

According to the RAMRIS-method, T2-weighted fat suppressed or short tau inversion recovery (STIR) sequences should be used to assess bone marrow edema (BME). Previous studies have demonstrated that a contrast enhanced T1-weigthed fat suppressed sequence has a strong correlation with T2-weighted fat suppressed sequences.(19-21) In addition, the European Society of Musculoskeletal Radiology (ESSR) Arthritis Subcommittee also recommends the use of contrast enhanced T1-weighted fat suppressed sequences for depiction of BME.(22) We therefore used the contrast enhanced T1-weighted fat suppressed sequence as it allowed a shorter scan time and has a higher signal to noise ratio. We thus believe that this did not influence the reliability of scoring BME, although we did not strictly follow the RAMRIS-protocol for depicting BME.

The MTP-images were acquired after gadolinium contrast was given for the acquisition of hand images. The time window between contrast administration and imaging of the foot was approximately 12 minutes. Previously it was shown that small time variations are not of major importance to measured synovial membrane volumes, as during a one-hour post-contrast follow up period the measured volumes remained almost unchanged.(36) We therefore believe this has not influenced our results.

In conclusion, scoring of status and change scores of BME, synovitis, tenosynovitis and erosions of the MTP-joints according to the RAMRIS was reliable. This is encouraging for the use of the scoring system also for MTP-joints in trials in early phases of RA.

REFERENCES

1.   Colebatch AN, Edwards CJ, Ostergaard M, van der Heijde D, Balint PV, D'Agostino MA, et al. Eular recommendations for the use of imaging of the joints in the clinical management of rheumatoid arthritis. Ann Rheum Dis 2013;72:804-14.

2.   Ostergaard M, Peterfy C, Conaghan P, McQueen F, Bird P, Ejbjerg B, et al. Omeract rheumatoid arthritis magnetic resonance imaging studies. Core set of mri acquisitions, joint pathology definitions, and the omeract ra-mri scoring system. J Rheumatol 2003;30:1385-6.

3.   Woodworth TG, Morgacheva O, Pimienta OL, Troum OM, Ranganath VK, Furst DE. Examining the validity of the rheumatoid arthritis magnetic resonance imaging score according to the omeract filter-a systematic literature review. Rheumatology (Oxford) 2017;56:1177-88.

4.   Haavardsholm EA, Ostergaard M, Ejbjerg BJ, Kvan NP, Uhlig TA, Lilleas FG, et al. Reliability and sensitivity to change of the omeract rheumatoid arthritis magnetic resonance imaging score in a multireader, longitudinal setting. Arthritis Rheum 2005;52:3860-7.

5.   Dakkak YJ, van der Heijde DM, Reijnierse M, van der Helm-van Mil AHM. Validity of the rheumatoid arthritis mri score applied to the forefeet using the omeract filter: A systematic literature review. RMD Open 2018;4:e000796.

6.   Boutry N, Larde A, Lapegue F, Solau-Gervais E, Flipo RM, Cotten A. Magnetic resonance imaging appearance of the hands and feet in patients with early rheumatoid arthritis. J Rheumatol 2003;30:671-9.

7.   Calisir C, Murat Aynaci AI, Korkmaz C. The accuracy of magnetic resonance imaging of the hands and feet in the diagnosis of early rheumatoid arthritis. Joint Bone Spine 2007;74:362-7.

8.   Landewe RB, Strand V, Conaghan PG, van der Heijde D. Damage and progression on radiographs in individual joints: Data from pivotal randomized controlled trials. J Rheumatol 2011;38:2018-22.

9.   van der Heijde DM, van Leeuwen MA, van Riel PL, Koster AM, van 't Hof MA, van Rijswijk MH, et al. Biannual radiographic assessments of hands and feet in a three-year prospective followup of patients with early rheumatoid arthritis. Arthritis Rheum 1992;35:26-34.

10.   Ostergaard M, Peterfy CG, Bird P, Gandjbakhch F, Glinatsi D, Eshed I, et al. The omeract rheumatoid arthritis magnetic resonance imaging (mri) scoring system: Updated recommendations by the omeract mri in arthritis working group. J Rheumatol 2017.

11.   Wells G, Beaton DE, Tugwell P, Boers M, Kirwan JR, Bingham CO, 3rd, et al. Updating the omeract filter: Discrimination and feasibility. J Rheumatol 2014;41:1005-10.

12.   Glinatsi D, Bird P, Gandjbakhch F, Mease PJ, Boyesen P, Peterfy CG, et al. Validation of the omeract psoriatic arthritis magnetic resonance imaging score (psamris) for the hand and foot in a randomized placebo-controlled trial. J Rheumatol 2015;42:2473-9.

13.   Baan H, Bezooijen R, Avenarius JK, Dubbeldam R, Drossaers-Bakker WK, van de Laar MA. Magnetic resonance imaging of the rheumatic foot according to the ramris system is reliable. J Rheumatol 2011;38:1003-8.

14.   Ejbjerg BJ, Vestergaard A, Jacobsen S, Thomsen HS, Ostergaard M. The smallest detectable difference and sensitivity to change of magnetic resonance imaging and radiographic scoring of structural joint damage in rheumatoid arthritis finger, wrist, and toe joints: A comparison of the omeract rheumatoid arthritis magnetic resonance imaging score applied to different joint combinations and the sharp/van der heijde radiographic score. Arthritis Rheum 2005;52:2300-6.

15.   Nieuwenhuis WP, van Steenbergen HW, Mangnus L, Newsum EC, Bloem JL, Huizinga TWJ, et al. Evaluation of the diagnostic accuracy of hand and foot mri for early rheumatoid arthritis. Rheumatology (Oxford) 2017;56:1367-77.

16.   van Steenbergen HW, van Nies JA, Huizinga TW, Bloem JL, Reijnierse M, van der Helm-van Mil AH. Characterising arthralgia in the preclinical phase of rheumatoid arthritis using mri. Ann Rheum Dis 2015;74:1225-32.

17.   Haavardsholm EA, Ostergaard M, Ejbjerg BJ, Kvan NP, Kvien TK. Introduction of a novel magnetic resonance imaging tenosynovitis score for rheumatoid arthritis: Reliability in a multireader longitudinal study. Ann Rheum Dis 2007;66:1216-20.

18.   Ostergaard M, Edmonds J, McQueen F, Peterfy C, Lassere M, Ejbjerg B, et al. An introduction to the eular-omeract rheumatoid arthritis mri reference image atlas. Ann Rheum Dis 2005;64 Suppl 1:i3-7.

19.   Mayerhoefer ME, Breitenseher MJ, Kramer J, Aigner N, Norden C, Hofmann S. Stir vs. T1-weighted fat-suppressed gadolinium-enhanced mri of bone marrow edema of the knee: Computer-assisted quantitative comparison and influence of injected contrast media volume and acquisition parameters. J Magn Reson Imaging 2005;22:788-93.

20.   Schmid MR, Hodler J, Vienne P, Binkert CA, Zanetti M. Bone marrow abnormalities of foot and ankle: Stir versus t1-weighted contrast-enhanced fat-suppressed spin-echo mr imaging. Radiology 2002;224:463-9.

21.   Stomp W, Krabben A, van der Heijde D, Huizinga TW, Bloem JL, van der Helm-van Mil AH, et al. Aiming for a shorter rheumatoid arthritis mri protocol: Can contrast-enhanced mri replace t2 for the detection of bone marrow oedema? Eur Radiol 2014;24:2614-22.

22.   Sudol-Szopinska I, Jurik AG, Eshed I, Lennart J, Grainger A, Ostergaard M, et al. Recommendations of the essr arthritis subcommittee for the use of magnetic resonance imaging in musculoskeletal rheumatic diseases. Semin Musculoskelet Radiol 2015;19:396-411.

23.   van Der Heijde D, Boonen A, Boers M, Kostense P, van Der Linden S. Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. Rheumatology (Oxford) 1999;38:1213-20.

24.   van Tuyl LH, van der Heijde D, Knol DL, Boers M. Chronological reading of radiographs in rheumatoid arthritis increases efficiency and does not lead to bias. Ann Rheum Dis 2014;73:391-5.

25.   Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155-63.

26.   Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.

27.   Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D. Deciding on progression of joint damage in paired films of individual patients: Smallest detectable difference or change. Ann Rheum Dis 2005;64:179-82.

28.   Durez P, Malghem J, Nzeusseu Toukap A, Depresseux G, Lauwerys BR, Westhovens R, et al. Treatment of early rheumatoid arthritis: A randomized magnetic resonance imaging study comparing the effects of methotrexate alone, methotrexate in combination with infliximab, and methotrexate in combination with intravenous pulse methylprednisolone. Arthritis Rheum 2007;56:3919-27.

29.   Genovese MC, Kavanaugh A, Weinblatt ME, Peterfy C, DiCarlo J, White ML, et al. An oral syk kinase inhibitor in the treatment of rheumatoid arthritis: A three-month randomized, placebo-controlled, phase ii study in patients with active rheumatoid arthritis that did not respond to biologic agents. Arthritis Rheum 2011;63:337-45.

30.   Ostergaard M, Emery P, Conaghan PG, Fleischmann R, Hsia EC, Xu W, et al. Significant improvement in synovitis, osteitis, and bone erosion following golimumab and methotrexate combination therapy as compared with methotrexate alone: A magnetic resonance imaging study of 318 methotrexate-naive rheumatoid arthritis patients. Arthritis Rheum 2011;63:3712-22.

31.   Dias EM, Lukas C, Landewe R, Fatenejad S, van der Heijde D. Reliability and sensitivity to change of the simple erosion narrowing score compared with the sharp-van der heijde method for scoring radiographs in rheumatoid arthritis. Ann Rheum Dis 2008;67:375-9.

32.   Lukas C, Braun J, van der Heijde D, Hermann KG, Rudwaleit M, Ostergaard M, et al. Scoring inflammatory activity of the spine by magnetic resonance imaging in ankylosing spondylitis: A multireader experiment. J Rheumatol 2007;34:862-70.

16

33. Dakkak YJ, Boeters DM, Boer AC, Reijnierse M, van der Helm-van Mil AHM. What is the additional value of mri of the foot to the hand in undifferentiated arthritis to predict rheumatoid arthritis development? Arthritis Res Ther 2019;21:56.

34. Glinatsi D, Bird P, Gandjbakhch F, Haavardsholm EA, Peterfy CG, Vital EM, et al. Development and validation of the omeract rheumatoid arthritis magnetic resonance tenosynovitis scoring system in a multireader exercise. J Rheumatol 2017.

35. Niemantsverdriet E, van der Helm-van Mil AHM. Imaging detected tenosynovitis of metacarpophalangeal and wrist joints: An increasingly recognised characteristic of rheumatoid arthritis. Clin Exp Rheumatol 2018;36 Suppl 114:131-8.

36. Ostergaard M, Klarlund M. Importance of timing of post-contrast mri in rheumatoid arthritis: What happens during the first 60 minutes after iv gadolinium-dtpa? Ann Rheum Dis 2001;60:1050-4.

Table 1. Interreader intraclass correlation coefficients and average status scores according to the RAMRIS in early arthritis patients (n=441 in total) and patients with clinically suspect arthralgia (n=82)

Results are based on a mean-rating, absolute agreement, two-way mixed effects model.
RAMRIS: Rheumatoid Arthritis Magnetic Resonance Imaging Score; n: number of patients; k: number of readers; MTP: metatarsophalangeal; MCP: metacarpophalangeal; BME: bone marrow edema; ICC: intraclass correlation coefficients; CI: confidence interval; IQR: interquartile range; SD: standard deviation.
Consecutive early arthritis patients were scored by four readers, the first 215 consecutive patients were scored by reader 1 and 2, the remaining 226 consecutive patients by reader 3 and 4. 82 consecutive patients with clinically suspect arthralgia were scored by reader 5 and 6. Of these 82 patients, 40 were randomly selected and in addition also scored by 7 other readers.

18

Table 2. Intrareader intraclass correlation coefficients and average status scores according to the RAMRIS in early arthritis patients (n=15) by two readers

Results are based on a single-measure, absolute agreement, two-way mixed effects model.
Based on the repeated scoring of randomly selected patients (patients are scored twice by each reader).
n: number of patients; MTP: metatarsophalangeal; MCP: metacarpophalangeal; BME: bone marrow edema; ICC: interclass correlation coefficient; CI: confidence interval; IQR: interquartile range; SD: standard deviation.

19

Table 3. Change scores intraclass correlation coefficients and average scores according to the RAMRIS in early arthritis patients (n=30) from baseline until 12 months of follow-up by two readers

IQR: interquartile range; SD: standard deviation; ICC: intraclass correlation coefficient; CI: confidence interval; SDC: smallest detectable change; n: number; MTP: metatarsophalangeal; MCP: metacarpophalangeal; BME: bone marrow edema. The change scores of individual readers (reader 1 and reader 2) are depicted, in addition to the average change over time of both readers ('median change' and 'mean change'). Change ICCs are based on a mean-rating, absolute agreement, two-way mixed effects model. 'Change, n' is defined as the numbers of patients with change of more than an average of both readers of >0 or >0.5.'Change >SDC, n' represents the number of patients with change after using the SDC as a cut-off.

Figure 1. Flowchart of patient selection

The 441 early arthritis patients included the following diagnoses: RA according to the 2010 classification criteria (n=157), unclassified arthritis (n=148), psoriatic arthritis or spondyloarthritis (n=45), inflammatory osteoarthritis (n=23), crystalarthropathy (n=21), remitting seronegative symmetrical synovitis with pitting edema (n=12), reactive arthritis (n=7) and another diagnoses (n=28). The 30 early arthritis patients that were used to determine the reliability of change scores included the following diagnoses: RA according to the 2010 classification cirteria (n=21), unclassified arthritis (n=6), inflammatory osteoarthritis (n=2), and remitting seronegative symmetrical synovitis with pitting edema (n=1). 29 out of the 30 patients received DMARDs, one received NSAIDs only.

21

Figure 2. Bland-Altman plot depicting interreader agreement of status scores of the metatarsophalangeal-joints for the two groups of early arthritis patients and for patients with clinically suspect arthralgia

Bland-Altman plots of assessment of agreement of scores of the two readers. The Y-axes demonstrates the absolute difference between reader 1 minus reader 2. The X-axes denotes the average value between the two readers ((reader 1-reader 2)/2). The middle dotted line depicts the mean difference, the upper and lower dotted lines depict the ±95% limits of agreement. Different cohorts of patients are depicted; A: 215 early arthritis patients scored by reader 1 and 2, B: 226 early arthritis patients scored by reader 3 and 4, C: 82 patients with clinically suspect arthralgia scored by reader 5 and 6. From left to right the following MRI-lesions are depicted: bone marrow edema (BME), synovitis, tenosynovitis and erosions.

Figure 3. Bland-Altman plot depicting interreader agreement of change scores of the metatarsophalangeal-joints in early arthritis patients (n=30) from baseline until 12 months of follow-up

Bland-Altman plots of assessment of agreement of change scores of the two readers. The Y-axes demonstrates the difference in the change scores (12 months minus baseline score) for reader 1 minus reader 2. The X-axes denotes the average change score between the two readers ((change score reader 1 - change score reader 2)/2). The middle dotted line depicts the mean difference, the upper and lower dotted lines depict the ±95% limits of agreement. From left to right and top to bottom the following MRI-lesions are depicted: bone marrow edema (BME), synovitis, tenosynovitis and erosions.
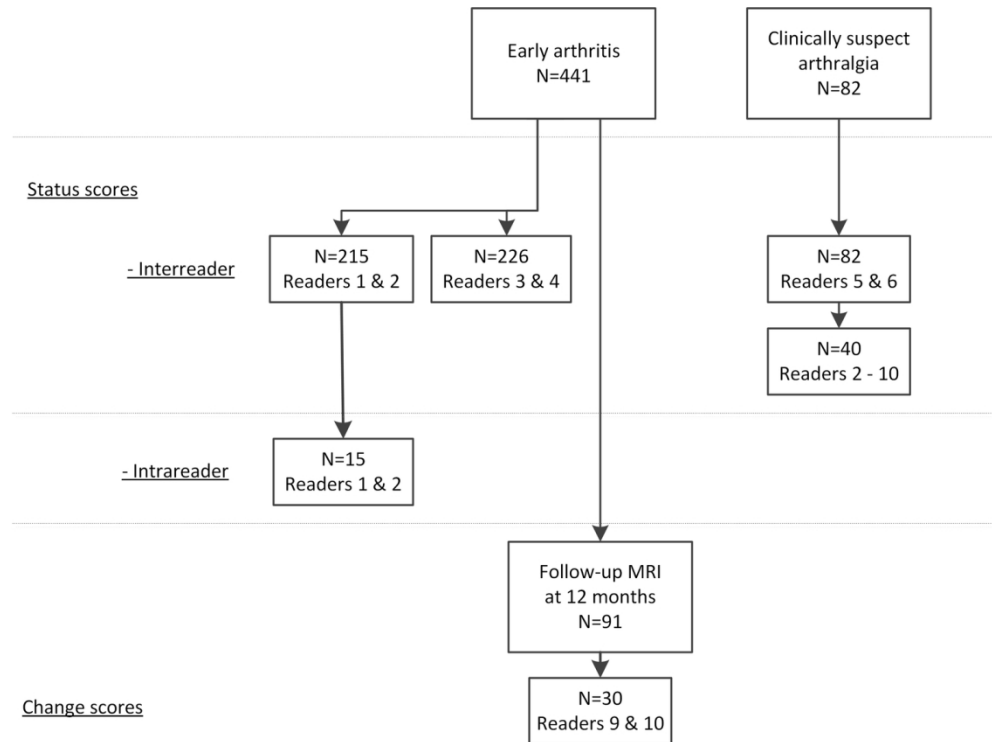
Figure 1. Flowchart of patient selection / The 441 early arthritis patients included the following diagnoses: RA according to the 2010 classification criteria (n=157), unclassified arthritis (n=148), psoriatic arthritis or spondyloarthritis (n=45), inflammatory osteoarthritis (n=23), crystalarthropathy (n=21), remitting seronegative symmetrical synovitis with pitting edema (n=12), reactive arthritis (n=7) and another diagnoses (n=28). The 30 early arthritis patients that were used to determine the reliability of change scores included the following diagnoses: RA according to the 2010 classification cirteria (n=21), unclassified arthritis (n=6), inflammatory osteoarthritis (n=2), and remitting seronegative symmetrical synovitis with pitting edema (n=1). 29 out of the 30 patients received DMARDs, one received NSAIDs only.
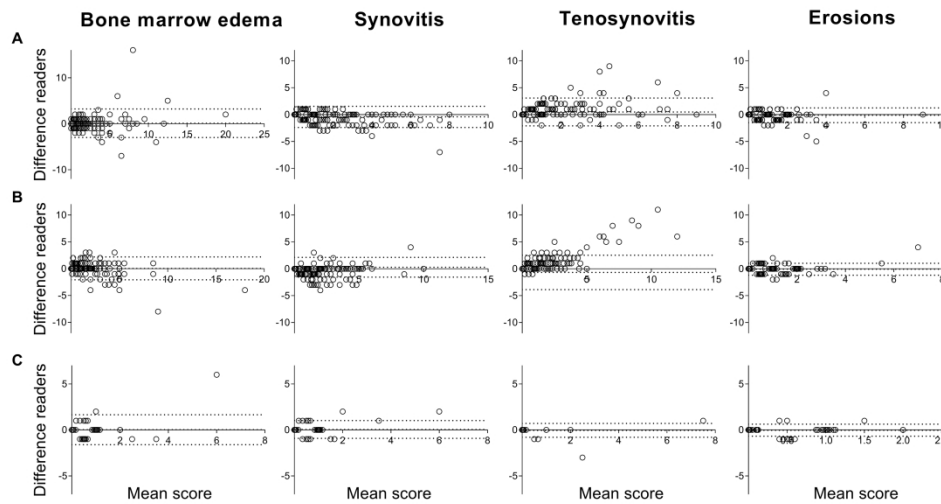
135x100mm (300 x 300 DPI)

Figure 2. Bland-Altman plot depicting interreader agreement of status scores of the metatarsophalangeal-joints for the two groups of early arthritis patients and for patients with clinically suspect arthralgia / Bland-Altman plots of assessment of agreement of scores of the two readers. The Y-axes demonstrates the absolute difference between reader 1 minus reader 2. The X-axes denotes the average value between the two readers ((reader 1-reader 2)/2). The middle dotted line depicts the mean difference, the upper and lower dotted lines depict the ±95% limits of agreement. Different cohorts of patients are depicted; A: 215 early arthritis patients scored by reader 1 and 2, B: 226 early arthritis patients scored by reader 3 and 4, C: 82 patients with clinically suspect arthralgia scored by reader 5 and 6. From left to right the following MRI-lesions are depicted: bone marrow edema (BME), synovitis, tenosynovitis and erosions.

180x98mm (600 x 600 DPI)

Figure 3. Bland-Altman plot depicting interreader agreement of change scores of the metatarsophalangeal-joints in early arthritis patients (n=30) from baseline until 12 months of follow-up / Bland-Altman plots of assessment of agreement of change scores of the two readers. The Y-axes demonstrates the difference in the change scores (12 months minus baseline score) for reader 1 minus reader 2. The X-axes denotes the average change score between the two readers ((change score reader 1 - change score reader 2)/2). The middle dotted line depicts the mean difference, the upper and lower dotted lines depict the ±95% limits of agreement. From left to right and top to bottom the following MRI-lesions are depicted: bone marrow edema (BME), synovitis, tenosynovitis and erosions.

97x77mm (600 x 600 DPI)

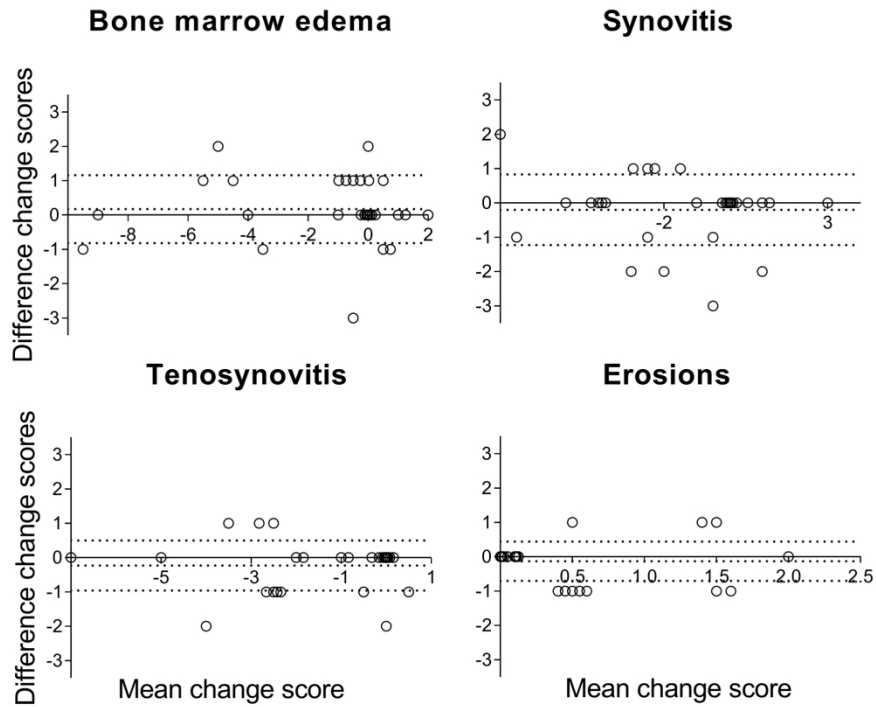**Table 1.** Interreader intraclass correlation coefficients and average status scores according to the RAMRIS in early arthritis patients (n=441 in total) and patients with clinically suspect arthralgia (n=82)

Patient population		Early arthritis n=215, k=2	Early arthritis n=226, k=2	Clinically suspect arthralgia n=82, k=2	Clinically suspect arthralgia n=40, k=9
MTP-joints					
BME	ICC (95%CI)	0.91 (0.90-0.93)	0.92 (0.90-0.94)	0.77 (0.64-0.85)	0.95 (0.93-0.97)
	Median (IQR, max)	0.5 (0-1.5; 20)	0.5 (0-1.5; 18)	0.0 (0-0.5; 6)	0.1 (0.0-0.4; 5.4)
	Mean (SD)	1.5 (2.6)	1.1 (2.0)	0.4 (0.9)	0.4 (0.9)
Synovitis	ICC (95%CI)	0.90 (0.84-0.94)	0.92 (0.88-0.94)	0.92 (0.87-0.95)	0.98 (0.97-0.99)
	Median (IQR, max)	1.0 (0-1.5; 8)	0.5 (0-1.5; 10)	0.0 (0-0.5; 6)	0.2 (0.0-0.8; 6.2)
	Mean (SD)	1.3 (1.7)	1.2 (1.7)	0.4 (0.9)	0.5 (1.0)
Tenosynovitis	ICC (95%CI)	0.85 (0.78-0.90)	0.80 (0.69-0.86)	0.96 (0.93-0.97)	0.99 (0.98-0.99)
	Median (IQR, max)	0.0 (0-1.5; 9)	0.0 (0-1.5; 12)	0.0 (0-0.0; 7.5)	0.0 (0.0-0.1; 6.8)
	Mean (SD)	1.1 (1.8)	1.1 (1.9)	0.2 (0.9)	0.3 (1.1)
Erosions	ICC (95%CI)	0.88 (0.84-0.91)	0.89 (0.86-0.92)	0.87 (0.81-0.92)	0.96 (0.93-0.97)
	Median (IQR, max)	0.5 (0-1.0; 9)	0.0 (0-1.0; 7)	0.0 (0-0.5; 2)	0.0 (0.0-0.7; 1.4)
	Mean (SD)	0.7 (1.0)	0.6 (0.9)	0.2 (0.5)	0.3 (0.5)
MCP-joints					
BME	ICC (95%CI)	0.60 (0.48-0.70)	0.90 (0.82-0.93)	0.78 (0.65-0.86)	0.93 (0.89-0.96)
	Median (IQR, max)	0 (0-0.5; 7.5)	0 (0-1; 14)	0 (0-0; 2)	0 (0-0; 0.9)
	Mean (SD)	0.6 (1.3)	1.0 (1.4)	0.1 (0.4)	0.1 (0.3)
Synovitis	ICC (95%CI)	0.91 (0.88-0.94)	0.92 (0.89-0.94)	0.91 (0.86-0.94)	0.99 (0.98-0.99)
	Median (IQR, max)	0 (0-1; 6)	0 (0-1; 9)	0 (0-0; 5)	0 (0-0; 4.9)
	Mean (SD)	0.9 (1.5)	0.8 (1.4)	0.2 (0.6)	0.2 (0.8)
Tenosynovitis	ICC (95%CI)	0.91 (0.88-0.93)	0.91 (0.87-0.94)	0.95 (0.92-0.97)	0.98 (0.96-0.99)
	Median (IQR, max)	0 (0-2.5; 8.5)	1.5 (0-3; 13)	0 (0-1; 6)	0.1 (0-1; 5.0)
	Mean (SD)	1.5 (1.9)	2.1 (2.4)	0.8 (1.4)	0.8 (1.3)
Erosions	ICC (95%CI)	0.93 (0.91-0.95)	0.92 (0.90-0.94)	0.93 (0.89-0.96)	0.97 (0.95-0.98)
	Median (IQR, max)	0 (0-1; 5)	0 (0-5; 5)	0 (0-1; 4)	0 (0-0.9; 1.9)
	Mean (SD)	0.6 (1.0)	0.7 (0.9)	0.5 (0.9)	0.4 (0.6)
Wrist					
BME	ICC (95%CI)	0.91 (0.88-0.93)	0.93 (0.75-0.97)	0.93 (0.90-0.96)	0.96 (0.94-0.98)
	Median (IQR, max)	1 (0-2.5; 24)	2 (0.5-4; 27)	0.5 (0-1.5; 11)	0.3 (0.1-1.2; 3.4)
	Mean (SD)	2.2 (3.4)	3.6 (4.8)	1.1 (1.7)	0.8 (1.0)
Synovitis	ICC (95%CI)	0.90 (0.87-0.92)	0.93 (0.91-0.95)	0.91 (0.87-0.94)	0.96 (0.95-0.98)
	Median (IQR, max)	1 (0-3; 8)	2 (0.5-3.5; 8)	0.5 (0-1.5; 5.5)	0.4 (0-1.1; 3.4)
	Mean (SD)	1.8 (2.1)	2.3 (2.1)	0.9 (1.1)	0.7 (0.9)
Tenosynovitis	ICC (95%CI)	0.93 (0.91-0.95)	0.96 (0.95-0.97)	0.97 (0.95-0.98)	0.99 (0.98-0.99)
	Median (IQR, max)	1 (0-4; 11.5)	1.5 (0-5; 13)	0 (0-0; 10)	0.1 (0-0.4; 9.2)
	Mean (SD)	2.5 (3.0)	2.8 (3.3)	0.6 (1.7)	0.8 (2.0)
Erosions	ICC (95%CI)	0.88 (0.84-0.91)	0.94 (0.92-0.95)	0.87 (0.79-0.92)	0.96 (0.94-0.98)
	Median (IQR, max)	1.5 (0.5-3.5; 11)	1.8 (0.5-3; 14)	1 (0.5-2.5; 7.5)	0.6 (0.2-2.3; 4.2)
	Mean (SD)	2.2 (2.2)	2.2 (2.2)	1.5 (1.5)	1.2 (1.2)

Results are based on a mean-rating, absolute agreement, two-way mixed effects model.
RAMRIS: Rheumatoid Arthritis Magnetic Resonance Imaging Score; n: number of patients; k: number of readers; MTP: metatarsophalangeal; MCP: metacarpophalangeal; BME: bone marrow edema; ICC: intraclass correlation coefficients; CI: confidence interval; IQR: interquartile range; SD: standard deviation.

Consecutive early arthritis patients were scored by four readers, the first 215 consecutive patients were scored by reader 1 and 2, the remaining 226 consecutive patients by reader 3 and 4. 82 consecutive patients with clinically suspect arthralgia were scored by reader 5 and 6. Of these 82 patients, 40 were randomly selected and in addition also scored by 7 other readers.

Table 2. Intrareader intraclass correlation coefficients and average status scores according to the RAMRIS in early arthritis patients (n=15) by two readers

Patient population		Early arthritis (n=15)	
		Reader 1	Reader 2
MTP-joints			
BME	ICC (95%CI)	0.96 (0.89-0.99)	0.98 (0.95-0.99)
	Median (IQR, max)	0.0 (0.0-2.5; 5)	0.0 (0.0-2.5; 8.5)
	Mean (SD)	1.1 (1.6)	1.4 (2.3)
Synovitis	ICC (95%CI)	0.90 (0.74-0.97)	0.98 (0.94-0.99)
	Median (IQR, max)	0.5 (0.0-1.0; 3.5)	1.0 (0.0-2.0; 5.5)
	Mean (SD)	0.8 (1.2)	1.3 (1.8)
Tenosynovitis	ICC (95%CI)	0.97 (0.91-0.99)	0.84 (0.58-0.94)
	Median (IQR, max)	0.0 (0.0-3.5; 8)	0.0 (0.0; 4.5)
	Mean (SD)	1.7 (2.9)	0.6 (1.4)
Erosions	ICC (95%CI)	0.71 (0.35-0.89)	0.92 (0.78-0.97)
	Median (IQR, max)	0.0 (0.0-1.0; 2)	0.0 (0.0-1.0; 4.5)
	Mean (SD)	0.4 (0.6)	0.7 (1.2)
MCP-joints			
BME	ICC (95%CI)	0.86 (0.61-0.95)	0.80 (0.51-0.93)
	Median (IQR, max)	0 (0-0.5; 2.5)	0 (0-0.5; 2.5)
	Mean (SD)	0.4 (0.8)	0.3 (0.7)
Synovitis	ICC (95%CI)	0.88 (0.68-0.96)	0.96 (0.89-0.99)
	Median (IQR, max)	0 (0-1; 2)	0 (0-1; 4)
	Mean (SD)	0.5 (0.7)	0.9 (1.3)
Tenosynovitis	ICC (95%CI)	0.97 (0.91-0.99)	0.94 (0.83-0.98)
	Median (IQR, max)	1 (0-3.5; 4.5)	0 (0-2; 3.5)
	Mean (SD)	1.4 (1.8)	1.0 (1.3)
Erosions	ICC (95%CI)	0.94 (0.82-0.98)	0.99 (0.96-1.00)
	Median (IQR, max)	0 (0-1. 4.5)	0 (0-1; 5.5)
	Mean (SD)	0.7 (1.3)	0.8 (1.6)
Wrist			
BME	ICC (95%CI)	0.77 (0.45-0.92)	0.90 (0.72-0.96)
	Median (IQR, max)	0.5 (0-1.5; 3.5)	0 (0-2; 5)
	Mean (SD)	0.8 (1.0)	1.2 (1.6)
Synovitis	ICC (95%CI)	0.92 (0.71-0.98)	0.95 (0.81-0.98)
	Median (IQR, max)	0 (0-1.5; 3.5)	0.5 (0-1; 5)
	Mean (SD)	0.8 (1.4)	1.1 (1.5)
Tenosynovitis	ICC (95%CI)	0.99 (0.97-1.00)	0.99 (0.98-1.00)
	Median (IQR, max)	0 (0-1; 6)	0.5 (0-2; 6)
	Mean (SD)	1.1 (1.9)	1.4 (2.0)
Erosions	ICC (95%CI)	0.95 (0.86-0.98)	0.97 (0.90-0.99)
	Median (IQR, max)	1 (0-3.5; 6.5)	1 (0-3; 5)
	Mean (SD)	1.9 (2.1)	1.4 (1.7)

Results are based on a single-measure, absolute agreement, two-way mixed effects model.
Based on the repeated scoring of randomly selected patients (patients are scored twice by each reader).
n: number of patients; MTP: metatarsophalangeal; MCP: metacarpophalangeal; BME: bone marrow edema; ICC: interclass correlation coefficient; CI: confidence interval; IQR: interquartile range; SD: standard deviation

Table 3. Change scores intraclass correlation coefficients and average scores according to the RAMRIS in early arthritis patients (n=30) from baseline until 12 months of follow-up by two readers

	Median change (IQR)	Mean change (SD)	Median change (IQR)	Mean change (SD)	Median change (IQR; max)	Mean change (SD)	ICC change (95% CI)	SDC	Change >0, n(%)	Change >0.5, n (%)	Change >SDC, n (%)
	Reader 1	Reader 1	Reader 2	Reader 2	Average both readers	Average both readers					
MTP											
BME	0 (-2-0)	-1.2 (2.9)	0 (-1-0)	-1.4 (2.9)	0 (-1-0;-9.5)	-1.3 (2.9)	0.97 (0.94-0.99)	1.0	20 (67)	11 (37)	11 (37)
Synovitis	-1 (-4-0)	-1.6 (2.3)	-1 (-3-0)	-1.4 (2.5)	-0.75 (-3-0;-7)	-1.5 (2.4)	0.90 (0.90-0.98)	1.0	22 (73)	20 (67)	20 (67)
Tenosynovitis	-1 (-3-0)	-1.5 (1.9)	0 (-2-0)	-1.2 (1.9)	-0.25 (-2.5-0;-7)	-1.4 (1.8)	0.96 (0.91-0.98)	0.7	16 (53)	14 (47)	14 (47)
Erosions	0 (0-0)	0.3 (0.7)	0 (0-1)	0.4 (0.7)	0 (0-0.5;2)	0.4 (0.6)	0.77 (0.52-0.89)	0.6	11 (37)	5 (17)	5 (17)
MCP											
BME	0 (-2-0)	-0.9 (1.9)	0 (-2-0)	-0.8 (1.9)	0 (-1.5-0;-6.5)	-0.9 (1.8)	0.94 (0.88-0.97)	0.9	16 (53)	13 (43)	13 (43)
Synovitis	-2 (-3-0)	-1.3 (2.8)	-2 (-3-0)	-1.4 (2.7)	-1.5 (-3-0.5;-6.5)	-1.4 (2.7)	0.97 (0.93-0.98)	1.0	26 (87)	25 (83)	25 (83)
Tenosynovitis	0 (-3-0)	-1.4 (2.7)	0 (-2-0)	-1.3 (2.6)	0 (-2.5-0;-10)	-1.3 (2.6)	0.97 (0.94-0.99)	0.9	15 (50)	14 (47)	14 (47)
Erosions	0 (0-1)	0.4 (0.7)	0 (0-1)	0.4 (0.6)	0 (0-1;2)	0.4 (0.6)	0.94 (0.87-0.97)	0.3	11 (37)	9 (30)	11 (37)
Wrist											
BME	-1 (-3-0)	-2.8 (5.4)	-1 (-5-0)	-2.9 (4.8)	-1 (-4-0;-19)	-2.8 (5.0)	0.98 (0.96-0.99)	1.5	25 (83)	20 (67)	14 (47)
Synovitis	-1 (-4-0)	-1.6 (2.8)	-2 (-4-0)	-2.3 (2.8)	-1.5 (-4-0;-7.5)	-2.8 (5.0)	0.96 (0.81-0.98)	0.9	24 (80)	20 (67)	20 (67)
Tenosynovitis	-2 (-4-0)	-2.6 (3.3)	-2 (-5-0)	-2.6 (3.1)	-1.5 (-4.5-0;-10.5)	-2.6 (3.2)	0.98 (0.96-0.99)	0.9	23 (77)	19 (63)	19 (63)
Erosions	0 (0-1)	1.0 (2.3)	0 (0-1)	0.7 (1.6)	0 (0-1;8)	0.9 (1.9)	0.91 (0.81-0.96)	1.1	18 (60)	9 (30)	5 (17)

IQR: interquartile range; SD: standard deviation; ICC: intraclass correlation coefficient; CI: confidence interval; SDC: smallest detectable change; n: number; MTP: metatarsophalangeal; MCP: metacarpophalangeal; BME: bone marrow edema. The change scores of individual readers (reader 1 and reader 2) are depicted, in addition to the average change over time of both readers ('median change' and 'mean change'). Change ICCs are based on a mean-rating, absolute agreement, two-way mixed effects model. 'Change, n' is defined as the numbers of patients with change of more than an average of both readers of >0 or >0.5.'Change >SDC, n' represents the number of patients with change after using the SDC as a cut-off.