Validity and Responsiveness of Combined Inflammation and Combined Joint Damage Scores based on the OMERACT Rheumatoid Arthritis MRI Scoring System (RAMRIS)

Author Name	ORCID ID
Ulf Sundin	https://orcid.org/0000-0003-1860-6150
Mikkel Østergaard	https://orcid.org/0000-0003-3690-467X
Daniel Glinatsi	https://orcid.org/0000-0003-0452-8200
Anna-Birgitte Aga	
Kim Hørslev-Petersen	https://orcid.org/0000-0002-5475-7610
Merete L. Hetland	https://orcid.org/0000-0003-4229-6818
Kristian Stengard-Pedersen	https://orcid.org/0000-0003-0779-8054
Peter Junker	https://orcid.org/0000-0002-2684-9964
Bo J. Ejbjerg	
Paul Bird	https://orcid.org/0000-0003-3314-3270
Philip G. Conaghan	https://orcid.org/0000-0002-3478-5665
Siri Lillegraven	https://orcid.org/0000-0003-3615-0134
Espen A. Haavardsholm	https://orcid.org/0000-0002-1427-4745

Key Indexing Terms

Rheumatoid Arthritis, Magnetic Resonance Imaging, Clinical Trials, Outcome Assessment, OMERACT

Departments/Institutions

Department of Rheumatology, Diakonhjemmet Hospital; Institute of Health and Society, University of Oslo, Oslo, Norway; Copenhagen Center for Arthritis Research (COPECARE), Center for Rheumatology and Spine Diseases, Rigshospitalet, Glostrup; Department of Clinical Medicine, University of Copenhagen, Copenhagen; King Christian 10th Hospital for Rheumatic Diseases; University of Southern Denmark, Institute of Regional Health Research, Graasten; Institute of Clinical Medicine, Aarhus University Hospital, Århus; Odense University Hospital; Institute of Clinical Research, University of Southern Denmark, Odense; Zealand University Hospital, Køge, Denmark; University of New South Wales, Sydney, Australia; Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds; NIHR Leeds Biomedical Research Centre, Leeds, UK.

Financial Support

This project has received no outside funding

Conflicts of interest

No conflicts of interests to report

Authors

U. Sundin, MD, Research Fellow, Department of Rheumatology, Diakonhjemmet Hospital, and Institute of Health and Society, University of Oslo, Oslo, Norway.

M. Østergaard, MD, PhD, DMSc, Professor, COPECARE, Center for Rheumatology and Spine Diseases, Rigshospitalet, Glostrup, and the Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.

D. Glinatsi, MD, PhD, Postdoctoral Researcher, COPECARE, Center for Rheumatology and Spine Diseases, Rigshospitalet, Glostrup, Denmark.

A.B. Aga, MD, PhD, Senior Consultant and Postdoctoral Researcher, Department of Rheumatology, Diakonhjemmet Hospital Oslo, Norway.

K. Hørslev-Petersen, MD, DMSc, Professor, Senior Consultant, King Christian 10th Hospital for Rheumatic Diseases, and the Institute of Regional Health Research, University of Southern Denmark, Graasten, Denmark.

M.L. Hetland, MD, PhD, DMSc, Professor, COPECARE, Center for Rheumatology and Spine Diseases, Rigshospitalet, Glostrup, and the Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.

K. Stengaard-Pedersen, MD, DMSc, Professor Emeritus, Institute of Clinical Medicine, Aarhus University Hospital, Aarhus, Denmark.

P. Junker, MD, DMSc, Professor, Department of Rheumatology, Odense University Hospital and Institute of Clinical Research, University of Southern Denmark, Odense, Denmark.

B.J. Ejbjerg, MD, PhD, Chief Consultant, Zealand University Hospital, Køge, Denmark.

P. Bird, BMed (Hons), FRACP, PhD, Grad Dip MRI, Associate Professor, University of NSW, Sydney, Australia.

P.G. Conaghan, MB, BS, PhD, FRACP, FRCP, Professor of Musculoskeletal Medicine, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre, Leeds, United Kingdom.

S. Lillegraven, MD, MPH, PhD, Postdoctoral Researcher, Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway.

E.A. Haavardsholm, MD, PhD, Professor, Department of Rheumatology, Diakonhjemmet Hospital, and Institute of Health and Society, University of Oslo, Oslo, Norway.

Correspondence

Ulf Sundin, MD, Diakonhjemmet Sykehus, Diakonveien 12, 0370 Oslo, Norway. Uffe.sundin@gmail.com.

Running Head

Combined RA MRI Scores

Word count

1505/1500

Abstract

Objective

The OMERACT RA MRI scoring system (RAMRIS) is used in clinical RA trials. We have explored methods to combine the RAMRIS features into valid and responsive scores for a) inflammation and b) joint damage.

Methods

We used data from three large randomized early RA trials to assess five methods to develop a combined score for inflammation based on RAMRIS bone marrow edema, synovitis and tenosynovitis scores, and a combined joint damage score based on erosions and joint space narrowing. Methods included unweighted summation, normalized summation, and three different variants of weighted summation of the RAMRIS features. We used a derivation cohort to calculate summation weights to maximize the responsiveness of the combined score. Construct validity of the combined scores was examined by assessing correlations to imaging, clinical and biochemical measures. Responsiveness was tested by calculating the standardized response mean and the relative efficiency of each score in a validation cohort.

Results

Patient characteristics, as well as baseline and follow-up RAMRIS scores were comparable between cohorts. All combined scores were significantly correlated to other imaging, clinical and biochemical measures. Inflammation scores combined by normalized and weighted summation had significantly higher responsiveness in comparison to unweighted summation, with standardized response mean (95% CI) for unweighted summation 0.62 (0.51-0.73), normalized summation 0.73 (0.63-0.83), and weighted summation 0.74 (0.65-0.84). For the damage score, there was a trend towards higher responsiveness for weighted summation.

Conclusion

Combined MRI scores calculated by normalized or weighted summation of individual MRI pathologies were valid and responsive.

Introduction

Magnetic resonance imaging (MRI) allows detailed assessment of the synovial joint. In rheumatoid arthritis (RA), MRI is more sensitive than radiography for detecting bone erosions and cartilage loss (1-3), and can visualize the inflammatory lesions that precede joint destruction (4-8).

MRI features are frequently used as outcome measures in RA clinical trials (8, 9). Outcome Measures in Rheumatology (OMERACT) is an independent initiative to develop and validate outcome measures for clinical trials in rheumatic diseases (10, 11). The OMERACT RA MRI Scoring system (RAMRIS) outlines semi-quantitative scoring of five RA pathologies: bone erosions, joint space narrowing (JSN), synovitis, tenosynovitis and bone marrow edema in the wrist and metacarpophalangeal joints (2, 12, 13). However, the primary interest in clinical studies might be the total inflammatory activity or the progression of total structural joint damage.

The objective of this study was to develop and validate two combined MRI scores, one for inflammation and one for joint damage, derived from the five RAMRIS pathology scores, with emphasis on responsiveness and construct validity.

Materials and methods

Validation and derivation cohorts

We used data from the ARCTIC (14) trial as a derivation cohort for the combined scores. Performance of the scores was assessed in a validation cohort of pooled data from CIMESTRA (15) and OPERA (16). ARCTIC was a 24-month randomized clinical trial, studying ultrasound for treatment decision-making. Participants (n=230) were disease modifying anti-rheumatic drug (DMARD)-naïve early RA patients aged 18-75 years fulfilling 2010 ACR/EULAR criteria, with indication for DMARD-treatment. Both CIMESTRA and OPERA were randomized controlled trials. CIMESTRA studied treatment with methotrexate and intra-articular betamethasone in early RA, and the additional effect of adding cyclosporine to the regimen. OPERA studied the effect of adding adalimumab to methotrexate and intra-articular triamcinolone as first-line therapy in early RA. Participants (CIMESTRA n=160, OPERA n=180) were > 17 years, fulfilled the 1987 ACR criteria, and had moderate to severe disease activity.

Written informed consent was obtained from all participants. The trials were approved by the local ethics committees (approval reference numbers: ARCTIC: 2010/744; CIMESTRA: M-1959-98; OPERA: VEK-20070008).

Imaging

MRI of one hand (acquisition as outlined in the RAMRIS core set (12)) was performed together with conventional radiographs of hands and feet at baseline and 12 months in all three trials. A single reader (CIMESTRA/OPERA: DG, ARCTIC: US) blinded to treatment arm and clinical data scored the MRI images according to RAMRIS, with known chronological order. Reliability of scorings was overall very good (intra- and inter-reader comparisons for ARCTIC: appendix table 3, intra-reader for CIMESTRA/OPERA: previously published (17)).

Radiographs were scored according to the van der Heijde-modified Sharp score. In ARCTIC, ultrasound was performed yearly for all patients according to a validated scoring system (18).

Clinical parameters

Tender and swollen joint counts, pain, patient and physician global, and C-reactive protein were registered at each visit. In ARCTIC, erythrocyte sedimentation rate was also analyzed. Physical function was assessed by the Health Assessment Questionnaire in CIMESTRA and OPERA, and by the Patient-Reported Outcomes Measurement Information 20 item short form in ARCTIC.

Calculation of combined scores

We categorized RAMRIS scores as either inflammation (synovitis, tenosynovitis, BME) or damage (erosions, JSN), and calculated the combined score for each category. Calculation was done using five different approaches, aiming to find which method would provide the most responsive combined score.

Approach 1. Unweighted summation: Combined scores were calculated by numerical summation of the RAMRIS scores for each category. These scores were used as reference.

Additionally, we tested several methods for transformation of the RAMRIS scores, before summation:

Approach 2. Normalized summation: The RAMRIS scores differ in range and will therefore have a disproportionate part of the total score if summarized without transformation. To counteract this, scores were transformed to the same range before summation.

Approach 3. Weighted summation: Each RAMRIS score was transformed by a multiplication factor (weight). To maximize responsiveness, weights were calculated in a data driven approach to give the highest standardized response mean (SRM) to the resulting score in the derivation cohort. To make the system more adaptable, each RAMRIS score was divided into

three anatomical areas, which were weighted individually. The areas and corresponding weights are shown in the appendix, box 1.

Approach 4. Adjusted weighted summation: To simplify the weighting system, data-derived weights from approach 3) were rescaled to whole numbers according to rank. Adjustment of +/- 1 step was allowed to optimize performance (appendix box 1).

Approach 5. Single site weighted summation: As in 3), but weights were calculated for each individual bone, joint and tendon.

Statistical analysis

Baseline characteristics were described as proportions or median values as appropriate. Construct validity of the suggested combined MRI scores was tested by calculating the Spearman correlation coefficients to established disease measures. Responsiveness was tested by calculating the SRM $\left(SRM = \frac{mean\ score\ change}{SD_{mean\ score\ change}}\right)$ for the suggested combined scores, the RAMRIS scores, and radiographic parameters. Relative efficiency was computed for each combined score with unweighted summation as reference $\left(RE = \left(\frac{\mathrm{SRM}i}{\mathrm{SRM}ref}\right)^2\right)$. Confidence intervals for SRM and relative efficiency were estimated by bootstrapping with 5000 replications. Only patients where all variables were available for baseline and the 12-month visit were included. Data analyses were undertaken using STATA v.14 (StataCorp, College Station, USA).

Results

Patient characteristics

Data from 194 patients from the ARCTIC trial (derivation cohort), and 195 patients from CIMESTRA and OPERA (validation cohort) were used. A larger proportion of the patients in the derivation cohort were positive for anti-cyclic citrullinated peptide, (82% vs. 61.5%, $p < 0.001$), and disease activity variables were somewhat higher in the validation cohort (**appendix table 1**). Duration of symptoms at inclusion was longer in the derivation cohort (median 166 vs 91 days, $p < 0.001$). Otherwise, patient characteristics were comparable between the cohorts.

MRI parameters

Baseline scores for synovitis were slightly higher in the validation cohort. Median one-year changes of inflammatory scores were similar in both cohorts. Baseline median erosion scores were similar in both cohorts, while the JSN score was higher in the validation cohort. The median one-year change for both erosions and JSN were comparable between the cohorts (**appendix table 2**).

Construct validity

All combined scores were significantly correlated to other imaging, clinical and biochemical measures. MRI inflammation scores were most strongly associated to ultrasound inflammation parameters, while associations between MRI damage scores and radiographic measures were overall moderate (**table 1**).

Responsiveness

For inflammation, relative efficiency for normalized summation (approach 2), weighted summation (approach 3) and adjusted weighted summation (approach 4) were statistically significantly superior to unweighted summation (approach 1), when tested in the validation

cohort (**figure 1**). Approaches 3 and 4 provided the numerically highest SRM values (**table 2**), however differences between approaches 2, 3 and 4 were not statistically significant. For damage, no approach was significantly superior to unweighted summation, although approach 4 provided the highest SRM values.

Discussion

We have developed and tested combined MRI scores capturing the principal pathogenic constructs of RA: inflammation and damage. For clinical trial settings, these two measures might be more important than the scores of the individual MRI lesions.

In previous studies, combined scores have been obtained through slightly differing methods (3, 17, 19). To ensure comparability between studies, and to avoid biased reporting, there is need for consensus regarding which method to use (20).

It could be argued that if responsiveness were the sole priority, it would be easiest to use only the most responsive single pathology; e.g. tenosynovitis in the present study. However, that would discard a large proportion of MRI information. By weighted summation, we could obtain responsive combined scores, while still covering the full spectrum of pathology. Approaches using complex weightings derived from data resulted in the numerically most responsive scores, but the gain was marginal compared to the simpler normalization approach.

The strengths of these analyses include the large datasets, with baseline and 1-year follow-up MRI data of 289 patients from three RCTs in early RA. By separating our data in derivation and validation cohorts, we were able to assess the validity and generalizability of our proposed combined scores with higher confidence than if only one dataset had been used. Limitations include the lack of opportunity to examine the discriminative properties of the combined scores, as none of the original trials showed significant group differences for clinical or MRI endpoints. A dataset with clinical differences between the treatment arms is needed to examine this. The SRM values of our scores were relatively low compared to a similar study (19). This might be explained by limited changes in RAMRIS scores during the follow-up, especially for joint damage.

In conclusion, we found that combined MRI scores for inflammation and joint damage can be responsive and valid. Our data indicate that the responsiveness of combined scores for inflammation could be improved by using normalized or weighted summation of the RAMRIS pathologies, rather than unweighted summation. However, our results do not support promoting one of these approaches over another. For the combined damage scores, there was a trend favoring weighted summation, but results were inconclusive. The discriminative properties of the scores need to be tested in placebo-controlled clinical trials.

Acknowledgements

We thank Joe Sexton for help and advice on statistical calculations and support on using statistical software, and Lena Bugge Nordberg and Nina Paulshus Sundlisæter for help with the ARCTIC database. PGC is supported in part by the National Institute for Health Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. We also thank all investigators, study personnel and patients who have contributed to the clinical trials that this study is based on.

References

1. Dohn UM, Ejbjerg B, Boonen A, Hetland ML, Hansen MS, Knudsen LS, et al. No overall progression and occasional repair of erosions despite persistent inflammation in adalimumab-treated rheumatoid arthritis patients: Results from a longitudinal comparative mri, ultrasonography, ct and radiography study. Ann Rheum Dis 2011;70:252-8.

2. Ostergaard M, Bøyesen P, Eshed I, Gandjbakhch F, Lillegraven S, Bird P, et al. Development and preliminary validation of a magnetic resonance imaging joint space narrowing score for use in rheumatoid arthritis: Potential adjunct to the omeract ra mri scoring system. J Rheumatol 2011;38:2045-50.

3. Peterfy CG, Olech E, DiCarlo JC, Merrill JT, Countryman PJ, Gaylis NB. Monitoring cartilage loss in the hands and wrists in rheumatoid arthritis with magnetic resonance imaging in a multi-center clinical trial: Impress (nct00425932). Arthritis Res Ther 2013;15:R44.

4. Boyesen P, Haavardsholm EA, Ostergaard M, van der Heijde D, Sesseng S, Kvien TK. Mri in early rheumatoid arthritis: Synovitis and bone marrow oedema are independent predictors of subsequent radiographic progression. Ann Rheum Dis 2011;70:428-33.

5. Conaghan PG, O'Connor P, McGonagle D, Astin P, Wakefield RJ, Gibbon WW, et al. Elucidation of the relationship between synovitis and bone damage: A randomized magnetic resonance imaging study of individual joints in patients with early rheumatoid arthritis. Arthritis Rheum 2003;48:64-71.

6. Haavardsholm EA, Bøyesen P, Østergaard M, Schildvold A, Kvien TK. Magnetic resonance imaging findings in 84 patients with early rheumatoid arthritis: Bone marrow oedema predicts erosive progression. Ann Rheum Dis 2008;67:794-800.

7. Hetland ML, Stengaard-Pedersen K, Junker P, Ostergaard M, Ejbjerg BJ, Jacobsen S, et al. Radiographic progression and remission rates in early rheumatoid arthritis - mri bone oedema and anti-ccp predicted radiographic progression in the 5-year extension of the double-blind randomised cimestra trial. Ann Rheum Dis 2010;69:1789-95.

8. Peterfy C, Ostergaard M, Conaghan PG. Mri comes of age in ra clinical trials. Ann Rheum Dis 2013;72:794-6.

9. American College of Rheumatology Rheumatoid Arthritis Clinical Trials Task Force Imaging G, Outcome Measures in Rheumatology Magnetic Resonance Imaging Inflammatory Arthritis Working G. Review: The utility of magnetic resonance imaging for assessing structural damage in randomized controlled trials in rheumatoid arthritis. Arthritis Rheum 2013;65:2513-23.

10. Boers M, Kirwan JR, Tugwell P, Beaton D, Bingham CO, 3rd, Conaghan PG, et al. The omeract handbook. OMERACT; 2018 [updated 2018; cited 2018-09-03]; Available from: www.omeract.org/resources.

11. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: Omeract filter 2.0. J Clin Epidemiol 2014;67:745-53.

12. 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.

13. 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;44:1688-93.

14. Haavardsholm EA, Aga AB, Olsen IC, Lillegraven S, Hammer HB, Uhlig T, et al. Ultrasound in management of rheumatoid arthritis: Arctic randomised controlled strategy trial. BMJ 2016;354:i4205.

15. Hetland ML, Stengaard-Pedersen K, Junker P, Lottenburger T, Ellingsen T, Andersen LS, et al. Combination treatment with methotrexate, cyclosporine, and intraarticular betamethasone compared with methotrexate and intraarticular betamethasone in early active rheumatoid arthritis: An investigator-initiated, multicenter, randomized, double-blind, parallel-group, placebo-controlled study. Arthritis Rheum 2006;54:1401-9.

16. Horslev-Petersen K, Hetland ML, Junker P, Podenphant J, Ellingsen T, Ahlquist P, et al. Adalimumab added to a treat-to-target strategy with methotrexate and intra-articular triamcinolone in early rheumatoid arthritis increased remission rates, function and quality of life. The opera study: An investigator-initiated, randomised, double-blind, parallel-group, placebo-controlled trial. Ann Rheum Dis 2014;73:654-61.

17. Glinatsi D, Baker JF, Hetland ML, Horslev-Petersen K, Ejbjerg BJ, Stengaard-Pedersen K, et al. Magnetic resonance imaging assessed inflammation in the wrist is associated with patient-reported physical impairment, global assessment of disease activity and pain in early rheumatoid arthritis: Longitudinal results from two randomised controlled trials. Ann Rheum Dis 2017;76:1707-15.

18. Hammer HB, Bolton-King P, Bakkeheim V, Berg TH, Sundt E, Kongtorp AK, et al. Examination of intra and interrater reliability with a new ultrasonographic reference atlas for scoring of synovitis in patients with rheumatoid arthritis. Ann Rheum Dis 2011;70:1995-8.

19. Haavardsholm EA, Ostergaard M, Hammer HB, Boyesen P, Boonen A, van der Heijde D, et al. Monitoring anti-tnfalpha treatment in rheumatoid arthritis: Responsiveness of magnetic resonance imaging and ultrasonography of the dominant wrist joint compared with conventional measures of disease activity and structural damage. Ann Rheum Dis 2009;68:1572-9.

20. 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.

Figure Legends

FIGURE 1: RELATIVE EFFICIENCY OF COMBINED SCORES FOR INFLAMMATION AND JOINT DAMAGE, VALIDATION COHORT. ERROR BARS REPRESENT 95% CI.

Tables and figures

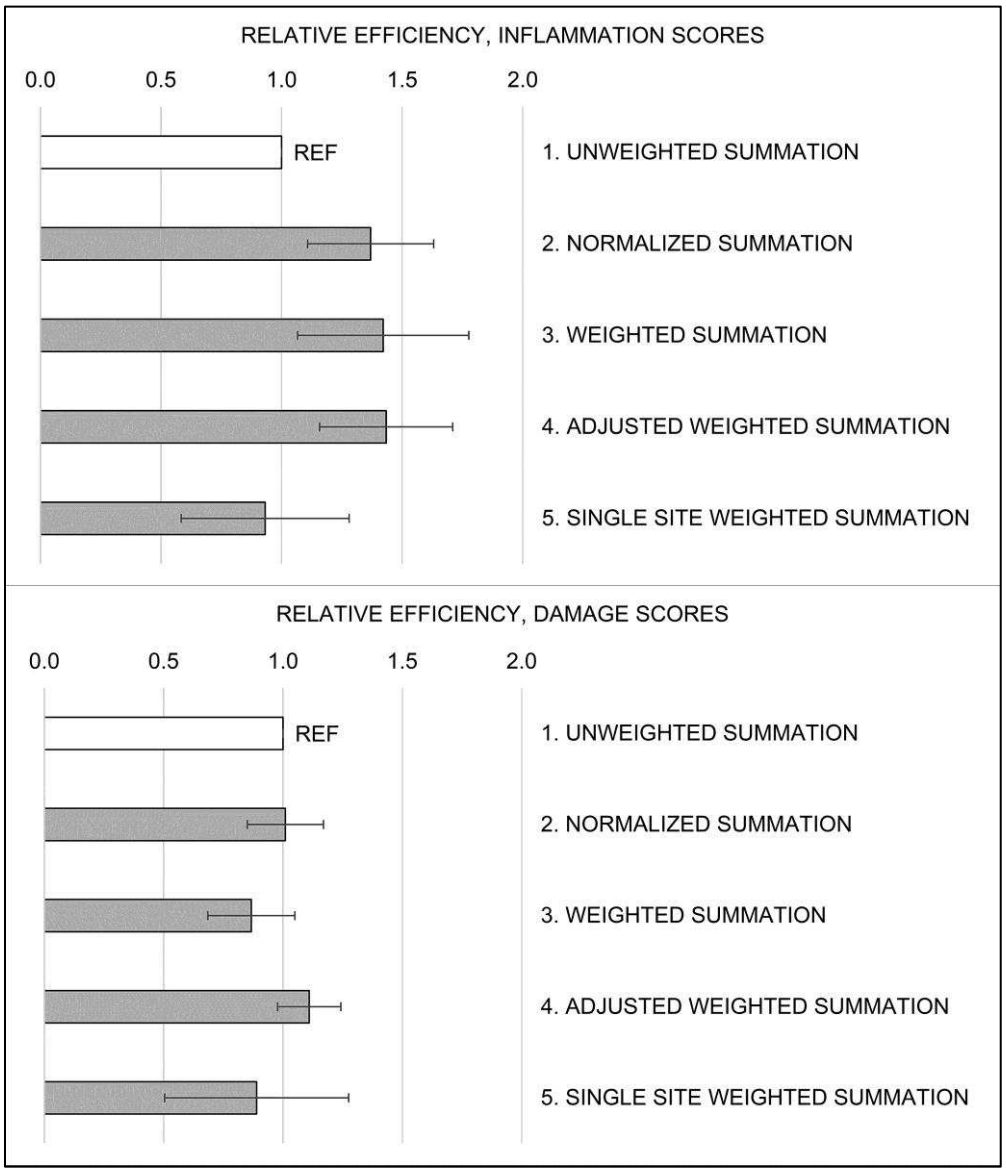
TABLE 1: SPEARMAN CORRELATION COEFFICIENTS BETWEEN MRI COMBINED SCORES AND CLINICAL, RADIOGRAPHIC (CR) AND ULTRASOUND PARAMETERS, ALL COHORTS

Inflammation scores	**DAS28**	**TJC28**	**SJC28**	**Pt.gl**	**Ph.gl**	**CRP**	**ESR***	**PFTS*‡**	**USPD***	**USBM***
1. Unweighted Summation	0.32	0.14	0.42	0.12	0.32	0.40	0.24	-0.26	0.49	0.54
2. Normalized Summation	0.36	0.18	0.46	0.14	0.35	0.43	0.24	-0.25	0.52	0.55
3. Weighted Summation	0.27	0.13	0.38	0.11	0.29	0.33	0.17	-0.29	0.47	0.52
4. Adjusted Weighted Summation	0.30	0.15	0.42	0.11	0.31	0.37	0.19	-0.28	0.50	0.54
5. Single Site Weighted Summation	0.29	0.14	0.36	0.11	0.29	0.32	0.21	-0.26	0.46	0.48
Damage scores	**CR Erosions**		**CR JSN**		**CR Total**					
1. Unweighted Summation	0.43		0.33		0.40					
2. Normalized Summation	0.41		0.32		0.38					
3. Weighted Summation	0.43		0.34		0.40					
4. Adjusted Weighted Summation	0.43		0.33		0.40					
5. Single Site Weighted Summation	0.42		0.37		0.43					

*ARCTIC only. ‡ Negative correlation due to inverse scale of PROMIS T-score. Abbreviations: CR: Conventional Radiography, CR scored by van der Heijde modified Sharp score (hands, wrists and feet). CRP: C-Reactive Protein (mg/L). DAS: Disease Activity Score (28-joints, range 0-10). Pt/Ph.gl: Patient's/Physician's Global Assessment Visual Analog Scale (range 0-100). PFTS: PROMIS T-score. SJC: Swollen Joint Count (28 joints). TJC: Tender Joint Count (28 joints). USPD/BM: Ultrasound Power Doppler/B-Mode. All coefficients significant at the 0.05 level.

TABLE 2: STANDARDIZED RESPONSE MEANS OF COMBINED SCORES, INDIVIDUAL MRI PATHOLOGIES, RADIOGRAPHIC (CR) AND CLINICAL PARAMETERS (95% CI)		
Inflammatory Measures	**Derivation Cohort**	**Validation Cohort**
1. Unweighted Summation	0.78 (0.70-0.85)	0.62 (0.51-0.73)
2. Normalized Summation	0.82 (0.74-0.90)	0.73 (0.63-0.83)
3. Weighted Summation	0.84 (0.76-0.92)	0.74 (0.64-0.84)
4. Adjusted Weighted Summation	0.84 (0.76-0.92)	0.74 (0.65-0.84)
5. Single Site weighted Summation	1.10 (0.99-1.21)	0.60 (0.50-0.70)
RAMRIS Synovitis	0.74 (0.65-0.83)	0.65 (0.54-0.76)
RAMRIS Tenosynovitis	0.81 (0.73-0.89)	0.76 (0.66-0.86)
RAMRIS Bone Marrow Edema	0.36 (0.29-0.42)	0.13 (0.02-0.24)
Damage Measures		
1. Unweighted Summation	0.35 (0.21-0.49)	0.43 (0.35-0.52)
2. Normalized Summation	0.29 (0.14-0.43)	0.44 (0.38-0.50)
3. Weighted Summation	0.43 (0.31-0.55)	0.40 (0.30-0.51)
4. Adjusted Weighted Summation	0.38 (0.25-0.51)	0.46 (0.36-0.56)
5. Single Site weighted Summation	0.58 (0.44-0.71)	0.41 (0.34-0.48)
RAMRIS Erosion	0.35 (0.26-0.45)	0.36 (0.27-0.45)
RAMRIS JSN	0.23 (0.10-0.35)	0.35 (0.30-0.40)
CR Erosion	0.52 (0.44-0.59)	0.19 (0.10-0.28)
CR JSN	0.35 (0.30-0.40)	0.20 (0.14-0.26)
CR Total	0.55 (0.48-0.61)	0.26 (0.19-0.34)
Abbreviations: RAMRIS: OMERACT RA MRI Scoring System. CR: Conventional Radiography, CR Measures measured by van der Heijde modified Sharp score (hands, wrists and feet).		

FIGURE 1: RELATIVE EFFICIENCY OF COMBINED SCORES, VALIDATION COHORT



Error bars representing 95% CI

Appendix

APPENDIX TABLE 1: BASELINE CHARACTERISTICS

	Derivation Cohort*	Validation Cohort**	p-value	CIMESTRA	OPERA
N	194	195		123	72
Age at baseline, median (IQR)	53 (42, 62)	53 (41, 62)	0.92	53 (42, 62)	54 (41, 62)
Female Sex, n (%)	121 (62.4%)	129 (66.2%)	0.44	84 (68.3%)	45 (62.5%)
Symptom Duration (days), median (IQR)	166.5 (85, 311)	91 (56, 133)	<0.001	98 (77, 147)	66.5 (42, 119)
Pos. anti CCP Status, n (%)	159 (82.0%)	120 (61.5%)	<0.001	73 (59.3%)	47 (65.3%)
TJC28, median (IQR)	6 (2, 11)	10 (6, 16)	<0.001	9 (5, 16)	10.5 (6.5, 16.5)
SJC28, median (IQR)	5 (3, 10)	8 (4, 12)	<0.001	9 (4, 12)	7 (4, 12)
Patient global, median (IQR)	48 (29, 67)	54 (31, 74)	0.025	51 (29, 72)	57.5 (41.5, 77)
Physician global, median (IQR)	35 (23, 50)	57 (41, 68)	<0.001	58 (41, 68)	55 (41.5, 69)
Joint pain VAS, median (IQR)	45 (27, 65)	50 (31, 70)	0.074	48 (28, 70)	52 (37, 73)
CRP, median (IQR)	7 (3, 18)	18.3 (8, 40)	<0.001	19.6 (9, 41)	15 (7, 38)
DAS28, median (IQR)	4.4 (3.6, 5.0)	5.3 (4.6, 5.9)	<0.001	5.2 (4.5, 5.9)	5.45 (4.8, 6)

*Derivation cohort: ARCTIC. **Validation cohort: CIMESTRA+OPERA. P-value: Probability of no difference between cohorts, tested by Wilcoxon rank-*sum test for median values, and by Pearson's chi*-square for proportions.
Abbreviations: Anti CCP: anti-cyclic citrullinated peptide. CRP: C-Reactive Protein (mg/L). DAS: Disease Activity Score (28-joints, range 0-10). SJC: Swollen Joint Count (28 joints). TJC: Tender Joint Count (28 joints). VAS: Visual Analog Scale (0-100).

APPENDIX TABLE 2A: BASELINE CONVENTIONAL RADIOGRAPHY (CR) AND MRI SCORES, MEDIAN VALUES (IQR)

	Derivation Cohort*	Validation Cohort**	p-value	CIMESTRA	OPERA
N	194	195		123	72
RAMRIS Synovitis	5 (3, 9)	8 (5.5, 11)	<0.001	8 (5, 11)	9 (6, 12)
RAMRIS Tenosynovitis	4 (2, 10)	4.5 (2, 11)	0.92	3 (1, 8)	9 (4, 14)
RAMRIS Bone marrow edema	1 (0, 5)	0 (0, 2)	<0.001	0 (0, 2)	0 (0, 3)
RAMRIS Erosion	1 (0, 3)	2 (0, 3)	0.98	1 (0, 3)	2 (1, 4)
RAMRIS JSN	15 (10, 20)	0 (0, 0)	<0.001	0 (0, 0)	0 (0, 0)
CR Erosion	3 (1, 5)	1 (0, 4)	<0.001	1 (0, 6)	1 (0, 2)
CR JSN	1 (0, 3.5)	0 (0, 3)	0.01	0 (0, 3)	2 (0, 3.5)
CR Total	4.5 (1.5, 9)	3 (0, 7)	<0.001	3 (0, 8)	3 (1, 6)

APPENDIX TABLE 2B: 1 YEAR CHANGE IN CR AND MRI SCORES, MEDIAN VALUES (IQR)

	Derivation Cohort*	Validation Cohort**	p-value	CIMESTRA	OPERA
N	194	195		123	72
RAMRIS Synovitis	-2 (-5, 0)	-2 (-5, 0)	0.82	-2.5 (-5, 0)	-2.5 (-4, 0.5)
RAMRIS Tenosynovitis	-3 (-7, 0)	-3 (-8, 0)	0.97	-2 (-6, 0)	-5 (-10, -1)
RAMRIS Bone marrow edema	0 (-2, 0)	0 (-1, 0)	0.36	0 (-1, 0)	0 (-2, 0)
RAMRIS Erosion	0 (0, 1)	0 (0, 1)	0.32	0 (0, 1)	0 (0, 1)
RAMRIS JSN	0 (0, 1)	0 (0, 0)	0.011	0 (0, 0)	0 (0, 0)
CR Erosion	0.5 (0, 1)	0 (0, 0)	<0.001	0 (0, 0)	0 (0, 1)
CR JSN	0 (0, 0)	0 (0, 0)	0.003	0 (0, 0)	0 (0, 0)
CR Total	0.5 (0, 1.5)	0 (0, 1)	<0.001	0 (0, 0)	0 (-0.5, 1)

*Derivation cohort: ARCTIC. **Validation cohort: CIMESTRA+OPERA. P-value: Probability of no difference between cohorts, tested by Wilcoxon rank-sum test.

APPENDIX TABLE 3: RELIABILITY OF MRI READINGS IN ARCTIC.
INTRA-CLASS CORRELATIONS, TWO-WAY MIXED-EFFECTS MODEL, ABSOLUTE AGREEMENT

Inter-reader	Baseline	1 year	Change
Synovitis	0.98	0.96	0.87
Tenosynovitis	0.97	0.94	0.90
Bone marrow edema	0.96	0.95	0.94
Erosion	0.92	0.97	0.94
JSN	0.60	0.43	0.61
Intra-reader	**Baseline**	**1 year**	**Change**
Synovitis	0.95	0.98	0.88
Tenosynovitis	0.94	0.97	0.88
Bone marrow edema	0.89	0.97	0.86
Erosion	0.98	0.95	0.81
JSN	0.93	0.92	0.95

Inter-reader comparisons: 12 patients scored separately by D.Glinatsi and U.Sundin. Intra-reader comparisons: 12 patients scored on separate occasions by U.Sundin.

APPENDIX BOX 1: ANATOMICAL ZONES FOR APPROACH 3-4, AND WEIGHTS APPLIED BY APPROACH 4.

Synovitis	**Zone 1**	**Zone 2**	**Zone 3**	**Total**
Joints	Radioulnar Radiocarpal	CMC/IC	MCP 2-5	
RAMRIS max. score	6	3	12	21
Approach 4 weight	2	1	1	
Approach 4 max. score	12	3	12	27

Tenosynovitis	**Zone 1**	**Zone 2**	**Zone 3**	**Total**
Tendons	Wrist extensor compartment 1-6	Wrist flexor compartment 1-3	Flexor tendon sheaths 2-5 at MCP-level	
RAMRIS max. score	18	9	12	39
Approach 4 weight	4	4	3	
Approach 4 max. score	72	36	36	144

Bone marrow edema	**Zone 1**	**Zone 2**	**Zone 3**	**Total**
Ossicles	Radius Ulna Scaphoid Lunate Pisiform Triquetrum	Trapezium Trapezoid Capitate Hamate Proximal metacarpals 1-5	Distal metacarpals 2-5 Proximal phalanges 2-5	
RAMRIS max. score	18	27	24	69
Approach 4 weight	1	1	1	
Approach 4 max. score	18	27	24	69

Erosions	**Zone 1**	**Zone 2**	**Zone 3**	**Total**
Ossicles	Radius Ulna Scaphoid Lunate Pisiform Triquetrum	Trapezium Trapezoid Capitate Hamate Proximal metacarpals 1-5	Distal metacarpals 2-5 Proximal phalanges 2-5	
RAMRIS max. score	60	90	80	230
Approach 4 weight	3	1	3	
Approach 4 max. score	180	90	240	510

Joint Space Narrowing	**Zone 1**	**Zone 2**	**Zone 3**	**Total**
Joints	Radio-Scaphoid Radio-Lunate Scapho-Lunate Lunato-Triquetral	Trapezium-Scaphoid Trapezoid-Scaphoid Trapezium-Trapezoid Trapezoid-Capitate Capitate-Scaphoid Capitate-Lunate Capitate-Hamate Hamato-Triquetral CMC 1-5	MCP 2-5	
RAMRIS max. score	16	52	16	84
Approach 4 weight	2	2	4	
Approach 4 max. score	32	104	64	200

Weights for approach 4 were obtained by ranking and rescaling the data-derived weights from approach 3 to whole numbers (range 1-4). Adjustment of +/- 1 step was allowed to optimize performance.