

Computer-Assisted Image Analysis in Assessment of Peripheral Joint MRI in Inflammatory Arthritis: A Systematic Review and Meta-analysis

Arya Haj-Mirzaian,¹ Olga Kubassova,² Mikael Boesen,³ John Carrino,⁴ and Paul Bird⁵ 

Objective. To summarize the feasibility of computer-assisted quantification of joint pathologies on magnetic resonance imaging (MRI) in patients with inflammatory arthritis by evaluating the published data on reliability, validity, and feasibility.

Methods. A systematic literature search was performed for original articles published from January 1, 1985, to January 1, 2021. We selected studies in which patients with inflammatory arthritis were enrolled, and arthritis-related structural damage/synovitis in peripheral joints was assessed on non-contrast-enhanced, contrast-enhanced (CE), or dynamic CE (DCE)-MRI using (semi)automated methods. Data were pooled using random-effects model.

Results. Twenty-eight studies consisting of 1342 MRIs were included (mean age, 54.8 years; 66.7% female; duration of arthritis, 3.6 years). Among clinical/laboratory factors, synovial membrane volume (SV) was moderately correlated with erythrocyte sedimentation rate (ESR) level ($P < 0.01$). Pooled analysis showed an overall excellent intra- and inter-reader reliability for computer-aided quantification of bone erosion volume (BEV; $r = 0.97$ [95% CI: 0.92–0.99], 0.93 [0.87–0.97]), SV ($r = 0.98$ [95% CI: 0.90–0.99], 0.86 [0.78–0.91]), and DCE-MRI perfusion parameters ($r = 0.96$ –0.99). Meta-regression showed that computer-aided and manual methods provide comparable reliability ($P > 0.05$). Computer-aided measurement of BEV ($r = 0.92$), SV ($r = 0.82$), and DCE-MRI biomarkers ($r = 0.72$ N-total; $r = 0.74$ N-plateau; $r = 0.64$ N-washout) were significantly correlated with the Rheumatoid Arthritis Magnetic Resonance Imaging Score (RAMRIS; $P < 0.01$), allowing for earlier assessment of drug efficacy. On average, (semi) automated analysis of BEV/SV took 17 minutes (vs. 9 minutes for the RAMRIS) and DCE-MRI took 4 minutes (vs. 33 minutes for manual assessment).

Conclusion. Computer-aided image quantification technologies demonstrate excellent reliability and validity when used to quantify MRI pathologies of peripheral joints in patients with inflammatory arthritis. Computer-aided evaluation of inflammatory arthritis is an emerging field and should be considered as a viable complement to conventional observer-based scoring methods for clinical trials application.

INTRODUCTION

Early diagnosis of inflammatory arthritis and accurate monitoring of disease progression are central for optimizing management of patients with rheumatoid arthritis (RA) and other types of arthritis (1,2). The pivotal role of advanced imaging modalities, especially magnetic resonance imaging (MRI) and dynamic contrast-enhanced (DCE) MRI, in the diagnosis of early disease and monitoring response to therapy has been confirmed by

numerous studies (3–5). MRI provides precise measurements for bony erosions, bone marrow edema (BME), and joint synovitis, which are considered to be predictive biomarkers for long-term clinical outcomes (5–7). One example of quantitative scoring is the Rheumatoid Arthritis Magnetic Resonance Imaging Score (RAMRIS) system, which has been developed by the Outcome Measures in Rheumatology (OMERACT) group for quantification of bone erosion, BME, and synovitis on MRI (8). The RAMRIS is well established, but the responsiveness can be limited by the

¹Arya Haj-Mirzaian, MD, MPH: Johns Hopkins Hospital, Baltimore, Maryland; ²Olga Kubassova, MD, PhD: Image Analysis Group, London, United Kingdom; ³Mikael Boesen, MD, PhD: University Hospital Bispebjerg and Frederiksberg; The Parker Institute, Copenhagen, Denmark; ⁴John Carrino, MD, MPH: Hospital for Special Surgery, Hackensack, New Jersey; ⁵Paul Bird, MD, PhD: University of New South Wales, Sydney, New South Wales, Australia.

Address correspondence to Paul Bird, MD, PhD, University of New South Wales, 4/19 Kensington Street, Kogarah NSW, Sydney 2217 Australia. Email: p.bird@unsw.edu.au.

Submitted for publication November 21, 2021; accepted in revised form April 15, 2022.

discreet nature of the score. As a complement measure, DCE-MRI allows comprehensive assessment of perfusion changes on a continuous scale and accurate visualization of the distribution and the degree of histological synovial inflammation (9,10). DCE-MRI data are routinely analyzed using the region of interest (ROI) method. A small ROI is manually placed by a reader in the area with maximal visual enhancement, and perfusion parameters are calculated within the predefined ROI (11).

Despite the promising results regarding the diagnostic/prognostic value of MRI-based biomarkers, the clinical utility of user-dependent image assessment is limited because of several factors. Semiquantitative MRI scoring systems can be insensitive for the identification of subtle changes in early arthritis (8,12). Traditional user-dependent assessment of DCE-MRI can be time-consuming and challenging as the reliability of the ROI method is dependent on the expertise of the reader (13,14). To overcome the inherent limitations of manual image assessment, several computer-aided methods have been developed. Semiautomatic quantification of bone erosion volume (BEV), BME volume, and synovial membrane volume (SV) has the potential to facilitate the clinical utility of MRI by alleviating the time and cost burden of manual scoring. In addition, semiautomatic quantification of DCE-MRI allows user-independent extraction of perfusion parameters by model-based approaches, such as pharmacokinetic and heuristic analysis, in a short amount of time on the entire image (9,10); these methods have the advantage of better reproducibility and time efficiency.

Several semiautomatic methods have been proposed for the assessment of peripheral joint MRI in inflammatory arthritis. However, the overall feasibility, reliability, and validity of these computer-aided image analyses have not been comprehensively assessed.

Our objective was to perform a research synthesis using systematic review and meta-analysis to summarize the published data on the computer-aided image quantification technologies for evaluating MRI features of inflammatory arthritis in peripheral joints with a focus on reliability, validity, and feasibility. These parameters were assessed using the OMERACT filter definitions (15). Feasibility in the OMERACT filter encompasses the practical considerations of using an instrument, including its ease of use, time to complete, monetary costs, and interpretability of the question(s) included in the instrument. In this analysis, feasibility was restricted to time taken for image assessment. Other measures of feasibility were not available in the articles.

MATERIALS AND METHODS

The study design was a research synthesis of the published literature. This systematic review and meta-analysis was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (16).

Literature search. A systematic literature search was performed on Medline, EMBASE, and Cochrane databases for original English-language articles published from January 01, 1985, to January 01, 2021. Various combinations of search terms representing “inflammatory arthritis,” “MRI,” “computer-aided methods,” and “peripheral joints” were used to retrieve all relevant studies; details of searching strategy and keywords are summarized in Supplementary Table 1. Search terms were developed with the assistance of a professional health informaticist and were approved by all authors. The reference lists of included articles were manually searched to identify any missing articles. All references were imported to the Covidence online platform, and all parts of the study selection and data extraction were conducted using the Covidence tool (<https://www.covidence.org/reviews/64844>).

Study selection. Studies were included if they met the following inclusion criteria: 1) enrolled patients with any type of inflammatory arthritis; 2) measured arthritis-related structural damage/inflammation in peripheral joints using non-CE, CE MRI, and/or DCE-MRI; and 3) used computer-aided image analysis. We excluded review articles, conference abstracts, case reports, editorials, letters, theses, methodology articles, studies performed on animals/cadavers, studies focused on other joints than peripheral joints, and studies focused on other imaging modalities than MRI. One reviewer (Arya Haj-Mirzaian, a radiology resident with 7 years of research experience) screened titles and abstracts of all retrieved citations. The full text of potentially eligible articles that passed title/abstract screening were obtained and assessed by the same reviewer in duplicate, and the final eligible studies were selected.

Data extraction. The following data were extracted from all included studies by the same reviewer: study design, patient enrollment, year of publication, number of patients, demographics, imaging modality, joint regions (hand, wrist, metacarpophalangeal [MCP], or knee joints), (semi)automated measures (ie, SV, BEV, and DCE-MRI parameters—maximum enhancement [ME], initial rate enhancement [IRE], time of onset of enhancement [Tonset], Gadolinium counts [total, persistent, plateau, and wash-out], and enhancement pattern), results of manual methods (ie, the RAMRIS for bone erosions and synovitis), time needed to perform image assessment, histopathological/laboratory findings, and clinical outcomes (eg, disease activity score [DAS]-28).

Finally, intrareader reliability (ie, degree of agreement among repeated measurements by a single reader) and inter-reader reliability (ie, degree of agreement among different readers in case of semiautomated methods) for each computer-aided method, the correlation measures between computer-aided and manual MRI-based measurements, and the correlation measures between imaging biomarkers and clinical/histopathological/laboratory outcomes were extracted.

Quality assessment. The quality of each included study was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2 (18). Risk of bias (ROB) and applicability of study was assessed in the following four domains: patient selection, index test (ie, computer-aided measurements), references standard (ie, established MRI scores and/or clinical outcomes), and flow and timing. The questions of the QUADAS-2 tool have been modified based on our research aims; details are summarized in Supplementary Table 2. Studies were rated to have low, unclear, or high ROB/applicability by combining the results of all domains (18).

Statistical analysis. Analyses were performed using Comprehensive Meta-Analysis version 3 (Biostat). Forest plots and inconsistency index with Q test and I^2 statistics were used to estimate the degree of between-study heterogeneity (19,20). High statistical heterogeneity was defined as $I^2 > 75%$ (19–21). Funnel plots were reconstructed for visual assessment of publication bias. Plot asymmetry was quantified using the Begg and Mazumdar rank correlation test (21). The significance level was set at P value < 0.05 . Results were adjusted using the trim and fill method, when the Begg and Mazumdar test suggested the possibility of publication bias (22).

The following correlation coefficients were transformed to normally distributed z values and pooled using a random-effects model: 1) intraclass correlation co-efficient (ICC) for intra- and inter-reader reliability of each computer-aided measurement; 2) correlation coefficients between computer-aided measures and user-dependent RAMRIS findings; and 3) correlation coefficients between computer-aided measurements and clinical/histopathological/laboratory findings. The weighted pooled correlation coefficients, along with 95% confidence interval (95% CI), were calculated and reported. Results were presented narratively when it was not feasible to perform meta-analysis due to the limited data (ie, fewer than three studies). The impact of covariates on outcome measures was evaluated using subgroup analysis and/or univariate meta-regression.

RESULTS

Study selection

Figure 1 illustrated the PRISMA flow chart summarizing the study selection process. A total of 311 published records were identified through the initial literature search. After title and abstract screening, 72 articles were considered eligible to be assessed at the level of full text. Four articles were added through manual search of bibliography; three of them have been excluded at the level of title/abstract screening. From these records, 44 were excluded for the following reasons: wrong outcome ($n = 23$), majority of these studies have not reported our outcomes of interest; wrong study design ($n = 18$), not using MRI or

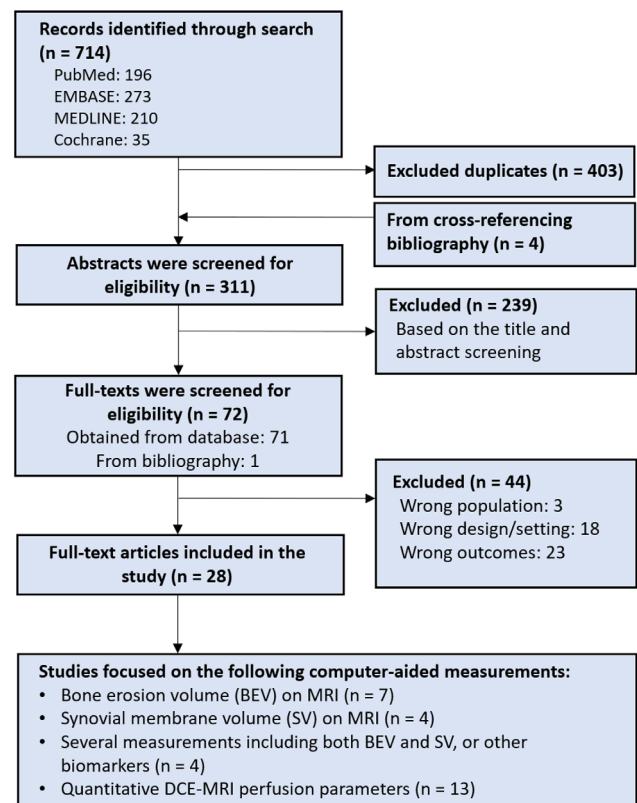


Figure 1. PRISMA flowchart diagram for study selection. BEV, bone erosion volume; DCE, dynamic contrast-enhanced; MRI, magnetic resonance imaging; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analysis; SV, synovial membrane volume.

computer-aided image analysis; and wrong patient population ($n = 3$), not including patients with inflammatory arthritis. The list of excluded studies and the main reason of exclusion are summarized in Supplementary Table 3. Ultimately, 28 articles including 1342 MRI examinations were deemed eligible for inclusion.

Study characteristics

The main characteristics of 28 included studies are summarized in Table 1 and Supplementary Table 4. Twenty-two of the studies were cross sectional; six were longitudinal studies. Most studies examined patients with RA, although a few studies also included patients with psoriatic arthritis and undifferentiated arthritis. Hand, wrist, MCP, and knee joints were studied in 5, 19, 8, and 3 studies, respectively (some studies assessed more than one region). Eleven studies used 1.5 T magnetic resonance (MR) scanner, whereas six studies used 3 T MR scanner. Sixteen studies reported findings of non-CE and/or CE MRI, and 15 studies reported results of DCE-MRI (some studies used both methods). MRI slice thickness varies across included studies, ranging from 0.7 to 5 mm. Among included studies, 19 used semiautomated methods, 9 used fully automated methods, and

Table 1. General characteristics of the included studies in meta-analysis

Study	Publication year	Joint region(s)	Main outcome (semi) automated measures	No. of subjects	Female (%)	Age (y, mean)	Duration of disease (y, mean)	Imaging (T and slice thickness)	Manual measurement(s)	(Semi)automated measurement(s)
1. Bird et al (1)	2002	MCP	Bone erosion volume	5	N/A	N/A	N/A	MRI (1.5T; 1 and 3mm)	OMERACT MRI scoring	OSIRIS software, semi-automated
2. Poh et al (2)	2019	MCP	Bone erosion volume	32	78.1	50.5	4.0	MRI (1.5T; 1 mm)	OMERACT MRI scoring	OSIRIS software, semi-automated
3. Døhn et al (3)	2007	MCP	Bone erosion volume	17	76.5	52.0	N/A	MRI (0.6T; N/A)	OMERACT MRI scoring	OSIRIS software, semi-automated
4. Bird et al (4)	2003	Hand and Wrist	Bone erosion volume, Synovial volume	12	50	56.0	6.0	MRI (1.5T; 3-4 mm)	OMERACT MRI scoring	OSIRIS software, semi-automated
5. Yang et al (5)	2015	Hand and Wrist	Bone erosion volume, Synovial volume, BME volume	16	81	52.9	7.6	MRI and DCE-MRI (3T; 1 mm)	OMERACT MRI scoring and manual volume	Semi-automated
6. Emond et al (6)	2012	MCP	Bone erosion volume	40	72	42 to 81	N/A	MRI (1T; 1 mm)	Manual volume	Semi-automated
7. Tomizza et al (7)	2015	MCP	Bone erosion volume	100	70.6	57.4	4.8	MRI (1T; 1 mm)	No	Semi-automated
8. Bird et al (8)	2005	Wrist and MCP	Bone erosion volume	18	N/A	N/A	N/A	MRI (1.5T; 3 mm)	OMERACT MRI scoring	OSIRIS software, semi-automated
9. Chand et al (9)	2011	Wrist	Synovial volume	38	63.2	57.1	6.7	MRI (3T; 2 mm)	OMERACT MRI scoring	OSIRIS software, semi-automated
10. Aizenberg et al (10)	2017	Wrist	BME volume	485	61.2	54.7	<2	MRI (1.5T; 3 mm)	OMERACT MRI scoring	Automated
11. Crowley et al (11)	2011	Wrist	Bone erosion volume, BME volume	22	68.2	52	<2	MRI (3T; 1.5 - 2.1 mm)	OMERACT MRI scoring	OSIRIS software, semi-automated
12. Czaplicka et al (12)	2015	Wrist	Synovial volume	32	87.5	47	<5	CE-MRI and DCE-MRI (0.2T; 0.7 mm)	OMERACT MRI scoring	Automated
13. Klarlund et al (13)	1999	MCP	Synovial volume	42	N/A	55	8	CE-MRI (1T; 3 mm)	Manual semi-quantitative score	XPrime software, Automated
14. Li et al (14)	2012	Wrist	BME volume and BME perfusion parameters	14	78.6	53.7	N/A	MRI and DCE-MRI (3T; 2 mm)	No	Automated
15. Stramare et al (15)	2014	Hand and Wrist	Synovial volume	16	87.5	50	17	CE-MRI (1.5T; 0.7 - 3 mm)	OMERACT RAMRIS and SAMIS scoring	Automated
16. Boesen et al (16)	2011	Wrist	Synovitis perfusion parameters	46	76.1	56	5.5	DCE-MRI (0.2T; 5 mm)	Manual measurement (REE and RE)	Dynamika RA Automated

(Continued)

Table 1. (Cont'd)

Study	Publication year	Joint region(s)	Main outcome (semi) automated measures	No. of subjects	Female (%)	Age (y, mean)	Duration of disease (y, mean)	Imaging (T and slice thickness)	Manual measurement(s)	(Semi)automated measurement(s)
17. Cirimino et al (17)	2012	Wrist	Synovitis perfusion parameters	17	76.5	51.6	N/A	DCE-MRI (0.2T; 5 mm)	No	Dynamika software, semi-automated
18. Cirimino et al (18)	2014	Wrist	Synovitis perfusion parameters	10	60	52.9	6.3	DCE-MRI (0.2T; 5 mm)	OMERACT RAMRIS scoring; Manual measurement (conventional ROI method; REE and RE)	Dynamika software, semi-automated
19. Orguc et al (19)	2013	Wrist and MCP	Synovitis perfusion parameters	40	67.5	51.1	2.4	Non-CE and DCE-MRI (1.5T; 2.5 mm)	OMERACT RAMRIS scoring	Automated
20. van der Leij et al (20)	2010	Knee	Synovitis perfusion parameters (TIC shapes)	10	50%	39	N/A	DCE-MRI (1.5T; 4 mm)	Manual measurement	Semi-automated
21. Boesen et al (21)	2012	Wrist	Synovitis and BME perfusion parameters	54	N/A	52	11	DCE-MRI (0.2T; 3 mm)	OMERACT RAMRIS scoring	Dynamika software, (semi) automated
22. Wojciechowski et al (22)	2013	Wrist	Synovitis perfusion parameters	59	N/A	47	N/A	DCE-MRI (0.2T; 3.5 mm)	OMERACT RAMRIS scoring	Dynamika software, (semi) automated
23. Axelsen et al (23)	2013	Knee	Synovitis perfusion parameters	12	N/A	70	5	DCE-MRI (1.5T; 5 mm)	No	Dynamika software, (semi) automated
24. Axelsen et al (24)	2012	Knee	Synovitis perfusion parameters	17	82.4	64	8	DCE-MRI (1.5T; 5 mm)	No	Dynamika software, (semi) automated
25. Zierhut et al (25)	2007	Wrist	Synovitis perfusion parameters (kinetic parameters)	12	58	43	N/A	DCE-MRI (1.5T; 3 mm)	Manual measurement	Kinetic parameters, automated
26. Meier et al (26)	2014	Wrist and Hand	Synovitis perfusion parameters	28	53.6	53	1	DCE-MRI (3T; N/A)	No	Automated
27. Kubassova et al (27)	2010	Wrist and Hand	Synovitis perfusion parameters	140	N/A	62.7	5	DCE-MRI (0.2T; 5 mm)	No	Dynamika software, (semi) automated
28. Sakashita et al (28)	2015	Wrist and Hand	Synovitis perfusion (TIC shapes)	8	87.5	57	<1	DCE-MRI (3T; 2 mm)	OMERACT RAMRIS scoring	Semi-automated

Abbreviations: BME, bone marrow edema; DCE, dynamic contrast-enhanced; MCP, metacarpophalangeal; MRI, magnetic resonance imaging; N/A, not applicable; OMERACT, outcome measurement in rheumatology clinical trials; RAMRIS, rheumatoid arthritis magnetic resonance imaging score; TIC, time intensity curve.

21 reported results of manual MRI-based assessments. The average number of enrolled patients in each study was 48 (range = 5-485). The mean age of participants was 54.8 years, with a female proportion of 66.7% and a mean duration of disease of 3.6 years.

Nine studies reported the reliability and/or validity of computer-aided measurement for BEV on MRI, 6 studies focused on SV, and 14 studies focused on DCE-MRI parameters. Among studies on DCE-MRI, most studies reported quantitative DCE-MRI perfusion markers and/or time intensity curve (TIC) shape categories using heuristic analysis rather than pharmacokinetic analysis. The correlation of (semi)automated measurements with manual measurements and clinical outcomes were reported in 11 and 12 studies, respectively. The time needed to perform measurement was reported in nine studies. Only four studies reported results of computer-aided BME volume measurement; meta-analysis was not performed on BME because of lack of data (24-27).

Quality assessment

Figure 2 summarizes the details of quality assessment using the QUADAS-2 tool. The overall low-moderate ROB was considered for the body of included literature. The overall ROB in the patient selection domain was unclear in most studies because they did not report details of patient recruitment (eg, consecutive vs. nonconsecutive). Regarding the index test domain, 10 studies had unclear ROB and 1 study had a high ROB; the main source of bias was unclear blindness of observer. The reference standard domain was judged to be high/unclear risk in two studies; in these studies, radiologists performing manual measurements were not blinded to (semi)automated measurements and/or clinical findings.

Meta-analysis

BEV. Intra- and inter-reader reliability. Nine studies (254 MRIs) reported the intra- and/or inter-reader reliability of computer-aided measurement of BEV (Supplementary Figures 1 and 2; 24,26,28-34). The weighted pooled ICC for intra- and inter-reader reliability was 0.97 (95% CI: 0.92-0.99; $P < 0.001$; $I^2 = 91.3\%$) and 0.93 (95% CI: 0.87-0.97; $P < 0.001$, $I^2 = 79.1\%$), respectively (Table 2, 3). Visual evaluation of funnel plot revealed a possibility of publication bias (Supplementary Figures 1 and 2); however, Begg and Mazumdar rank correlation test showed no evidence of publication bias ($P = 0.711$ and 0.266 , respectively). Pooled analyses were adjusted for publication bias using the trim and fill method, which showed equal ICC estimates (Table 2; $P > 0.05$). Restricting analysis to studies with prospective design ($P = N/A$, $P = 0.990$) and low ROB ($P = 0.687$, $P = 0.956$) resulted in almost the same ICC values for intra- and inter-reader reliability, respectively. Meta-regression analysis showed a significant negative association between the field strength of the MR scanner (tesla) and intrareader (Beta-coefficient \pm SEM = -0.726 ± 0.176 ;

$P < 0.001$) and inter-reader (Beta-coefficient \pm SEM = -0.384 ± 0.156 ; $P = 0.014$) reliability of (semi)automated BEV measurements (Figure 3). Furthermore, results of meta-regression showed that (semi)automated BEV calculation on MRI has higher inter-reader reliability in older patients (Beta-coefficient \pm SEM = 0.114 ± 0.050 ; $P = 0.023$). No other significant association was detected between reliability estimates and the following factors: year of publication ($P = 0.349$; $P = 0.696$), mean age ($P = 0.515$; significant for inter-reader, as noted earlier), female proportion ($P = 0.776$; $P = 0.816$), mean duration of disease ($P = 0.984$; $P = 0.269$), and slice thickness ($P = 0.396$; $P = 0.920$).

Comparing with manual MRI-based measurements. Seven studies (144 MRIs) reported the correlation coefficient of computer-aided BEV measurements with the RAMRIS and/or manual BEV measurements (Supplementary Figure 3; 24,26,28-32).

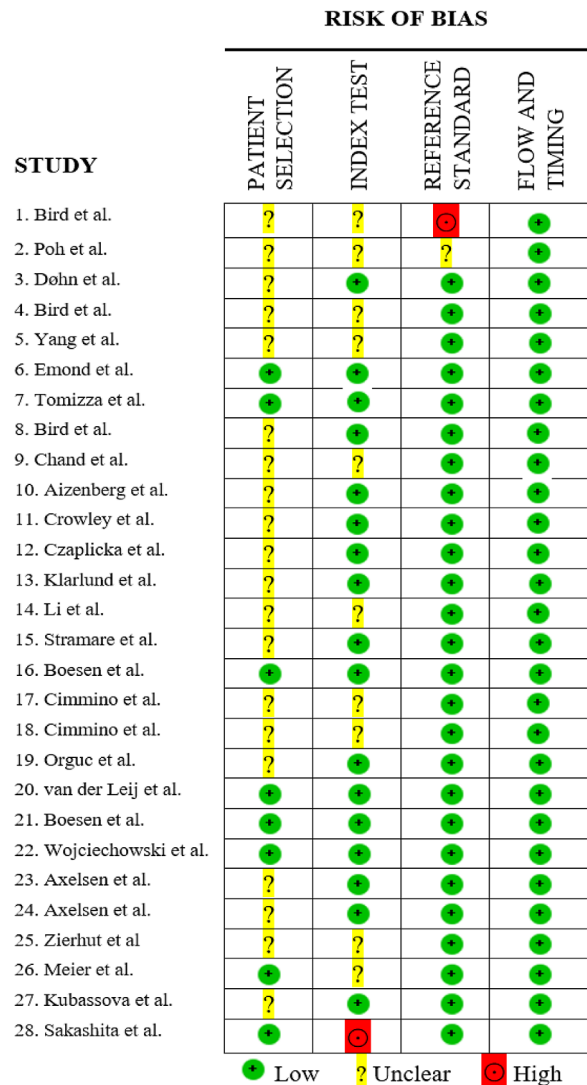


Figure 2. Risk of bias and quality assessment using the QUADAS-2 tool. No concern was detected regarding the applicability of patient selection, index test, and reference standard. QUADAS, Quality Assessment of Diagnostic Accuracy Studies.

Meta-analysis showed an excellent level of agreement between these methods with the weighted pooled ICC of 0.92 (95% CI: 0.80-0.97; $P < 0.001$; $I^2 = 84.1\%$; Table 2). No evidence of publication bias was detected (Supplementary Figure 3; $P = 0.764$); there was an equal adjusted correlation coefficient (Table 2; $P > 0.05$). Subgroup and meta-regression analyses showed no factor has a significant impact on the following pooled correlation values: study design ($P = 0.178$), ROB ($P = 0.178$), year of publication ($P = 0.057$), mean age ($P = 0.649$), female proportion ($P = 0.931$), mean duration of disease ($P = 0.558$), MRI tesla ($P = 0.623$), and slice thickness ($P = 0.185$).

Five studies (108 MRIs) reported intra- and inter-reader reliability of the RAMRIS and/or manual BEV measurements (24,26,31,32,34). Results showed an overall comparable intrareader reliability ($P = 0.626$; weighted pooled ICC of 0.96 [95% CI: 84-0.99] for manual measurements) and inter-reader reliability ($P = 0.273$; weighted pooled ICC of 0.89 [95% CI: 0.85-0.93] for manual measurements) between these two methods.

Correlation with clinical variables. Two studies (44 MRIs) displayed the correlation coefficient between the computer-aided BEV measurements and DAS-28 (29,31). The overall poor correlation was observed between BEV and DAS-28 (range of reported correlation coefficient = -0.10 to 0.94). Further analysis on other

clinical/histopathological/laboratory findings was not possible because of the limited number of studies.

Time needed to perform measurements. Time taken to perform (semi)automated BEV measurements was estimated to be 14 minutes (range = 2.6-21.5) based on the results of five studies (105 MRIs; 24,28,29,31,32), which was comparable to the time needed to perform the RAMRIS (10 minutes [range = 5-12]).

SV. Intra- and inter-reader reliability. Intra- and inter-reader reliability of MRI-based (semi)automated SV measurement was reported in four studies (108 MRIs; Supplementary Figures 4 and 5; 24,31,35,36). The weighted pooled ICC for intrareader reliability was 0.98 (95% CI: 0.90-0.99; $P < 0.001$, $I^2 = 93.4\%$) and for inter-reader reliability was 0.86 (95% CI: 0.78-0.91; $P < 0.001$; $I^2 = 27.8\%$; Table 2). Symmetrical distribution of studies was observed in the funnel plot (Supplementary Figures 4 and 5); Begg and Mazumdar test also suggested no risk of publication bias ($P > 0.05$). Meta-regression analyses showed a higher intrareader (Beta-coefficient \pm SEM = -0.955 ± 0.144 ; $P < 0.001$) and inter-reader (Beta-coefficient \pm SEM = -0.256 ± 0.144 ; $P = 0.076$; statistically approached significant) reliability for (semi)automated SV measurements in early stages of

Table 2. Reliability, correlation with clinical variables and established MRI scores, duration

Study	Intra-reader reliability	Inter-reader reliability	Correlation with clinical variables	Correlation with established manual MRI-based scores	Time of measures
1. Bird et al (1)	Yes	Yes	No	Yes (RAMRIS)	Yes
2. Poh et al (2)	Yes	Yes	Yes	Yes (RAMRIS)	Yes
3. Døhn et al (3)	Yes	No	No	Yes (RAMRIS)	No
4. Bird et al (4)	Yes	Yes	Yes	Yes (RAMRIS)	Yes
5. Yang et al (5)	Yes	Yes	No	Yes (RAMRIS/manual)	Yes
6. Emond et al (6)	Yes	Yes	No	Yes (manual)	Yes
7. Tomizza et al (7)	Yes	Yes	No	No	No
8. Bird et al (8)	No	Yes	No	No	No
9. Chand et al (9)	Yes	Yes	Yes	Yes (RAMRIS)	Yes
10. Aizenberg et al (10)	No	No	No	Yes (RAMRIS)	No
11. Crowley et al (11)	Yes	Yes	No	Yes (RAMRIS)	No
12. Czaplicka et al (12)	No	No	No	Yes (RAMRIS/manual)	No
13. Klarlund et al (13)	Yes	Yes	Yes	Yes (manual)	No
14. Li et al (14)	Yes	Yes	Yes	No	No
15. Stramare et al (15)	No	No	Yes	Yes	No
16. Boesen et al (16)	Yes	No	No	Yes	No
17. Cimmino et al (17)	No	No	Yes	No	No
18. Cimmino et al (18)	Yes	Yes	Yes	No	No
19. Orguc et al (19)	No	No	No	Yes	No
20. van der Leij et al (20)	Yes	No	No	No	No
21. Boesen et al (21)	No	No	Yes (no data)	Yes	Yes
22. Wojciechowski et al (22)	No	No	No	Yes	No
23. Axelsen et al (23)	Yes	Yes	No	No	No
24. Axelsen et al (24)	Yes	Yes	Yes	No	No
25. Zierhut et al (25)	No	No	No	Yes (manual)	No
26. Meier et al (26)	No	No	Yes	No	No
27. Kubassova et al (27)	No	No	No	No	Yes
28. Sakashita et al (28)	No	No	No	No	Yes

Abbreviations: MRI, magnetic resonance imaging; RAMRIS, rheumatoid arthritis magnetic resonance imaging score.

Table 3. Results of meta-analysis: Pooled ICC values for intra- and inter-reader reliability of each computer-aided measurement, correlation values between computer-aided measurements and manual scores/clinical findings, and time needed to perform each image analysis method

(Semi)automated measurements	DCE-MRI perfusion parameters						
	BEV	SV	ME	IRE	N-total	N-plateau	N-washout
Intra-reader reliability							
Pooled ICC (95% CI), I ² value	0.97 (0.92-0.99), 91%	0.98 (0.90-1.0), 93%	0.99 (0.82-1.0), 97%	0.99 (0.94-1.0), 97%	0.96 (0.85-0.99), 80%	N/A	N/A
Adjusted pooled ICC (95% CI)	0.97 (0.92-0.99)	0.92 (0.57-0.99)	0.89 (-0.30 to 0.99)	0.98 (0.73-0.99)	0.90 (0.70-0.97)	N/A	N/A
Number of studies (MRIs)	8 (244)	4 (108)	3 (68)	4 (85)	3 (65)	N/A	N/A
Inter-reader reliability							
Pooled ICC (95% CI), I ² value	0.93 (0.87-0.97), 79%	0.86 (0.78-0.91), 28%	N/A	N/A	N/A	N/A	N/A
Adjusted pooled ICC (95% CI)	0.93 (0.87-0.97)	0.83 (0.74-0.89)	N/A	N/A	N/A	N/A	N/A
Number of studies (MRIs)	8 (237)	4 (108)	N/A	N/A	N/A	N/A	N/A
Correlation with the RAMRIS or manual measurements							
Pooled ICC (95% CI), I ² value	0.92 (0.80-0.97), 84.1%	0.82 (0.74-0.87), 0%	Range = 0.05-0.80	Range = 0.18-0.34	Range = 0.42-0.75	Range = 0.41-0.77	Range = 0.44-0.62
Adjusted pooled ICC (95% CI)	0.92 (0.80-0.97)	0.82 (0.74-0.87)	N/A	N/A	N/A	N/A	N/A
Number of studies (MRIs)	7 (144)	5 (140)	2 (70)	2 (70)	2 (113)	2 (113)	2 (113)
Correlation with clinical variables							
Pooled ICC (95% CI), I ² value	N/A	ESR = 0.53 (0.28-0.72), 22.4%	N/A	DAS-28 Range of P value = 0.003-0.01	DAS-28 Range of P value = 0.003-0.01	N/A	N/A
Adjusted pooled ICC (95% CI)	N/A	ESR = 0.53 (0.28-0.72), 22.4%	N/A	N/A	N/A	N/A	N/A
Number of studies (MRIs)	N/A	ESR = 3 (70)	N/A	2 (27)	2 (27)	N/A	N/A
Time needed to perform measurements							
Pooled mean (range), min	13.97 (2.6-23.5)	15.14 (8.2-20)			4.0 (3.0-8.0)		
Number of studies (MRIs)	5 (105)	3 (66)			2 (194)		

Abbreviations: BEV, bone erosion volume; CI, confidence interval; DAS-28, disease activity score-28; DCE, dynamic contrast-enhanced; ESR, erythrocyte sedimentation rate; ICC, intraclass correlation co-efficient; IRE, initial rate enhancement; ME, maximum enhancement; MRI, magnetic resonance imaging; N/A, not applicable; RAMRIS, Rheumatoid Arthritis Magnetic Resonance Imaging Score; SV, synovial membrane volume.

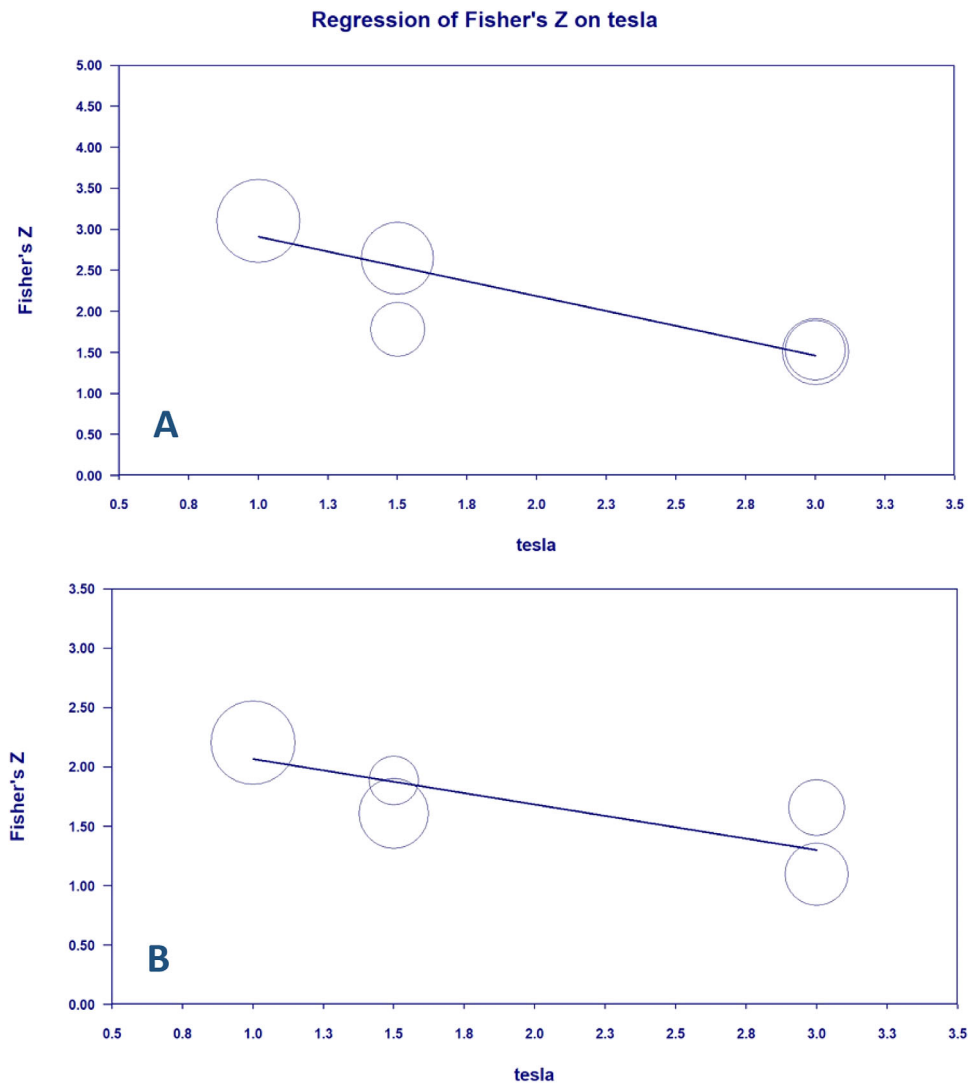


Figure 3. Meta-regression analysis for the association between the field strength of MR scanner (tesla) and intrareader (**A**) and inter-reader (**B**) reliability of computer-aided measurement of bone erosion volume on magnetic resonance imaging.

inflammatory arthritis (Figure 4). No further source of between-study heterogeneity was detected ($P > 0.05$ for all variables).

Comparing with manual MRI-based measurements. The correlation coefficient between computer-aided SV and the RAMRIS or manual measurements was reported in five studies (140 MRIs; Supplementary Figure 6; 24,31,35–37). Pooled analysis demonstrated an overall moderate level of agreement between (semi)automated and manual assessment of SV, with the weighted pooled ICC of 0.82 (95% CI: 0.75–0.87; $P < 0.001$; $I^2 = 0\%$; Table 2). No evidence of publication bias was detected ($P = 1.0$), with the equal adjusted ICC (Table 2; $P > 0.05$). No variable had a significant impact on the pooled correlation estimates ($P > 0.05$ for all variables).

The intra- and inter-reader reliability of the RAMRIS and/or manual assessment of synovitis was reported in five studies (156 MRIs; 24,31,35–37). Pooled analysis showed that intrareader ($P = 0.208$; pooled ICC of 0.94 [95% CI: 0.84–0.98] for

manual measurements) and inter-reader reliability ($P = 0.782$; pooled ICC of 0.85 [95% CI: 0.75–0.91] for manual measurements) of the RAMRIS and/or manual SV measurement was not significantly different from (semi)automated SV measurement.

Correlation with clinical variables. The correlation coefficient of computer-aided SV measurement with erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) levels was reported in three studies (70 MRIs; 31,35,36). (Semi)automated SV measures were moderately correlated with the serum ESR level with the pooled ICC of 0.53 (95% CI: 0.28–0.72, $P < 0.001$, $I^2 = 22.4\%$) (Supplementary Figure 7); no evidence of publication bias was detected ($P = 0.296$; Supplementary Figure 7). No significant correlation was detected between (semi)automated SV measures and CRP level (0.18 [95% CI: –0.15 to 0.47]; $P = 0.290$; $I^2 = 31.0\%$). The ICC values of 0.11 to 0.83 were reported for the correlation between (semi)automated SV and

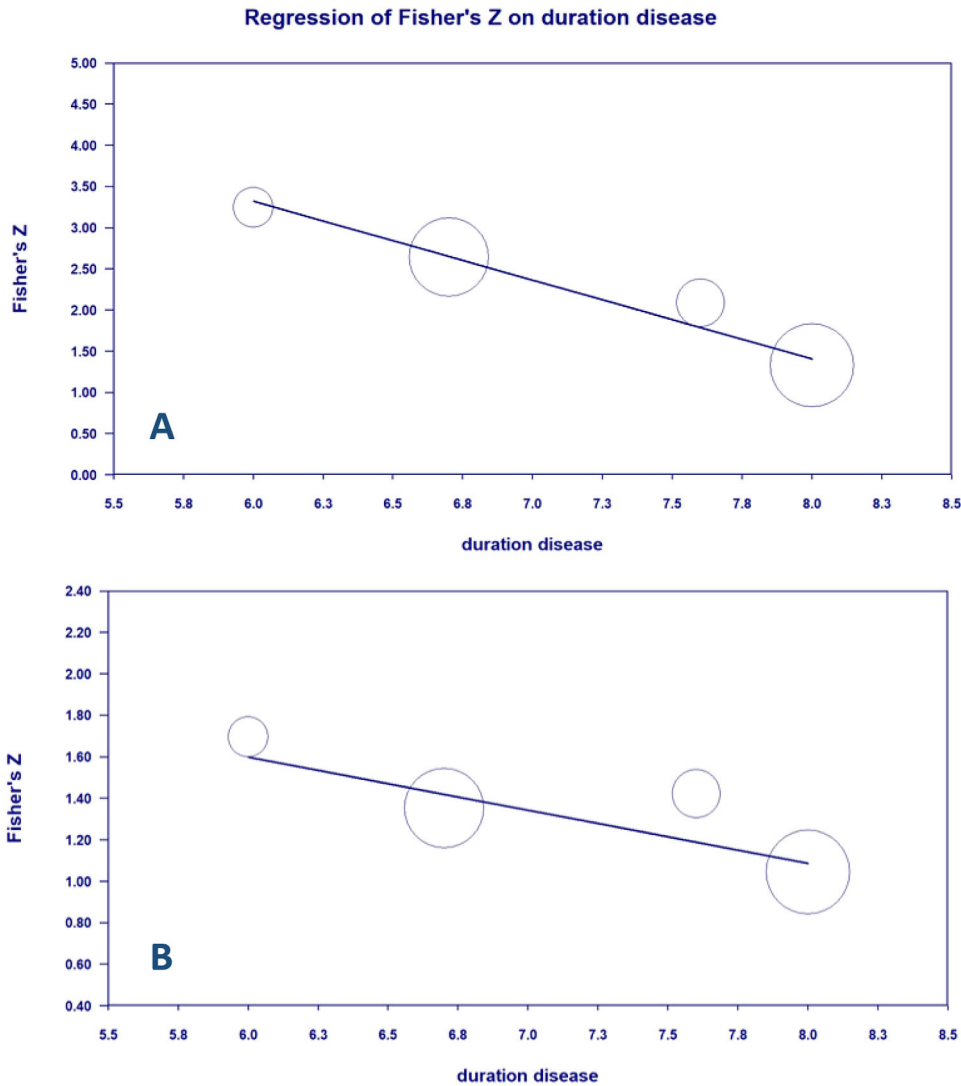


Figure 4. Meta-regression analysis for the association between the duration of disease, from the onset of inflammatory arthritis (years), and intra-reader (A) and inter-reader (B) reliability of computer-aided measurement of synovial membrane volume on magnetic resonance imaging.

total number of tender joints based on the results of two studies (28 MRIs; 31,35). Moreover, the overall poor correlation between disease duration and (semi)automated SV measures was reported based on the results of two studies, including 80 MRIs (range of reported ICC = 0.29-0.31; 35,36). Further analysis on other variables was not possible because of the low sample size.

Time needed to perform measurements. The average time needed to measure SV using computer-aided methods was 15 minutes (range = 8.2-20) based on the results of three studies (66 MRIs; 24,31,35), which was longer in comparison with the time taken to assess synovitis using the RAMRIS (7.1 minutes [range = 6-7.5]).

DCE-MRI perfusion parameters. Intra- and inter-reader reliability. The pooled intrareader reliability of quantitative DCE-MRI perfusion parameters for the assessment of synovitis across

four studies (85 MRIs) was 0.995 (0.821-1.0, $P < 0.001$, $I^2 = 97.0\%$) for ME, 0.997 (0.935-1.0, $P < 0.001$, $I^2 = 97.0\%$) for IRE, and 0.956 (0.854-0.987, $P < 0.001$, $I^2 = 80.1\%$) for N-total (Table 2, Supplementary Figure 8; 11,13,38,39). Funnel plots raised the possibility of publication bias for intrareader reliability of ME and IRE; however, correlation test demonstrated no evidence of publication bias ($P = 1.0$), with the adjusted ICC values of 0.890 (-0.296 to 0.996) for ME, 0.976 (0.725-0.998) for IRE, and 0.900 (0.701-0.969) for N-total. Intrareader reliability of other parameters, including T-onset, N-washout, N-plateau, IRE*N-total, and TIC shape categories were reported in fewer than two studies (Table 1); it was not feasible to conduct a pooled analysis.

Comparing with manual MRI-based measurements. A limited number of studies reported the correlation between the RAMRIS and DCE-MRI (semi)automated analysis. Results of three studies (129 MRIs; 24,40,41) showed an overall

poor-moderate level of agreement between DCE-MRI (semi) automated measures and the RAMRIS (for synovitis and BME); details are presented in Table 2. The correlation coefficient between the conventional ROI method and the DCE-MRI (semi)automated analysis was presented in two studies including 63 MRIs (14,39), with the ICC range of 0.31 to 0.92 for the correlation between rate of enhancement (RE)/rate of early enhancement (REE) and ME/IRE.

A few studies also presented the intrareader reliability of the conventional ROI method, for REE and RE (two studies including 56 MRIs; 14,42), and the RAMRIS (two studies including 26 MRIs; 24,42). Results showed that the ROI method (range of reported ICC, 0.02-0.99) and the RAMRIS (range of reported ICC, 0.90-0.94) had comparable intrareader reliability when compared with the computer-aided analysis of DCE-MRI (range of reported ICC = 0.989-0.99).

Correlation with clinical variables. The correlation between DCE-MRI measures and clinical/histopathological findings were assessed in four studies including 98 MRIs (39,42–44). However, the outcome measures were heterogenous across the included studies. Overall, two studies reported a significant correlation between DAS-28 and N-total measures with *P* values ranging from 0.003 to 0.01 (42,43). A significant correlation between IRE and clinical variables (including the number of swollen joints and DAS-28) was reported by two studies (range of *P* values, 0.003-0.01). Only one study reported results of correlation between histopathological findings and DCE-MRI (semi)automated measures (39).

Time needed to perform measurements. The time need to perform computer-aided DCE-MRI analysis was estimated to be 4 minutes (range = 3-8) based on the results of two studies (194 MRIs; 40,45), which was significantly shorter when compared with the manual assessment of DCE-MRI (30-45 minutes) and the RAMRIS (7-10 minutes).

DISCUSSION

The results from this research suggest an excellent intra- and inter-reliability for computer-aided image analysis, and we observed an excellent agreement between computerized and manual MRI-based measurements. Computerized methods, in particular for quantifying DCE-MRI parameters, potentially could reduce the time needed to perform image assessment. Computer-aided evaluation of imaging biomarkers of inflammatory arthritis on non-CE MRI, CE-MRI, and DCE-MRI could be considered as an efficient alternative to conventional observer-based methods.

Our literature search revealed a few narrative review articles about the role of computer-aided image analysis in patients with inflammatory arthritis (46–48). To the best of our knowledge, we are the first study to perform a systematic review and meta-analysis to assess the diagnostic value of computerized methods

by pooling available data and performing subgroup and meta-regression analyses. By estimating the overall pooled results of published literature and assessing potential biases, we could determine areas with a lack of evidence and guide future research on this topic.

In the first part of this research synthesis, we found an excellent overall intra- and inter-reliability ($r = 0.82$ to 0.99) for all computer-aided methods that were comparable or even higher when compared with the RAMRIS system and the ROI method (5,49,50). These findings are in line with prior hypotheses that computerized image analysis could improve the reproducibility of imaging findings by minimizing the impact of a reader's expertise. The majority of included studies used semiautomatic methods, which requires a reader to delineate regions of MRI that needs to be assessed. Semiautomatic image analysis might still have some limitations that are related to difficulties in determining the borders of joint pathologies. This fact could be the main reason that semi-automated methods had higher intrareader reliability ($r = 0.97$ to 0.98) in comparison with inter-reader reliability ($r = 0.86$ to 0.93). A few studies have introduced fully automatic methods for the assessment of BEV, SV, and perfusion parameters (12,25,27,36,44,51); however, we could not pool that data and compare the results of fully automated versus semiautomated methods because of small sample size.

Regarding DCE-MRI, when manual observer driven assessment is deployed, the size and position of the ROI will have significant impact on the diagnostic performance of this method (52,53). This issue has been minimized by a fully automatic voxel-based DCE-MRI data analysis with movement correction, developed by Kubassova et al (9). It has also been suggested that use of ROI optimizes the analysis as it allows to exclude large artifacts such as blood vessels (39,40). Larger studies are required to address these questions regarding the reliability and potential added value of fully automatic methods.

Furthermore, we assessed the possible impact of several variables on the reliability of computerized methods; computer-aided image analysis of BEV and SV were shown to have a higher reliability in older patients at early stage of disease and lower field strength of the MR scanner. It might be possible that developed computer-aided methods were validated by low field magnets and further investigation on high field scanner is required. There was a lack of data for the reliability of BME volume, tenosynovitis, and several DCE-MRI parameters and analytic methods such as TIC shape categories; we could not perform the pooled analysis based on the PRISMA protocol. BME has been considered as the strongest imaging biomarker for the prediction of disease worsening and development of bony erosions (54–56). Only two studies reported the reliability of computer-aided BME volume measurement, with the intrareader reliability of 0.92 to 0.99 and inter-reader reliability of 0.46 to 0.99 (26,27). Finally, it should be noted that our findings should be interpreted with caution, as there were only three to

eight studies in most parts of the analysis with moderate-high heterogeneity.

We evaluated the validity of computer-aided methods by pooling results of correlation analysis between (semi)automated measurements and manual measurements as well as clinical outcomes. We observed an overall excellent correlation ($r = 0.82-0.92$) between (semi)automated BEV/SV and RAMRIS. However, in studies of earlier disease, agreement between quantitative DCE-MRI perfusion parameters and RAMRIS (for synovitis) is poor to moderate. It can be concluded that quantitative DCE-MRI parameters might provide additional detailed information about the histological synovial inflammation, which can be used besides RAMRIS, especially for the diagnosis of early disease changes. Regarding the clinical variables, only a few studies reported the diagnostic value of (semi)automated measurements in predicting disease activity/progression (29,31,35,36). We only observed a moderate correlation between SV measures and ESR level. More study is required to answer the unresolved question regarding the role of (semi)automated image analysis in the prediction of disease course. Regarding BME assessment, only two studies focused on the validity of computer-aided BME volume measurement and reported the correlation coefficient of these measures with RAMRIS (range = 0.72-0.87; 24,26).

Feasibility assessment in this study was limited to time required to perform image analysis by the operator, and it is important to note that this information was only available in a small number of studies. Based on these limited results, we note that when a manual delineation of joints is required, then computer-aided BEV and SV measurements would require more time as compared with radiological scoring with RAMRIS. Automation of the segmentation of the erosion or perhaps even better segmentation of the bone volume instead of the erosion volume might be able to address this issue. In the majority of the assessed studies, manual segmentation of the erosion volume process was deployed (24,28,29,31,32).

For DCE-MRI, computer-aided data analysis was significantly faster when compared with manual methods (40,45). As such, computer-aided analysis of DCE-MRI data provides detailed information in a short time period. However, further validation and integration of the analysis method in a clinical working environment is required as these results were obtained from a limited number of studies in research studies.

Finally, only a few studies assessed the correlation between the imaging biomarkers (driven from computer-aided methods) and histopathological findings (39), where future studies need to clarify which imaging biomarker can accurately reflect the degree of disease activity.

This systematic review has several limitations. First, although we included 28 articles, a scant number of studies reported all outcomes of interests, and we only included 3 to 8 studies in each part of analysis. Second, a moderate-high level of heterogeneity was observed. To address this issue, we performed subgroup

and meta-regression analyses based on several variables, which could partly explain the partial impact of age, disease duration, and strength of scanner on the observed heterogeneity. Third, the funnel plot suggested presence of publication bias in a few parts of the analysis. To address this issue, we performed the Begg and Mazumdar test and the trim and fill method, and the final adjusted ICC values were reported, which were not significantly different from the crude estimates, enabling us to conclude that the pooled results are not biased. Fourth, we do not have sufficient evidence for the reliability and validity of several methods, such as (semi)automated BME volume and TIC shape categories.

Furthermore, the utility of (semi)automated image analysis in the assessment of disease progression, disease prediction, and treatment response has not been comprehensively assessed. For instance, Conaghan et al used computer-aided image analysis of follow-up MRIs to evaluate the impact of tofacitinib on progression of arthritis-related structural damages (57). The research studies examined in this review suggest that quantitative assessments allow for earlier separation of treatment groups from placebo. Cimmino et al assessed performance of rituximab with quantitative DCE-MRI to show that it can provide statistically significant efficacy as early as 4 weeks even in a small number of patients (42).

Fifth, non-English articles were excluded, which might result in overlooking several reports. Sixth, the majority of included studies only focused on patients with established RA; there was a lack of data on early RA and/or other types of inflammatory arthritis. Seventh, studies focused on axial joints and other imaging modalities such as ultrasonography, radiography, and CT scan have not been included. Eighth, the reliability and validity of computer-aided synovitis measurement has not been compared between non-CE and CE MRI because of the lack of evidence. Despite this limitation, however, this paper represents a robust examination of the available data and underscores the need for further research.

Feasibility was limited to time taken for image analysis by the operator and was only available in a small number of studies. Feasibility includes a large number of parameters, costs of equipment, training and calibration requirements as examples, and the results of the feasibility analysis in this paper should be viewed as limited to operator time.

In conclusion, computerized image analysis has excellent reliability and validity when compared with existing manual scoring methods such as RAMRIS in quantifying peripheral joint pathologies on MRI of patients with inflammatory arthritis.

With the emergence of quantitative imaging methods for assessment of scans acquired from patients with inflammatory arthritis, we expect that the research community will be developing other artificial intelligence (AI)-driven approaches, such as the ones based on deep learning algorithms. With the potential of rapid and reliable data analysis by computerized methodologies, the next step would be to explore their value in predicting disease progression and detection of early disease, especially through a

combination of imaging and clinical findings. Further studies are required to improve the diagnostic performance of available methods and define the optimal subgroup of patients who can benefit the most from these assessments.

ACKNOWLEDGMENTS

None.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Haj-Mirzaian had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Haj-Mirzaian, Kubassova, Boesen, Carrino, Bird.

Acquisition of data. Haj-Mirzaian, Kubassova, Boesen, Carrino, Bird.

Analysis and interpretation of data. Haj-Mirzaian, Kubassova, Boesen, Carrino, Bird.

REFERENCES

- Singh JA, Saag KG, Bridges Jr SL, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2016;68:1–26.
- Ledingham J, Snowden N, Ide ZJb. Diagnosis and early management of inflammatory arthritis. *BMJ* 2017;358.
- Kubassova O, Mobasheri A, Hinton M. The role of advanced MRI in the development of treat-to-target therapeutic strategies, patient stratification and phenotyping in rheumatoid arthritis. 2019. URL: https://www.ia-grp.com/ia_publications/the-role-of-advanced-mri-in-the-development-of-treat-to-target-therapeutic-strategies-patient-stratification-and-phenotyping-in-rheumatoid-arthritis/.
- Boesen M, Kubassova O, Sudoł-Szopińska I, Maas M, Hansen P, Nybing JD, et al. MR imaging of joint infection and inflammation with emphasis on dynamic contrast-enhanced MR imaging. *PET Clin* 2018;13:523–50.
- Boesen M, Østergaard M, Cimmino MA, Kubassova O, Jensen KE, Bliddal H. MRI quantification of rheumatoid arthritis: current knowledge and future perspectives. *Eur J Radiol* 2009;71:189–96.
- Šenolt L, Grassi W, Szodoray P. Laboratory biomarkers or imaging in the diagnostics of rheumatoid arthritis? [Original article]. *BMC Med* 2014;12:49.
- Conaghan PG, McQueen FM, Peterfy CG, Lassere MN, Ejbjerg B, Bird P, et al. The evidence for magnetic resonance imaging as an outcome measure in proof-of-concept rheumatoid arthritis studies. *J Rheumatol* 2005;32:2465–9.
- Ostergaard M, McQueen FM, Bird P, Ejbjerg B, Lassere MN, Peterfy CG, et al. Magnetic resonance imaging in rheumatoid arthritis advances and research priorities. *J Rheumatol* 2005;32:2462–4.
- Kubassova OA, Boyle RD, Radjenovic A. Quantitative analysis of dynamic contrast-enhanced MRI datasets of the metacarpophalangeal joints. *Acad Radiol* 2007;14:1189–200.
- Lavini C, de Jonge MC, van De Sande MG, Tak PP, Nederveen AJ, Maas MJMri. Pixel-by-pixel analysis of DCE MRI curve patterns and an illustration of its application to the imaging of the musculoskeletal system. *Magn Reson Imaging* 2007;25:604–12.
- Axelsen MB, Poggenborg RP, Stoltenberg M, Kubassova O, Boesen M, Hørslev-Petersen K, et al. Reliability and responsiveness of dynamic contrast-enhanced magnetic resonance imaging in rheumatoid arthritis. *Scand J Rheumatol* 2013;42:115–22.
- Orguc S, Tikiz C, Aslanalp Z, Erbay PD. Comparison of OMERACT-RAMRIS scores and computer-aided dynamic magnetic resonance imaging findings of hand and wrist as a measure of activity in rheumatoid arthritis. *Rheumatol Int* 2013;33:1837–44.
- Boesen M, Kubassova O, Parodi M, Bliddal H, Innocenti S, Garlaschi G, et al. Comparison of the manual and computer-aided techniques for evaluation of wrist synovitis using dynamic contrast-enhanced MRI on a dedicated scanner. *Eur J Radiol* 2011;77:202–6.
- Huang J, Stewart N, Crabbe J, Robinson E, McLean L, Yeoman S, et al. A 1-year follow-up study of dynamic magnetic resonance imaging in early rheumatoid arthritis reveals synovitis to be increased in shared epitope-positive patients and predictive of erosions at 1 year. *Rheumatology (Oxford)* 2000;39:407–16.
- 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.
- Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264–9.
- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
- Higgins JP, Altman DG. Assessing risk of bias in included studies. *Cochrane Handbook for Systematic Reviews of Interventions: Cochrane book series* 2008. p. 187–241.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–60.
- McGrath TA, McInnes MD, Korevaar DA, Bossuyt PM. Meta-analyses of diagnostic accuracy in imaging journals: analysis of pooling techniques and their effect on summary estimates of diagnostic accuracy. *Radiology* 2016;281:78–85.
- Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? [Original article] *Stat Med* 2002;21:1559–73.
- Begg CB, Mazumdar MJ. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50:1088–101.
- Duval S, Tweedie R. A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *J Am Stat Assoc* 2000;95:89–98.
- Yang H, Rivoire J, Hoppe M, Srikhum W, Imboden J, Link TM, et al. Computer-aided and manual quantifications of MRI synovitis, bone marrow edema-like lesions, erosion and cartilage loss in rheumatoid arthritis of the wrist. *Skeletal Radiol* 2015;44:539–47.
- Aizenberg E, Roex EA, Nieuwenhuis WP, Mangnus L, van der Helm-van Mil AH, Reijnerse M, et al. Automatic quantification of bone marrow edema on MRI of the wrist in patients with early arthritis: a feasibility study. *Magn Reson Med* 2018;79:1127–34.
- Crowley AR, Dong J, McHaffie A, Clarke AW, Reeves Q, Williams M, et al. Measuring bone erosion and edema in rheumatoid arthritis: a comparison of manual segmentation and RAMRIS methods. *J Magn Reson Imaging* 2011;33:364–71.
- Li X, Yu A, Virayavanich W, Noworolski SM, Link TM, Imboden J. Quantitative characterization of bone marrow edema pattern in rheumatoid arthritis using 3 Tesla MRI. *J Magn Reson Imaging* 2012;35:211–7.
- Bird P, Ejbjerg B, McQueen F, Ostergaard M, Lassere M, Edmonds J. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Exercise 5: An international multicenter reliability study using computerized MRI erosion volume measurements. *J Rheumatol* 2003;30:1380–4.

29. Poh MQ, Lassere M, Bird P, Edmonds J. Reliability and longitudinal validity of computer-assisted methods for measuring joint damage progression in subjects with rheumatoid arthritis. *J Rheumatol* 2013;40:23–9.
30. Døhn UM, Ejbjerg BJ, Hasselquist M, Narvestad E, Szkudlarek M, Møller J, et al. Rheumatoid arthritis bone erosion volumes on CT and MRI: reliability and correlations with erosion scores on CT, MRI and radiography. *Ann Rheum Dis* 2007;66:1388–92.
31. Bird P, Lassere M, Shnier R, Edmonds J. Computerized measurement of magnetic resonance imaging erosion volumes in patients with rheumatoid arthritis: a comparison with existing magnetic resonance imaging scoring systems and standard clinical outcome measures. *Arthritis Rheum* 2003;48:614–24.
32. Emond P, Inglis D, Choi A, Tricta J, Adachi J, Gordon CL. Volume measurement of bone erosions in magnetic resonance images of patients with rheumatoid arthritis. *Magn Reson Med* 2012;67:814–23.
33. Tomizza MA, Jessome MA, Barbosa J, Beattie KA, Bensen WG, Bobba RS, et al. Assessing the reliability of a semiautomated segmentation algorithm for quantifying erosions in the metacarpophalangeal joints of patients with rheumatoid arthritis. *J Rheumatol* 2015;42:1582–6.
34. Bird P, Joshua F, Lassere M, Shnier R, Edmonds J. Training and calibration improve inter-reader reliability of joint damage assessment using magnetic resonance image scoring and computerized erosion volume measurement. *J Rheumatol* 2005;32:1452–8.
35. Chand AS, McHaffie A, Clarke AW, Reeves Q, Tan YM, Dalbeth N, et al. Quantifying synovitis in rheumatoid arthritis using computer-assisted manual segmentation with 3 tesla MRI scanning. *J Magn Reson Imaging* 2011;33:1106–13.
36. Klartund M, Østergaard M, Lorenzen IJ. Finger joint synovitis in rheumatoid arthritis: Quantitative assessment by magnetic resonance imaging. *Rheumatology (Oxford)* 1999;38:66–72.
37. Czaplicka K, Wojciechowski W, Włodarczyk J, Urbanik A, Tabor ZJ. Automated assessment of synovitis in 0.2 T magnetic resonance images of the wrist. *Comput Biol Med* 2015;67:116–25.
38. van der Leij C, van de Sande MG, Lavini C, Tak PP, Maas MJR. Rheumatoid synovial inflammation: Pixel-by-pixel dynamic contrast-enhanced MR imaging time-intensity curve shape analysis—a feasibility study. *Radiology* 2009;253:234–40.
39. Axelsen MB, Stoltenberg M, Poggenborg R, Kubassova O, Boesen M, Bliddal H, et al. Dynamic gadolinium-enhanced magnetic resonance imaging allows accurate assessment of the synovial inflammatory activity in rheumatoid arthritis knee joints: a comparison with synovial histology. *Scand J Rheumatol* 2012;41:89–94.
40. Boesen M, Kubassova O, Bouert R, Axelsen MB, Østergaard M, Cimmino MA, et al. Correlation between computer-aided dynamic gadolinium-enhanced MRI assessment of inflammation and semi-quantitative synovitis and bone marrow oedema scores of the wrist in patients with rheumatoid arthritis—a cohort study. *Rheumatology (Oxford)* 2012;51:134–43.
41. Wojciechowski W, Tabor Z, Urbanik A. Assessing synovitis based on dynamic gadolinium-enhanced MRI and EULAR-OMERACT scores of the wrist in patients with rheumatoid arthritis. *Clin Exp Rheumatol* 2013;31:850–6.
42. Cimmino MA, Parodi M, Zampogna G, Boesen M, Kubassova O, Barbieri F, et al. Dynamic contrast-enhanced, extremity-dedicated MRI identifies synovitis changes in the follow-up of rheumatoid arthritis patients treated with rituximab. *Biologics* 2014;32:647–52.
43. Cimmino MA, Barbieri F, Boesen M, Paparo F, Parodi M, Kubassova O, et al. Dynamic contrast-enhanced magnetic resonance imaging of articular and extraarticular synovial structures of the hands in patients with psoriatic arthritis. *J Rheumatol* 2012;89:44–8.
44. Meier R, Thuernel K, Noël PB, Moog P, Sievert M, Ahari C, et al. Synovitis in patients with early inflammatory arthritis monitored with quantitative analysis of dynamic contrast-enhanced optical imaging and MR imaging. *Radiol* 2014;270:176–85.
45. Kubassova O, Boesen M, Cimmino MA, Bliddal H. A computer-aided detection system for rheumatoid arthritis MRI data interpretation and quantification of synovial activity. *Eur J Radiol* 2010;74:e67–72.
46. Stoel BC. Artificial intelligence in detecting early RA. *Semin Arthritis Rheum (Elsevier)* 2019;49:S25–8.
47. Stoel B. Use of artificial intelligence in imaging in rheumatology—current status and future perspectives. *RMD Open* 2020;6:e001063.
48. Stafford I, Kellermann M, Mossotto E, Beattie R, MacArthur B, Ennis S. A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. *NPJ Digit Med* 2020;3:1–11.
49. 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.
50. Haavardsholm EA, Østergaard M, Ejbjerg BJ, Kvan NP, Uhlig TA, Lilleås 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.
51. Stramare R, Coran A, Faccineto A, Costantini G, Bernardi L, Botsios C, et al. MR and CEUS monitoring of patients with severe rheumatoid arthritis treated with biological agents: a preliminary study. *Radiol Med* 2014;119:422–31.
52. Cimmino MA, Innocenti S, Livrone F, Magnaguagno F, Silvestri E, Garlaschi G. Dynamic gadolinium-enhanced magnetic resonance imaging of the wrist in patients with rheumatoid arthritis can discriminate active from inactive disease. *Arthritis Rheum* 2003;48:1207–13.
53. McQueen FM, Crabbe J, Stewart N. Dynamic gadolinium-enhanced magnetic resonance imaging of the wrist in patients with rheumatoid arthritis: comment on the article by Cimmino et al. *Arthritis Rheum* 2004;50:674–5; author reply 675–676.
54. McQueen FM, Benton N, Perry D, Crabbe J, Robinson E, Yeoman S, et al. Bone edema scored on magnetic resonance imaging scans of the dominant carpus at presentation predicts radiographic joint damage of the hands and feet six years later in patients with rheumatoid arthritis. *Arthritis Rheum* 2003;48:1814–27.
55. 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.
56. Hetland ML, Ejbjerg B, Hørslev-Petersen K, Jacobsen S, Vestergaard A, Jurik AG, et al. MRI bone oedema is the strongest predictor of subsequent radiographic progression in early rheumatoid arthritis. Results from a 2-year randomised controlled trial (CIMESTRA). *Ann Rheum Dis* 2009;68:384–90.
57. Conaghan PG, Østergaard M, Bowes MA, Wu C, Fuerst T, van der Heijde D, et al. Comparing the effects of tofacitinib, methotrexate and the combination, on bone marrow oedema, synovitis and bone erosion in methotrexate-naïve, early active rheumatoid arthritis: results of an exploratory randomised MRI study incorporating semiquantitative and quantitative techniques. *Ann Rheum Dis* 2016;75:1024–33.