

## Original article

# Examining the validity of the rheumatoid arthritis magnetic resonance imaging score according to the OMERACT filter—a systematic literature review

Thasia G. Woodworth<sup>1,\*</sup>, Olga Morgacheva<sup>1,\*</sup>, Olga L. Pimienta<sup>2</sup>,  
Orrin M. Troum<sup>2</sup>, Veena K. Ranganath<sup>1</sup> and Daniel E. Furst<sup>1</sup>

## Abstract

**Objective.** To examine whether the RA MRI score (RAMRIS) for RA of the wrist/hand meets the OMERACT filter criteria—truth (validity), discrimination and feasibility.

**Methods.** We conducted a systematic literature review in PubMed and Scopus, from 1970 through June 2014, focused on MRI measures of synovitis, osteitis/bone marrow oedema, erosions and/or joint space narrowing in RA randomized controlled trials and observational studies with cohort size  $\geq 10$ . Strength of evidence was assessed using the Cochrane Handbook criteria.

**Results.** Of 634 MRI titles/abstracts, 202 met the review criteria, with 92 providing at least 1 type of validity. Four articles provided criterion validity, and 26 articles utilized RAMRIS to assess 1.5 T MRI images. Histopathology data showed inflammation corresponding to MRI of synovitis and osteitis. MRI erosions corresponded to those identified with CT. Content and construct validity for RAMRIS synovitis, osteitis and erosions were documented by correlations with clinical, laboratory and/or radiographic data. Each measure was sensitive to change and responsive to therapy. RAMRIS synovitis and osteitis were able to discriminate between the efficacy of treatments vs placebo in 12-week studies, whereas RAMRIS erosions required studies of  $\geq 24$  weeks.

**Conclusion.** RAMRIS synovitis, osteitis and erosions imaged with 1.5 T MRI are valid and useful for evaluating joint inflammation and damage for RA of the wrist/hand, according to the OMERACT filter.

**Key words:** rheumatoid arthritis, magnetic resonance imaging, MRI, rheumatoid arthritis magnetic resonance imaging score, RAMRIS, OMERACT filter, structural progression, clinical trials

### Rheumatology key messages

- The RA MRI score was validated for 1.5 T MRI RA hand/wrist images.
- The RA MRI score for synovitis and osteitis were able to discriminate between potent treatments vs placebo in 12-week studies.
- The RA MRI score for erosions was able to discriminate with respect to worsening joint damage, but only after  $\geq 24$  weeks.

<sup>1</sup>Department of Medicine, Division of Rheumatology, David Geffen School of Medicine, UCLA, Los Angeles and <sup>2</sup>Keck School of Medicine, University of Southern California, Santa Monica, CA, USA  
Submitted 22 April 2015; revised version accepted 1 November 2016

\*Thasia G. Woodworth and Olga Morgacheva contributed equally to this study.

Correspondence to: Daniel Furst, UCLA Medical School, Rheumatology Division, 1000 Veteran Ave, Room 32-59, Los Angeles, CA 90095-1670, USA. E-mail: defurst@mednet.ucla.edu

## Introduction

RA is a systemic autoimmune disease, characterized by persistent or recurring synovitis that is often associated with substantial joint damage and disability [1]. Pharmaceutical developers and regulatory agencies place a high value on labelling that presents strong evidence for limitation of joint damage [2–4]. Sharp scoring of hand and foot radiographs is considered the reference standard for assessment of joint damage. However,

Sharp scores are relatively unable to detect worsening erosions and joint space narrowing (JSN) over short periods [requiring randomized controlled trials (RCTs) of  $\geq 12$  months duration] [2–5]. To reliably detect the efficacy of novel treatments that limit joint damage, sensitive methods that measure joint inflammation and structural damage are essential.

Periodic evaluation of MRI in studies over 3–6 months has detected effects of treatments on both joint inflammation and damage [6–8]. Recent draft updates to the Food and Drug Administration (2013) and European Medicines Agency (November 2011) guidance documents acknowledge that MRI measures might be useful for evaluating RA joint damage in RCTs; however, they indicated that MRI methods are not sufficiently validated [3, 4].

We performed a systematic literature review (SLR) to determine the state of validation of MRI measurement methods. Each identified method was evaluated according to the OMERACT filter: truth (face, criterion, content and construct validity), discrimination and feasibility [9–11].

## Methods

### Search strategy

To identify validated MRI measurement methods for RA, we conducted an SLR consistent with Preferred Reporting Items for Systematic Reviews and Meta-analyses methodology. Our initial search (1970–2011) in PubMed with Cochrane hedge, used the terms: RA, AND MRI, AND specific terms: for example, synovitis, JSN, erosions, osteitis or bone marrow oedema (BMO) AND humans AND RCTs, clinical studies. Based on the results demonstrating that RAMRIS was the only method with validation data, we updated our search adding RAMRIS to the search terms to July 2011 to July 2014. We also examined bibliographies of the ultimately selected articles. For the update, we used the same search strategy in Scopus prior to finalizing this review (see Search Strategy for RA Imaging Markers, available at *Rheumatology* Online).

### Article identification and evaluation

Independent pairs of authors evaluated the titles and abstracts using the following criteria: RA patients  $\geq 18$  years of age; MRI performed as part of a clinical study;  $\geq 10$  patients; and English. We excluded reviews, abstracts and letters to the editor. Prior to initiating the selection process, author pairs achieved a 95% consistency for article identification. Consensus for any disagreements was achieved by discussion among the authors.

Level of evidence was determined using the modified Cochrane Back Review Group criteria [12]: highest level: single/multicentre (RCTs) describing 1.5 T MRI acquisition and scoring together with statistical analyses and sufficient clinical, laboratory and/or radiographic data to examine validity; moderate level: single-centre RCTs and longitudinal observational studies (LOSs) with cohort size  $\geq 10$ , and/or limited description of MRI and/or scoring method or/and statistical method, and/or clinical data; low

level: articles with less than moderate evidence were rejected. We limited our selection to articles describing results with 1.5 T MRI, since image quality with low-field MRI is evolving. Articles describing criterion validity (e.g. histopathology, multiparameter/micro-CT) for at least one MRI feature were also included, regardless of cohort size).

### Data extraction

Each author extracted identified articles using a standardized form, entering data for MRI field strength, acquisition method and sequences, use of gadolinium (Gd), study design, patient characteristics, measurement method(s) and statistics used with clinical, laboratory and radiographic data associated with the measure [9–11] (Table 1). To be included, an article had to contain at least one validity criterion for at least one MRI-assessed feature.

### Analysis and reporting methods

Descriptions of each type of validity data were tabulated by study. To optimize our ability to identify MRI measurement methods with adequate validity data, we enumerated the number of studies using each method. To examine content validity, age range, RA duration, erosions, RF and ACPA status were recorded for each study. Intra- and inter-class correlations (ICCs) and kappa statistics were included to describe the reliability of the measure, together with Spearman's correlation coefficients examining relationships to clinical data. Wilcoxon's rank sum test was also used to evaluate sensitivity to change and discrimination. Image acquisition and scoring data were tabulated when available to assist in estimating feasibility.

## Results

A total of 634 titles and abstracts were examined, and 202 articles were extracted: 92 included at least 1 MRI measurement method and at least 1 validity criterion. Since RAMRIS was the only consistently used method, analysis was limited to studies reporting data to determine the validity, sensitivity to change/responsiveness, discrimination and feasibility of RAMRIS. The majority of studies reported RAMRIS measurement for 1.5 T MRI images (26 articles between 1999 and June 2014). There are 10 RCTs, 6 LOSs, 7 cross-sectional studies and 1 validation study, as well as 2 that examined criterion validity. No articles explicitly examined aspects of face validity. Seventeen articles reported use of RAMRIS in low-field MRI studies but were not included, as image quality and reliability of measurement have been evolving (see Figure 1).

### Criterion validity

Four articles provided moderate to high-level evidence for 1.5 T imaging criterion validity of synovitis, osteitis/BMO and/or erosions (Table 2). Based on histopathology of an inflamed RA joint visualized on MRI before the specimen was obtained, there were correlations with abnormalities as follows: one for synovitis/JSN [13], two for osteitis [14,

**TABLE 1** The OMERACT filter: truth (validity), discrimination, feasibility

Type of validity	Definition	Data extracted
Face	Expert opinion on credibility of the measure	MRI image of RA joint feature (erosion, osteitis, synovitis, JSN) to be measured
Criterion	Estimate the extent a measure agrees with gold standard—visualizes the MRI joint feature	Histology of lesions seen on MRI; contrast enhancement MRI of lesions consistent with increased tissue vascularity; X-ray/CT vs MRI
Content	Measure describes full spectrum of a disease; for example, features of patient population, including joint inflammation, deformity/damage	For each study's patient population: age range, gender distribution, RF/ACPA status, disease duration, Sharp vdH scores, treatment status/disease activity
Construct-convergent	Compares correlation coefficients between scores on the same health component (e.g. inflammation or damage), as measured by two different methods	Correlation coefficients between joint feature measurement and clinical/lab measures of inflammation such as swollen joint count, ESR/CRP, DAS28, or damage (erosions) with HAQ
Construct-divergent(discriminant)	Discriminant validity (or divergent validity) tests that constructs that should not relate to the measure, in fact, have no relationship (e.g. damage vs inflammation)	Correlation coefficients low/non-significant between inflammation joint feature measurement and measures of remission/low disease activity—e.g. DAS28, HAQ, etc.
Construct-predictive	Statistical correlation between scores for a single health component, as measured by two different instruments	Correlation coefficients between MRI measure and a clinical or other imaging measure assessed in comparable time frame; for example, MRI features predict Sharp vdH scores
Reliability	Repeatability, consistency and reproducibility (repeated measures yields the same result)	Inter-rater and intra-reader reliability measurement of scoring consistency between and within MRI readers—inter/intra-class correlation coefficients (ICCs), kappa statistics
Sensitivity to change	A measure's ability to detect a change between (relevant) time points	Change in joint inflammation and damage features detected between different time points: smallest detectable difference, minimal detectable change
Responsiveness	A measure's ability to detect clinically relevant change with treatment	Identified data documenting statistically significant changes in relation to treatment introduction or change
Discrimination	A measure's ability to distinguish between different treatments or features that influence outcome	Statistically significant differences in joint biomarker change, identifying greater efficacy for one intervention vs another
Feasibility	Essential element in determining usefulness and ability to use a measure reliably	Can measure be applied easily, given constraints of availability, time, money, interpretability?

Data taken from [9–11]. Sharp vdH scores: Sharp van der Heijde damage scores.

15] and one for erosions [15]. Micro-CT studies comparing CT erosions to MRI erosions were also accepted.

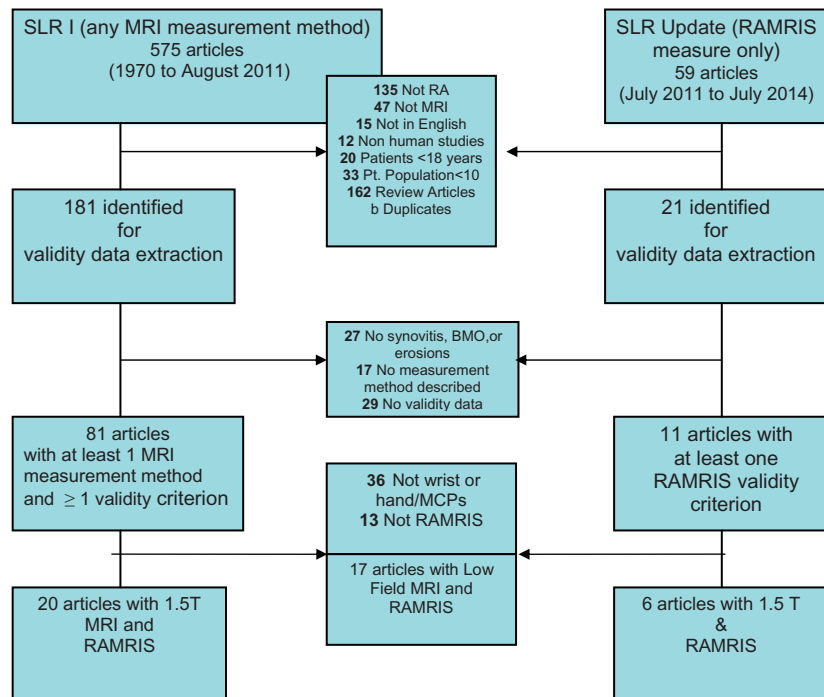
#### Synovitis

Ostendorf *et al.* [13] reported synovitis histopathology for second MCP miniarthroscopy within 24 h of 1.5 T MRI with dynamic contrast enhancement. MRI synovial vascularity and proliferation correlated with synovial hyperaemia and thickening on miniarthroscopy, respectively ( $P = 0.0038$ ;  $P = 0.0063$ ). Bony changes by miniarthroscopy correlated with JSN on MRI ( $P = 0.0015$ ).

#### Osteitis

Bone marrow oedema (BMO—called osteitis throughout the rest of this article) seen on MRI correlated with histopathology of those lesions in total joint replacement specimens [14, 15]. In 4 patients and 7 bones, McQueen *et al.* [14] examined MRI pre-surgery vs histological evaluation at orthopaedic surgery. High-grade MRI bone oedema was strongly associated with histology showing an inflammatory infiltrate consistent with osteitis. Similar findings were reported by Jimenez-Boj *et al.* [15] for lesions identified on 1.5 T MRI without Gd enhancement.

Fig. 1 Article selection



### Erosions

Multiparameter CT and microCT are recognized to be especially sensitive for identification of bony abnormalities (Table 2). Albrecht *et al.* [16] reported that MRI-visualized erosions were readily seen with CT or microCT [16].

### Is RAMRIS scoring of 1.5 T MRI images of the RA hand/wrist valid, and does it meet the OMERACT filter?

OMERACT methods development conferences for RAMRIS were held in 2000, 2004 and 2008, and two atlases were published in 2005, standardizing RAMRIS and documenting that experts agree on the features to be measured, providing evidence of face validity [17–22] (Table 3). An overview of the numbers of articles that show how RAMRIS meets the OMERACT filter criteria is provided by category in Table 4. Supplementary Tables S1–3, available at *Rheumatology* Online, provide detailed data and references.

### Content validity

RAMRIS measures RA joint inflammation and erosions regardless of patient age, disease duration, disease activity and treatment status/response. Evidence of content validity is provided by seven studies describing that RAMRIS can measure synovitis, osteitis and erosions in early and established RA. RAMRIS performs equally regardless of treatment status/prior response, RF/ACPA status, age [23–29] or damage (erosions) [27–29]. There are 11 reports describing the application of RAMRIS separately in

patients with early RA, as well as two articles describing use in established RA patients. No analyses explicitly describe the impact of these contextual factors (see supplementary Tables S1–3, available at *Rheumatology* Online).

### Construct validity

Construct validity includes convergence, correlations with clinical and laboratory results that measure joint inflammation (e.g. synovitis and osteitis with ESR, CRP, DAS28, etc.) or damage (e.g. RAMRIS erosions correlates with HAQ-DI).

It also includes discriminant or divergent validity (e.g. no correlation would be expected between RAMRIS synovitis and osteitis scores and DAS remission). For predictive validity, RAMRIS measures should correlate with subsequent joint damage such as the van der Heijde modified Sharp radiographic scores (vdHSSs) (Table 4 and supplementary Tables S1–3, available at *Rheumatology* Online).

### Convergent vs divergent (discriminant) construct validity

#### Synovitis

Six studies support convergent construct validity, demonstrating RAMRIS synovitis correlations with ESR, DAS28, HAQ-DI and ACR response [ $r=0.21-0.6$  ( $P < 0.05$ )] [6, 8, 26, 27, 30, 31]. Three studies reported lack of correlations between RAMRIS synovitis and HAQ-DI, DAS28 remission and ACR remission ( $P=0.22-0.60$ ), suggesting divergent construct validity for active RAMRIS synovitis compared with remission and full function [26, 27, 29].

### Osteitis

Three studies support this construct measurement, correlating with clinical and laboratory measures of inflammation: DAS28, CRP, swollen joints and grip strength ( $r = 0.22-0.32$ ,  $P = 0.001-0.05$ ) [6, 8, 31]. Discriminant construct validity was demonstrated in three studies of a DMARD inadequate responder (DMARD-IR) subset ( $P = 0.08-0.96$ ) [8, 29, 31], whereas there was no correlation with ACR response or DAS remission.

### Erosions

One study supported construct validity for RAMRIS erosions, describing correlations with DAS28 and CRP (Spearman's  $r = 0.23$ ;  $P < 0.001$ ) [8], while another study, using DAS28/CRP/ESR, showed no such correlations ( $P = 0.16-0.46$ ) [21]. As with osteitis, no correlations were found in DMARD-IR patients [8].

### Predictive construct validity

For predictive construct validity for 1.5 T MRI, McQueen *et al.* [21] demonstrated that a preliminary BMO/osteitis score predicted the 6-year total Sharp X-ray score (hands and feet,  $P = 0.01$ ).

### Synovitis/osteitis/erosions

Three studies correlated vdHSS-based radiographic progression with RAMRIS synovitis (Spearman's  $r = 0.25-0.5$ ,  $P < 0.001-0.05$ ), osteitis (Spearman's  $r = 0.48$ ,  $P < 0.05$ ) and predicted vdHSS erosion score [odds ratio (OR) = 2.5 (1.0-6.1), and erosions correlated vdHSS (Spearman's  $r = 0.27-0.76$ ,  $P < 0.001-0.05$ )] [6-8].

### Reliability and reproducibility

Inter and intra-reader consistency (reliability) and reproducibility (repeatability of measurement in short time frame, or/and by more than one observer) was assessed in a range of articles—RCTs, LOSs, cross-sectional studies, as well as one validation study. Overall, reliability and reproducibility were good to excellent (Table 4 and supplementary Tables S1-3, available at *Rheumatology* Online).

### Synovitis

The measurement of synovitis was reliable and reproducible. In 12 articles, the intra-rater ICCs were (0.77-0.98) and kappa ( $k$ ) statistics were (0.81-0.88) [6, 7, 26, 28, 29, 32-38]. In 10 articles, the inter-rater reliability ICCs were (0.68-0.97,  $k = 0.74$  and  $r = 0.79-0.97$ ) [24, 26, 28, 30, 32, 34-37, 39].

### Osteitis

Twelve articles of MRI document intra-rater ICCs ranging from 0.65 to 0.94 and/or  $k$ 's of 0.6-0.94 [6, 7, 26, 28, 29, 32-38]. Inter-reader ICCs were 0.68-0.97,  $r = 0.73-0.95$  and/or  $k$ 's of 0.58 in 10 articles [24, 26, 28, 30, 32, 34-37, 40].

### Erosions

In 11 studies, intra-rater ICCs ranged between 0.51 and 0.99 and  $k$ 's 0.67-0.87 [6, 7, 26, 28, 33, 34-38], whereas in

**TABLE 2** Criterion validity for 1.5 T MRI synovitis, osteitis and erosions

Strength of evidence	Correlative evidence	MRI features assessed	Study design	Reference
Moderate to high (localization of MRI lesions vs MA unclear)	Correlation between DCE, macroscopic synovial vascularity/hyperaemia, MRI synovitis, clinical and radiographic findings	Synovitis—synovial DCE, synovial proliferation, joint space narrowing	Histology synovitis by MA (second MCP) vs DCE MRI, vs clinical and radiographic findings—22 estRA pts	Ostendorf <i>et al.</i> [13]
MRI images comparable among pts—Modest	MRI bone oedema (osteitis) associated with subchondral inflammatory cell infiltrate	Osteitis	Histological specimens (7 bones from 4 TJR pts) vs MRI osteitis in 11 estRA pts	McQueen <i>et al.</i> [14]
High, due to specific localization of lesions in joint vs MRI	MRI bone erosions and edema reflect replacement of bone marrow fat by inflammatory cells	Erosions, osteitis	Histology sequential joint sections—TJR (12 joints from 3 estRA pts) vs MRI vs/PIPs	Jimenez-Boj <i>et al.</i> [15]
High	MRI erosion scores strongly correlated with erosion volumes on micro CT	Erosions	RAMRIS erosion scores vs microCT	Albrecht <i>et al.</i> [16]

DCE: dynamic gadolinium contrast enhancement; ERA: early RA; estRA: established RA; MA: miniarthroscopy; TJR: total joint replacement.

**TABLE 3** RAMRIS scoring of synovitis, osteitis and erosions<sup>a</sup>

Feature	Description [22–28]	Scoring [17] (RAMRIS units)
Synovitis	Soft tissue with increased thickness/volume (T1-weighted image) and water content (high signal in fat-suppressed T2-weighted image). For the wrist, this feature is assessed in three wrist regions (distal radio-ulnar, radio-carpal, intercarpal–metacarpal joints) of the dominant or most inflamed wrist. For the hand, it is evaluated in MCPs 2–5 of the dominant or most inflamed side with or without Gd enhancement (signal intensity increase at 4–5 min post injection)	0 (normal) to 3 (mild, moderate, severe) for each region/joint; maximum: 21
Osteitis	Identified within the subchondral trabecular bone as a lesion with ill-defined margins and signal characteristics consistent with increased water content (may also be seen in association with erosion); a high signal with or without Gd on fat-suppressed T2-weighted and short tau inversion recovery (STIR) MR images, low signal on T1-weighted images. Each bone of the dominant or most inflamed hand-wrist is scored separately	0 (normal), (1) 1–33% of bone, (2) 34–66% of bone and (3) 67–100% of bone showing increased water content. Maximum: 69 (45 for wrist alone)
Erosions	Sharply marginated bone lesions in a juxta-articular location, visible in two planes, with a cortical break area in at least one plane and loss of normal low signal intensity of cortical bone on T1-weighted images (loss of high signal on T2-weighted images). As with osteitis, each bone of the wrist-hand is scored separately	0–10, according to 10% increments of bone eroded. Maximum: 230 (150 for wrist alone)

<sup>a</sup>There is no agreed RAMRIS measure for joint space narrowing.

12 studies, the inter-rater ICC range was 0.81–0.96 and  $r = 0.84$ –0.92 [24, 26, 28, 30, 32, 34–37, 39–41].

### Sensitivity to change

Sensitivity to change describes the difference between two time points within a treatment or population, independent of differences between treatments (see change in statistics described in supplementary Tables S1–S3 available at *Rheumatology* Online).

#### Synovitis

Data for sensitivity to change for the RAMRIS synovitis measurement was provided in 14 studies: 4 LOSs, 9 RCTs and 1 validation study. The time points analysed were 4–6, 12–18, 24 and 52–54 weeks.

Significant change was seen in two single-centre RCTs after only 4–6 weeks of treatment: one evaluating a TNFi, and one evaluating double-filtration plasmapheresis (DFPP) [26, 42]. Four other TNFi studies demonstrated synovitis improvement in 12–18 weeks. One example is a study of infliximab vs MTX in which median (min, max) synovitis decreased, respectively, by  $-7$  ( $-12$ , 1) and  $-1$  ( $-4$ , 2), ( $P = 0.003$ –0.05) [32, 34, 43–45]. Other studies showed improvement after 52–54 weeks [45–47].

#### Osteitis

Similar to synovitis, seven RCTs described change in RAMRIS osteitis at time points ranging from 4 to 54 weeks. The majority of RCTs evaluated TNFi treatment vs MTX while patients were on background DMARDs.

In one study with infliximab, a significant decrease in osteitis was already seen at week 4, and confirmed at week 16 [43]. In another four TNFi RCTs, the change was seen at 12–18 weeks, and in two golimumab RCTs changes were seen at 24 weeks,  $P < 0.001$  [35, 37]. In two other infliximab studies there was significant reduction in osteitis—median (min, max):  $-9$  ( $-25$ , 5) ( $P < 0.05$ ) at 52–54 weeks [46, 47]. However, significant osteitis change was not seen at 6 weeks in two other TNFi studies [26, 27] or with DFPP ( $P = 0.18$ ) [42]. In an abatacept trial, a decrease in osteitis of  $-1.94$  (0.86) was seen with abatacept/MTX at 18 weeks [23, 32]. In a MRI measurement validation study reporting 1-year change detected with RAMRIS in four early RA (ERA) and six established RA (estRA) patients by multiple readers, the smallest detectable difference at 12 months for osteitis was 2.73 (ERA), and 3.68 for established RA patients [36].

#### Erosions

While erosions are seen on MRI with higher sensitivity compared with X-ray, change in RAMRIS erosions has not been consistently shown. Although effects to limit worsening may be discernible at 24 weeks with RAMRIS and vdHSSs, healing is not readily demonstrable with these methods, raising questions about the effects of treatment vs measurement methods. The magnitude of change/absence of change at 6–54 weeks in nine RCTs and five LOSs was  $<1\%$  of the range and independent of disease duration, present or prior treatment/response and

TABLE 4 Overview of number of articles providing each type of validity data<sup>a</sup>

Biomarker	Construct (no. of reports)	Predictive	Sensitivity to change/ responsiveness	Discrimination	Intra-rater reliability	Inter-rater reliability	
Synovitis	(9)	+3 VdHSS	+4 LOSs +8 RCTs -1 RCTs +1 validation study	+6 RCTs -3 RCTs +1 validation study -1 cross-sectional	+12	+10	
	+ 3 DAS28						
	+ 4 CRP/-1 CRP						
	+2 ESR						
	+1 HAQ-DI/-1 HAQ-DI						
	+1 SJC						
	+1 ACR20						
	+1 Grip Strength						
	-1 Pt. Pain						
	+1 InYKL/+1 InMMP						
	-1 DAS28 remission						
	-1 ACR remission						
Osteitis/BMO	(3)	+3 VdHSS	+4 LOSs +7 RCTs -3 RCTs +1 validation study	+7 RCTs -3 RCTs +1 cross-sectional +1 validation study	+12	+10	
	+1 DAS28/-1 DAS28						
	+2 CRP/-1 CRP						
	+1 SJC						
	+1 Grip Strength						
	-1 InYKL						
	+1 InMMP						
	-1 DAS remission						
	-1 ACR remission						
	Erosions	(3)	+2 VdHSS -1 VdHSS +1 USGS inflammation -1 DXR-BMD	+1 LOS -3 LOS +5 RCTs -4 RCTs +1 validation study	+5 RCTs -4 RCTs +1 validation study +1 cross-sectioned	+11	+12
		+1 DAS28/-2 DAS28					
		+1 CRP/-2 CRP					
-1 ESR							

<sup>a</sup>Negative studies for content validity report data for only a subpopulation of RA patients and, thus, not including a wide range of patients; for example, only early, or only established RA, or only MTX-inadequate responders. + Articles providing statistically significant results supporting type of validity. - Articles that provide data for type of validity, but result is not statistically significant. LOS: longitudinal observational study; RCT: randomized controlled trial; SJC: swollen joint count; USGS: grey scale ultrasound; YKL: human cartilage glyco-protein-39 (marker of cartilage injury).

disease activity (supplementary Table S3, available at *Rheumatology* Online).

RAMRIS erosions did not change statistically in 6 weeks in two studies [26, 27] or in 12–18 weeks in seven studies [6, 7, 21, 23, 30, 33, 35]. There is some inconsistency at 3, 6 and 12 months in LOSs [6, 7, 25, 38]. In two LOSs, changes were discernible at 12 weeks, whereas in two others there were no detectable changes [6, 38]. For the studies discerning changes, the standardized response means were  $>0.23$ – $0.32$ , and the patients had either early or established disease with high disease activity [6].

At 24 weeks, three RCTs and one LOS found small changes [6, 23, 35, 42, 45], whereas in one RCT and three LOSs no change was seen [7, 25, 33]. For example, in a denosumab RCT, an increase in RAMRIS erosions at 24 weeks was detected in all three arms, but was  $0.06/0.13/1.75$  U high-dose vs low-dose vs placebo [41].

Even at 52 weeks, RAMRIS did not detect a change in erosions consistently. Of seven studies (three RCTs and four LOSs [6, 7, 25, 37, 41, 43, 45]), in only three were there changes in RAMRIS erosions [6, 38, 42].

#### Discrimination: does RAMRIS differentiate efficacy between therapies?

Discrimination refers to the ability to differentiate between therapies. Among nine RCTs, discrimination using synovitis and osteitis was usually seen by 12–18 weeks. In contrast, RAMRIS erosions were discriminating at 24–52 weeks (occasionally at 12–18 weeks) (see statistics described in supplementary Tables S1–3, available at *Rheumatology* Online).

#### Synovitis

Six of nine RCTs provide evidence that change in RAMRIS synovitis can discriminate between treatments at all time points. Some change in RAMRIS synovitis is seen as early as 6 weeks following initiation of a TNFi, and consistently at 12 weeks [23, 26, 27, 30, 32, 34, 42, 43, 45]. In golimumab studies in estRA and ERA patients, differences in synovitis could be discriminated at 12 weeks, and the mean (s.d.) change in estRA was  $-1.77$  (2.54) and in ERA,  $-1.92$  (3.09), compared with placebo,  $-0.15$  (2.75) and  $+0.14$  (2.98), respectively ( $P < 0.001$  in both studies) [33, 35]. By 24 weeks RAMRIS synovitis was consistently able to be discriminated in all studies [23, 26, 27, 30, 32, 34, 42, 43, 45]. For example, DFPP was able to be discriminated from control with a median difference of 7 by RAMRIS ( $P < 0.001$ ) [48].

#### Osteitis

None of three RCTs evaluating either TNFi or DFPP reported reduced osteitis vs control at 4–6 weeks [48, 26, 27]. In six of seven RCTs, RAMRIS osteitis was able to discriminate between active treatment and placebo at 12–18 weeks [24, 30, 33, 35, 45, 44]. A single study of a SyK inhibitor (SyKi) was not able to discriminate between active treatment and placebo. RAMRIS osteitis changes were  $-0.2$  (SyKi) vs  $+1.2$  (control) ( $P = 0.058$ ) [30].

Change in RAMRIS osteitis measured at 24 and 52 weeks in five RCTs could consistently discriminate

between the experimental agent and the control [32, 35, 45, 44, 48].

#### Erosions

Discrimination was evaluable in nine RCTs at time points across 6–54 weeks. The majority of RCTs examined change in erosions with TNFi treatment in patients while on background DMARDs.

Differences in RAMRIS erosions scoring did not discriminate at 6 weeks [26, 27] and were inconsistent at 12–18 weeks (two studies showed discrimination and two did not) [21, 30, 33, 35, 44]. At 24 weeks discrimination was generally achieved. In a TNFi study in MTX-naïve patients, the mean change (s.d.) reported for TNFi vs placebo was  $-0.40$  (4.31) vs  $-0.24$  (6.31), ( $P = 0.010$ ) [34]. In a trial of denosumab, RAMRIS detected limitation of erosions with a mean change for the high dose vs placebo of  $0.06$  vs  $1.75$ , respectively ( $P = 0.007$ ) [41]. At 52 weeks, patients treated with infliximab/MTX vs infliximab/placebo demonstrated an erosion change of median (IQR) 1 (3) ( $P = 0.05$ ) [45].

#### Feasibility

We could not find specific data on the day-to-day issues of feasibility, neither time to perform MRIs, to quality assure images, nor importantly, the cost of these key activities. Nevertheless, there is good evidence from RCTs that MRIs have been done on all continents and in several multicentre studies. Thus, while feasible in dedicated centres during adequately funded clinical trials, based on the fact these studies have been done, feasibility outside that venue has not been examined.

## Discussion

This SLR represents a critical examination of the published data regarding the state of validation of RAMRIS scoring of the hands and wrists in RA. It provides evidence for the validity of RAMRIS for measurement of synovitis and osteitis, in the hands/wrists of RA patients treated in clinical trials. Further, these data describe RAMRIS responsiveness and sensitivity to change together with evidence that RAMRIS can discriminate efficacy among therapies. Specific data demonstrating feasibility for clinical use were not found, but evidence for utility to assess whether an intervention limits joint inflammation and damage in well-controlled clinical trials appears sufficient to apply RAMRIS measures as end points.

Three prior reviews examined the validity and usefulness of RAMRIS. In 2008, Hodgson *et al.* briefly described the state of RAMRIS validation, and observed that reliability and discriminatory validity of scoring remained to be verified [49–53]. In 2010, Suter *et al.* [54] reported an SLR using Cochrane methods to specifically examine the role of MRI in the diagnosis and prognosis of ERA and showed its ability to predict later radiographic damage, but they did not examine validity. Our SLR extends the earlier results, by presenting validation evidence not previously described, which supports many of the recommendations



of the 2013 ACR Clinical Trials Task Force Imaging Group and OMERACT Inflammatory Arthritis Working Group article advocating RAMRIS measurement in clinical trials [55].

Following on the ACR Task Force report, this SLR further documents the ability of RAMRIS measures to discriminate between treatments, including only four of the same studies [23, 30, 32, 34, 41] that have applied RAMRIS measures for synovitis, BMO/osteitis and/or erosions. We identified 10 published RCTs [23, 26, 27, 30, 32, 34, 41, 42, 45, 48] that reported at least 1 RAMRIS measure and used high-field MRI, only 8 of which reported data for all 3 RAMRIS measures [23, 26, 27, 30, 32, 34, 45, 48]. Importantly, not all measures perform consistently in these studies.

Sensitivity to change and responsiveness are strongly supported for RAMRIS synovitis and osteitis, which improves within 12–18 weeks with potent treatments. Further, changes in RAMRIS synovitis and osteitis usually discriminate between treatments in this time frame [30, 35, 45, 33, 44, 48]. Importantly, in terms of detecting clinical benefit, high RAMRIS synovitis and osteitis predicted subsequent radiographic joint damage, and improvements predicted limitation of damage [6–8, 56].

In contrast, for erosions, sensitivity to change is largely detected as worsening, so that limitation of damage with improvement in synovitis and osteitis results in smaller changes and in a minority of patients, even at 24 weeks and beyond. Although, in one of the RCTs using a RAMRIS erosion end point [34], discrimination was seen at 12 weeks in MTX-naïve patients treated with golimumab/MTX, in four other studies at least 24–52 weeks were required to discriminate treatment efficacy [34, 41, 43, 45].

Thus, examination of whether RAMRIS erosion scoring meets the OMERACT filter is challenging because current treatment recommendations quickly and effectively limit erosions measured by vdHSSs of hand/feet radiographs. Worsening of erosions is seen in only a small minority of patients, and improvement is uncommon, with changes typically small at the group level.

Of particular importance, and not previously carefully examined, the time to improvement is different for different features of joint damage and inflammation assessed by RAMRIS. Some of the differences may derive from patient population and/or sample size, as well as by study design, when measures are taken at different times (e.g. at 6, 12, 18, as well as at 24 weeks vs only at 24 weeks or only at 52–54 weeks). RAMRIS synovitis scores decreased as early as 4–6 weeks. These rigorously conducted studies also reported smallest detectable differences, minimal clinical differences and standardized response means [26, 27, 35, 42]. Significant improvement in osteitis was seen by 12–18 weeks and occasionally 24 weeks with various biologics with or without MTX background therapy [23, 26, 27, 30, 32, 42, 43, 45].

These findings provide useful data for the preparation and conduct of studies designed to guide decisions regarding future development of novel RA treatments. The magnitude, variability, and time to observed changes in

RAMRIS synovitis and osteitis are reasonably characterized, and we believe they may be used as end points in studies of 3–4 months duration. If a trial design requires erosions as an end point, studies of 6–12 months may be required.

### Limitations

The results described here are based only on published studies in peer-reviewed journals, and only those published in English. Potentially relevant data not yet published (e.g. studies reported in recent abstracts and presentations) are not included [5]. In addition, some inevitable heterogeneity was introduced because there is variability in selection of the hand and/or wrist imaged (right/left, dominant/non-dominant), MRI acquisition (e.g. standardization of positioning, quality control of images), and importantly, whether contrast was used to assess inflammation.

Further, to fully substitute for radiographic vdHSSs in studies aimed at regulatory approval and labelling for novel treatments, a measure for JSN is desirable. While there are at least three JSN measures proposed for RAMRIS, consensus and validation have not been achieved thus far [5, 34, 46, 57]. There is also ongoing research for future inclusion of tenosynovitis to strengthen the measurement of joint inflammation within RAMRIS [37, 47, 58, 59].

Since completion of our search and analysis, two articles describing RCTs using 1.5 T RAMRIS end points and included in the ACR Task Force report have been published [60, 61]. The data further validate these results, and also provide additional evidence of feasibility.

Further, we could not fully address feasibility in quantitative terms as none of the articles reported costs of equipment and standardized acquisition, nor any information regarding patient burden. The requirement for i.v. contrast (Gd) to accurately measure synovitis likely impacts feasibility significantly [40].

### Conclusions

This SLR provides evidence that RAMRIS measurement of synovitis, osteitis and erosions meets the OMERACT filter for use in RA clinical trials evaluating 1.5 T MRI images of RA hand/wrist. However, timing of the measurements to meet validation criteria is variable. Importantly, data from published RCTs provide evidence that RAMRIS measures are responsive and can generally discriminate efficacy between treatment groups within 12–18 weeks. Since changes in erosions are small, 24–52 weeks are usually necessary for discrimination of effective treatment. Periodic systematic evaluation and standardization to optimize image quality and scoring precision appear necessary for assuring optimal RAMRIS performance.

All other authors have declared no conflicts of interest.

### Supplementary data

Supplementary data are available at *Rheumatology* Online.

## Acknowledgements

We gratefully acknowledge Rikke Ogawa, Lisa Federer and Bethany Myers (University of California, Los Angeles research librarians) for their help with the data searches, Gabriel Valdivia (executive assistant) for his invaluable help with preparation of the manuscript and bibliography, and Nashla Barroso for her assessment of article quality, without whom this article could not have been completed.

*Funding:* Bristol Myers Squibb provided funding for the literature review.

*Disclosure statement:* D.E.F. has received grant/research support from Abbvie, Actelion, Amgen, BMS, National Institutes of Health, Novartis, Pfizer and Roche/Genentech and has consulted for Abbvie, Actelion, Amgen, BMS, Cytos, Novartis, Pfizer and Roche/Genentech. V.R. has received grants from Genentech and Pfizer and was on the advisory board for BMS.

## References

- Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. *Lancet* 2010;376:1094–108.
- Smolen JS, Han C, Bala M *et al.* Evidence of radiographic benefit of treatment with infliximab plus methotrexate in rheumatoid arthritis patients who had no clinical improvement: a detailed subanalysis of data from the anti-tumor necrosis factor trial in rheumatoid arthritis with concomitant therapy study. *Arthritis Rheum* 2005;52:1020–30.
- Smolen JS, Avila JC, Aletaha D. Tocilizumab inhibits progression of joint damage in rheumatoid arthritis irrespective of its anti-inflammatory effects: disassociation of the link between inflammation and destruction. *Ann Rheum Dis* 2012;71:687–93.
- Døhn UM, Ejlertsen B, Hasselquist M *et al.* Detection of bone erosions in rheumatoid arthritis wrist joints with magnetic resonance imaging, computed tomography and radiography. *Arthritis Res Ther* 2008;10:R25.
- Peterfy C, Østergaard M, Conaghan PG. MRI comes of age in RA clinical trials. *Ann Rheum Dis* 2013;72:794–6.
- 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.
- Bøyesen P, Haavardsholm EA, Østergaard M *et al.* MRI in early rheumatoid arthritis: synovitis and bone marrow oedema are independent predictors of subsequent radiographic progression. *Ann Rheum Dis* 2011;70:428–33.
- Emery P, van der Heijde D, Østergaard M *et al.* Exploratory analyses of the association of MRI with clinical, laboratory and radiographic findings in patients with rheumatoid arthritis. *Ann Rheum Dis* 2011;70:2126–30.
- Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol* 1998;25:198–9.
- Bellamy N. Science of assessment. *Ann Rheum Dis* 2005;64:ii42–5.
- Boers M, Kirwan JR, Wells G *et al.* Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *J Clin Epidemiol* 2014;67:745–53.
- Furlan AD, Pennick V, Bombardier C, van Tulder M. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine (Phila Pa 1976)* 2009;34:1929–41.
- Ostendorf B, Peters R, Dann P *et al.* Magnetic resonance imaging and miniarthroscopy of metacarpophalangeal joints: sensitive detection of morphologic changes in rheumatoid arthritis. *Arthritis Rheum* 2001;44:2492–502.
- McQueen FM, Gao A, Østergaard M *et al.* High-grade MRI bone oedema is common within the surgical field in rheumatoid arthritis patients undergoing joint replacement and is associated with osteitis in subchondral bone. *Ann Rheum Dis* 2007;66:1581–7.
- Jimenez-Boj E, Nöbauer-Huhmann I, Hanslik-Schnabel B *et al.* Bone erosions and bone marrow edema as defined by magnetic resonance imaging reflect true bone marrow inflammation in rheumatoid arthritis. *Arthritis Rheum* 2007;56:1118–24.
- Albrecht A, Finzel S, Englbrecht M *et al.* The structural basis of MRI bone erosions: an assessment by microCT. *Ann Rheum Dis* 2013;72:1351–7.
- Østergaard M, Edmonds J, McQueen F *et al.* An introduction to the EULAR-OMERACT rheumatoid arthritis MRI reference image atlas. *Ann Rheum Dis* 2005;64 (Suppl 1):i3–7.
- Bird P, Conaghan P, Ejlertsen B *et al.* The development of the EULAR-OMERACT rheumatoid arthritis MRI reference image atlas. *Ann Rheum Dis* 2005;64 (Suppl 1):i8–10.
- Østergaard M, Peterfy C, Conaghan P *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.
- Ejlertsen B, McQueen F, Lassere M *et al.* The EULAR-OMERACT rheumatoid arthritis MRI reference image atlas: the wrist joint. *Ann Rheum Dis* 2005;64 (Suppl 1):i23–47.
- McQueen FM, Benton N, Perry D *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.
- Freeston JE, Bird P, Conaghan PG. The role of MRI in rheumatoid arthritis: research and clinical issues. *Curr Opin Rheumatol* 2009;21:95–101.
- Conaghan PG, Durez P, Alten RE *et al.* Impact of intravenous abatacept on synovitis, osteitis and structural damage in patients with rheumatoid arthritis and an inadequate response to methotrexate: the ASSET randomised controlled trial. *Ann Rheum Dis* 2013;72:1287–94.
- Kosta PE, Voulgari PV, Zikou AK, Drosos AA, Argyropoulou MI. The usefulness of magnetic resonance imaging of the hand and wrist in very early rheumatoid arthritis. *Arthritis Res Ther* 2011;13:R84.

- 25 Kosta PE, Voulgari PV, Zikou AK *et al.* Effect of very early treatment in rheumatoid arthritis on bone oedema and synovitis, using magnetic resonance imaging. *Scand J Rheumatol* 2012;41:339–44.
- 26 Lisbona MP, Maymo J, Perich J *et al.* Etanercept reduces synovitis as measured by magnetic resonance imaging in patients with active rheumatoid arthritis after only 6 weeks. *J Rheumatol* 2008;35:394–7.
- 27 Lisbona MP, Maymo J, Perich J, Almirall M, Carbonell J. Rapid reduction in tenosynovitis of the wrist and fingers evaluated by MRI in patients with rheumatoid arthritis after treatment with etanercept. *Ann Rheum Dis* 2010;69:1117–22.
- 28 Wieners G, Detert J, Streitparth F *et al.* High-resolution MRI of the wrist and finger joints in patients with rheumatoid arthritis: comparison of 1.5 Tesla and 3.0 Tesla. *Eur Radiol* 2007;17:2176–82.
- 29 Brown AK, Quinn MA, Karim Z *et al.* Presence of significant synovitis in rheumatoid arthritis patients with disease-modifying antirheumatic drug-induced clinical remission: evidence from an imaging study may explain structural progression. *Arthritis Rheum* 2006;54:3761–73.
- 30 Genovese MC, Kavanaugh A, Weinblatt ME *et al.* An oral Syk kinase inhibitor in the treatment of rheumatoid arthritis: a three-month randomized, placebo-controlled, phase II study in patients with active rheumatoid arthritis that did not respond to biologic agents. *Arthritis Rheum* 2011;63:337–45.
- 31 Syversen SW, Haavardsholm EA, Bøyesen P *et al.* Biomarkers in early rheumatoid arthritis: longitudinal associations with inflammation and joint destruction measured by magnetic resonance imaging and conventional radiographs. *Ann Rheum Dis* 2010;69:845–50.
- 32 Conaghan PG, Emery P, Østergaard M *et al.* Assessment by MRI of inflammation and damage in rheumatoid arthritis patients with methotrexate inadequate response receiving golimumab: results of the GO-FORWARD trial. *Ann Rheum Dis* 2011;70:1968–74.
- 33 Narváez J, Narváez JA, de Albert M, Gomez-Vaquero C, Nolla JM. Can magnetic resonance imaging of the hand and wrist differentiate between rheumatoid arthritis and psoriatic arthritis in the early stages of the disease? *Semin Arthritis Rheum* 2012;42:234–45.
- 34 Østergaard M, Emery P, Conaghan PG *et al.* Significant improvement in synovitis, osteitis, and bone erosion following golimumab and methotrexate combination therapy as compared with methotrexate alone: a magnetic resonance imaging study of 318 methotrexate-naive rheumatoid arthritis patients. *Arthritis Rheum* 2011;63:3712–22.
- 35 Haavardsholm EA, Østergaard M, Ejbjerg BJ *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.
- 36 Cyteval C, Miquel A, Hoa D *et al.* Rheumatoid arthritis of the hand: monitoring with a simplified MR imaging scoring method—preliminary assessment. *Radiology* 2010;256:863–9.
- 37 Haavardsholm EA, Østergaard M, Hammer HB *et al.* Monitoring anti-TNF $\alpha$  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.
- 38 Marzo-Ortega H, Tanner SF, Rhodes LA *et al.* Magnetic resonance imaging in the assessment of metacarpophalangeal joint disease in early psoriatic and rheumatoid arthritis. *Scand J Rheumatol* 2009;38:79–83.
- 39 Schirmer C, Scheel AK, Althoff CE *et al.* Diagnostic quality and scoring of synovitis, tenosynovitis and erosions in low-field MRI of patients with rheumatoid arthritis: a comparison with conventional MRI. *Ann Rheum Dis* 2007;66:522–9.
- 40 Østergaard M, Conaghan PG, O'Connor P *et al.* Reducing invasiveness, duration, and cost of magnetic resonance imaging in rheumatoid arthritis by omitting intravenous contrast injection – does it change the assessment of inflammatory and destructive joint changes by the OMERACT RAMRIS? *J Rheumatol* 2009;36:1806–10.
- 41 Cohen SB, Dore RK, Lane NE *et al.* Denosumab treatment effects on structural damage, bone mineral density, and bone turnover in rheumatoid arthritis: a twelve-month, multicenter, randomized, double-blind, placebo-controlled, phase II clinical trial. *Arthritis Rheum* 2008;58:1299–309.
- 42 Yu X, Wang L, Xu P *et al.* Effects of double filtration plasmapheresis, leflunomide, and methotrexate on inflammatory changes found through magnetic resonance imaging in early rheumatoid arthritis. *J Rheumatol* 2012;39:1171–8.
- 43 Quinn MA, Conaghan PG, O'Connor PJ *et al.* Very early treatment with infliximab in addition to methotrexate in early, poor-prognosis rheumatoid arthritis reduces magnetic resonance imaging evidence of synovitis and damage, with sustained benefit after infliximab withdrawal: results from a twelve-month randomized, double-blind, placebo-controlled trial. *Arthritis Rheum* 2005;52:27–35.
- 44 Keystone EC, *et al.* A faster clinical response to Certolizumab Pegol (CZP) treatment is associated with better 52-week outcomes in patients with Rheumatoid Arthritis (RA). *Arthritis Rheum* 2010;60:S372.
- 45 Durez P, Malghem J, Nzeusseu Toukap A *et al.* Treatment of early rheumatoid arthritis: a randomized magnetic resonance imaging study comparing the effects of methotrexate alone, methotrexate in combination with infliximab, and methotrexate in combination with intravenous pulse methylprednisolone. *Arthritis Rheum* 2007;56:3919–27.
- 46 Döhn UM, Conaghan PG, Eshed I *et al.* The OMERACT-RAMRIS rheumatoid arthritis magnetic resonance imaging joint space narrowing score: intrareader and interreader reliability and agreement with computed tomography and conventional radiography. *J Rheumatol* 2014;41:392–7.
- 47 McQueen FM. The MRI view of synovitis and tenosynovitis in inflammatory arthritis: implications for diagnosis and management. *Ann N Y Acad Sci* 2009;1154:21–34.
- 48 Quinn MA, Conaghan PG, O'Connor PJ *et al.* Very early treatment with infliximab in addition to methotrexate in early, poor-prognosis rheumatoid arthritis reduces magnetic resonance imaging evidence of synovitis and damage, with sustained benefit after infliximab withdrawal: results from a twelve-month randomized,

- double-blind, placebo-controlled trial. *Arthritis Rheum* 2005;52:27–35.
- 49 Hodgson RJ, O'Connor P, Moots R. MRI of rheumatoid arthritis image quantitation for the assessment of disease activity, progression and response to therapy. *Rheumatology* 2008;47:13–21.
- 50 Lassere M, McQueen F, & Østergaard M *et al.* OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 3: an international multicenter reliability study using the RA-MRI Score. *J Rheumatol* 2003;30:1366–75.
- 51 Conaghan P, Lassere M, Østergaard M *et al.* OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 4: an international multicenter longitudinal study using the RA-MRI Score. *J Rheumatol* 2003;30:1376–9.
- 52 Østergaard M, Klarlund M, Lassere M *et al.* Interreader agreement in the assessment of magnetic resonance images of rheumatoid arthritis wrist and finger joints—an international multicenter study. *J Rheumatol* 2001;28:1143–50.
- 53 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.
- 54 Suter LG, Fraenkel L, Braithwaite RS. Role of magnetic resonance imaging in the diagnosis and prognosis of rheumatoid arthritis. *Arthritis Care Res* 2011;63:675–88.
- 55 American College of Rheumatology Rheumatoid Arthritis Clinical Trials Task Force Imaging Group and Outcome Measures in Rheumatology Magnetic Resonance Imaging Inflammatory Arthritis Working Group. Review: the utility of magnetic resonance imaging for assessing structural damage in randomized controlled trials in rheumatoid arthritis. *Arthritis Rheum* 2013;65:2513–23.
- 56 Baker JF, Østergaard M, Emery P *et al.* Early MRI measures independently predict 1-year and 2-year radiographic progression in rheumatoid arthritis: secondary analysis from a large clinical trial. *Ann Rheum Dis* 2014;73:1968–74.
- 57 Peterfy CG, DiCarlo JC, Olech E *et al.* Evaluating joint-space narrowing and cartilage loss in rheumatoid arthritis by using MRI. *Arthritis Res Ther* 2012;14:R131.
- 58 Eshed I, Feist E, Althoff CE *et al.* Tenosynovitis of the flexor tendons of the hand detected by MRI: an early indicator of rheumatoid arthritis. *Rheumatology* 2009;48:887–91.
- 59 Navalho M, Resende C, Rodrigues AM *et al.* Bilateral MR imaging of the hand and wrist in early and very early inflammatory arthritis: tenosynovitis is associated with progression to rheumatoid arthritis. *Radiology* 2012; 264:823–33.
- 60 Peterfy C, Emery P, Tak PP *et al.* MRI assessment of suppression of structural damage in patients with rheumatoid arthritis receiving rituximab: results from the randomised, placebo-controlled, double-blind RA-SCORE study. *Ann Rheum Dis* 2016;75:170–7.
- 61 Østergaard M, Jacobsson LT, Schaufelberger C *et al.* MRI assessment of early response to certolizumab pegol in rheumatoid arthritis: a randomised, double-blind, placebo-controlled phase IIIb study applying MRI at weeks 0, 1, 2, 4, 8 and 16. *Ann Rheum Dis* 2015; 74:1156–63.