

# Interreader Agreement in the Assessment of Magnetic Resonance Images of Rheumatoid Arthritis Wrist and Finger Joints — An International Multicenter Study

MIKKEL ØSTERGAARD, METTE KLARLUND, MARISSA LASSERE, PHILIP CONAGHAN, CHARLES PETERFY, FIONA McQUEEN, PHIL O'CONNOR, RON SHNIER, NEIL STEWART, DENNIS McGONAGLE, PAUL EMERY, HARRY GENANT, and JOHN EDMONDS

**ABSTRACT.** Magnetic resonance imaging (MRI) allows direct visualization of inflammation and destruction in rheumatoid arthritis (RA) joints. However, MRI scoring methods have not yet been standardized or appropriately validated. Our aim was to examine interreader agreement for a simple system of scoring RA changes on MRI among 5 centers that had not undertaken intergroup calibration. MRI of RA wrist and metacarpophalangeal (MCP) joints were scored by experienced readers in 5 centers in different countries. In substudy 1, 5 sets of 2nd–5th MCP joints from UK [Technique A: 1.5 T, coronal and axial T1 and T2 spin-echo, –/+ fat saturation (FS), –/+ iv gadolinium (Gd)] were scored for synovitis (score 0–3) and bone lesions (0–3). In substudy 2, we evaluated 19 sets of 2nd–5th MCP joints [10 sets from UK (Technique A) and 9 sets from the US (Technique B: 1.5 T; coronal T1 spin-echo and T2\* gradient-echo + FS, no Gd)] and 19 wrist joints [9 from the US (Technique B) and 10 from Denmark (Technique C: 1.0 T; coronal and axial T1 spin-echo, no FS, –/+ Gd)]. Synovitis (0–3), bone lesions (0–3), and joint space narrowing (JSN, 0–3) were scored in each MCP joint and in 3 different regions of the wrist. Bone erosions and lesions in each bone were scored 0–5. Substudy 1 served to test and redesign the score sheets. In substudy 2, the scores of synovitis and bone lesions by the 5 groups were the same or differed by only one grade in 73% and 85% of joints, respectively. On MRI that included 2 imaging planes and iv Gd (Techniques A and C), these rates were 86% (synovitis) and 97% (bone lesions). Corresponding intraclass correlation coefficients (quadratic weighted kappas) were 0.44–0.68, mean 0.58 (synovitis), and 0.44–0.69, mean 0.62 (bone lesion), i.e., in the moderate to good range. Unweighted kappa values were in the low to moderate range, generally lowest for JSN (< 0.20), better for synovitis and bone erosions, and best for bone lesions, being generally highest for MRI with 2 planes pre- and post-Gd and in MCP joints compared with wrists. These preliminary results suggest that the basic interpretation of MRI changes in RA wrist and MCP joints is relatively consistent among readers from different countries and medical backgrounds, but that further training, calibration, and standardization of imaging protocols and grading schemes will be necessary to achieve acceptable intergroup reproducibility in assessing synovitis and bone destruction in RA multicenter studies. (*J Rheumatol* 2001;28:1143–50)

## Key Indexing Terms:

MAGNETIC RESONANCE IMAGING  
WRIST FINGER JOINTS

RHEUMATOID ARTHRITIS  
OUTCOME ASSESSMENT

From the Department of Rheumatology and Danish Research Centre of Magnetic Resonance, Hvidovre Hospital, University of Copenhagen; Department of Rheumatology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark; Rheumatology Research Unit, University of Leeds, Leeds, UK; Department of Radiology, University of California San Francisco; Synarc Inc., San Francisco, CA, USA; Department of Rheumatology, St. Georges Hospital, University of NSW, Sydney, Australia; Department of Molecular Medicine, Auckland School of Medicine, University of Auckland, Auckland, New Zealand; Department of Radiology, Leeds General Infirmary, Leeds, UK; and the Sydney Imaging Group, Sydney, Australia.

M. Østergaard, MD, PhD, DMSc, Senior Registrar in Rheumatology, Department of Rheumatology, Rigshospitalet, and Danish Research Centre of Magnetic Resonance, Hvidovre Hospital; M. Klarlund, MD, PhD, Senior Registrar in Rheumatology, Department of Rheumatology, Hvidovre Hospital; M. Lassere, MB, BS, PhD, FRACP, FAFPHM, Staff Specialist in Rheumatology, Department of Rheumatology, St. Georges Hospital, University of NSW; P. Conaghan, MB, BS, FRACP, Senior Lecturer in Rheumatology, Rheumatology Research Unit, University of Leeds; C. Peterfy, MD, PhD, Chief Medical Officer, Synarc Inc., Assistant

Clinical Professor of Radiology, University of California San Francisco; F. McQueen, MD, FRACP, Senior Lecturer in Rheumatology, Department of Molecular Medicine, Auckland School of Medicine, University of Auckland; P. O'Connor, MB, BS, MRCP, FRCR, Consultant Skeletal Radiologist, Department of Radiology, Leeds General Infirmary; R. Shnier, MB, BS, FRACR, National Director of Diagnostic Imaging, Mayne Nickless Sydney Imaging Group; N. Stewart, MBChB, FRACP, Consultant Musculoskeletal Radiologist, University of Auckland; D. McGonagle, MB, BCh, BAO, MRCP, Lecturer in Rheumatology Rheumatology Research Unit, University of Leeds; P. Emery, MA, MD, FRCP, ARC Professor in Rheumatology, Rheumatology Research Unit, University of Leeds; H. Genant, MD, MD, FACR, FRCR, Professor of Radiology, Medicine and Orthopaedics, Department of Radiology, University of California San Francisco; J. Edmonds, MB, BS, FRACP, Professor of Rheumatology, Department of Rheumatology, St. Georges Hospital, University of NSW.

Address reprint requests to Dr. M. Østergaard, Danish Research Centre of Magnetic Resonance, Hvidovre Hospital, University of Copenhagen, Kettegaard Alle 30, DK-2650 Hvidovre, Denmark. E-mail: mo@dadlnet.dk

The potential of magnetic resonance imaging (MRI) to reveal early inflammatory and destructive change in RA has attracted considerable attention. Recent studies have found MRI superior to conventional radiography and clinical examination in its ability to reveal soft tissue and bone changes in knees<sup>1,2</sup>, wrists<sup>3,4</sup>, and finger joints<sup>5,6</sup>.

Several methods for qualitative, semiquantitative, and quantitative assessment of changes in cartilage, juxtaarticular bone, and particularly synovium have been suggested<sup>1,5,7-11</sup>. However, MRI scoring methods have not yet been standardized or appropriately validated. Our aim was to examine the interreader agreement among 5 centers that had not undertaken intergroup calibration for a simple MRI system of scoring inflammatory and destructive changes in the wrists and metacarpophalangeal (MCP) joints of patients with rheumatoid arthritis (RA).

## MATERIALS AND METHODS

**Design and patients.** In 2 substudies, MR images of RA wrists and/or metacarpophalangeal (MCP) joints were scored in 5 centers in different countries, followed by analysis of agreement. The 5 centers were Auckland, New Zealand (FM, NS), Hvidovre, Denmark (MK, MØ), Leeds, UK (PC, PO, DM, PE), San Francisco, USA (CP, HG), and Sydney, Australia (ML, RS, JE).

In Substudy 1 (June–November 1999), MR images of 2nd–5th MCP joints of 5 patients with RA from Leeds, UK, were evaluated. In substudy 2 (November 1999–May 2000), MR images of 10 sets of 2nd–5th MCP joints from Leeds (substudy 2A), 9 sets of wrist and 2nd–5th MCP joints from San Francisco (Substudy 2B), and 10 wrists from Hvidovre (Substudy 2C) were evaluated. All patients fulfilled the American College of Rheumatology 1987 criteria for RA<sup>12</sup>.

**MR imaging. Substudy 1 and 2A (Technique A).** MR images were obtained on a 1.5 T Philips MR Unit. Coronal and axial T1 weighted spin-echo images and coronal fat saturated T2 weighted images were obtained before intravenous (iv) gadolinium-DTPA contrast injection. After the contrast injection, the axial T1 weighted sequence was repeated and a fat saturated coronal T1 weighted sequence was obtained (details in Table 1).

**Substudy 2B (Technique B).** A 1.5 T General Electric Signa MR unit was used. Coronal T2\* weighted 3D gradient-echo and coronal T1 weighted

spin-echo sequences were obtained (Table 1). No contrast agent was administered. Substudy 2C (Technique C): A 1.0 T Siemens Impact MR unit was used to obtain contiguous axial and coronal T1 weighted spin-echo MR images before and after iv Gd-DTPA (Table 1). No fat saturated images were obtained. Representative examples of images by the different MRI techniques are shown in Figure 1.

**Development of MRI scoring method and analysis of images.** The scoring method for substudy 1 was designed by consensus among the investigators at a meeting in Glasgow, UK, June 1999. Based on the experiences from substudy 1, the scoring method was modified for substudy 2 at a meeting in Boston, USA, November 1999. Images were circulated by mail as hard-copy films. This limited viewing of the 3D sequences in substudy 2C to the original coronal plane. All images were scored independently in each of the 5 centers and the scores were recorded on paper score sheets. Readers were rheumatologists or radiologists, who were all experienced in MRI interpretation of RA joints.

**Scoring of MR images. Substudy 1.** Bone lesions, defined as any observable abnormality of the bone morphology or signal, were scored as absent (0) or present (1) in each MCP joint quadrant. Subsequently, a bone global score (0–3) was assigned each MCP joint as a whole. Synovitis was assessed in each MCP joint as a global score from 0 to 3, based on the size of areas with post-gadolinium enhancement. The observers first scored the coronal images alone. Subsequently the axial images were included and the observers noted whether this changed their scores.

**Substudy 2.** MCP joints: As described in Substudy 1, a synovitis global score (0–3) and a bone global score (0–3) were assigned to each MCP joint. This was supplemented with a joint space narrowing score (0–3). Further, a bone erosion score (0–5) and a bone lesion score (0–5) were assigned to the proximal and distal half of each MCP joint. Bone lesions included all bone abnormalities, as in substudy 1, while bone erosions were defined as sharply margined defects with a cortical break. Scores 0–5 referred to the percent involvement of the immediate (first 5 mm) subarticular bone: 0: 0%, 1: 1–20%, 2: 21–40%, 3: 41–60%, 4: 61–80%, 5: 81–100%.

**Wrist joints.** Global readings included 3 variables scored from 0 to 3 in three different regions: Synovitis global score (0–3, separately in the carpus, the first carpometacarpophalangeal joint, and the distal radioulnar joint), joint space narrowing score (0–3, separately in the radiocarpal joint, the midcarpal joint, and the carpometacarpal joint), and bone global score (0–3, separately in the carpus, the base of the metacarpals, and the radius/ulna).

Further, bone lesions and bone erosions were scored 0–5 according to

Table 1. MRI sequences, by center.

Center; MRI unit	Sequence	Plane	Fat Sat	Gd	TR, ms	TE, ms	ST, mm	Gap, mm	FOV, mm	Matrix	Time, min
Technique A (Leeds, UK); Philips Gyroscan 1.5 T											
	T1-SE	Cor	–	–	485	20	1.5	0.1	100	256×256	4.10
	T2-TSE-SPiR	Cor	+	–	2000	100	2.0	0.2	100	198×256	3.44
	T1-SE	Ax	–	–/+	485	20	1.5	0.1	100	205×256	3.21
	T1-SPiR	Cor	+	+	450	20	1.5	0.1	100	192×256	8.44
Technique B (San Francisco, USA); General Electric Signa 1.5 T											
	T1-SE	Cor	–	–	600	9	3.0	0.0	120	192×512	3.54
	3D-FGR	Cor	+	–	29.4	6.3	1.5	0.0	120	192×512	6.02
(flip = 20°)											
Technique C (Hvidovre, Denmark); Siemens Impact 1.0 T											
	T1-SE	Cor	–	–/+	600	15	3.0	0.0	140	192×256	3.53
	T1-SE	Ax	–	–/+	600–700	15	3.0	0.0	120	192×256	3.53–4.32

T1-SE: T1-weighted spin echo sequence; T2-TSE-SPiR: T2-weighted fat saturated turbo spin echo sequence; T1-SPiR: T1-weighted fat saturated sequence; 3D-FGR: T2\*-weighted 3-dimensional fast gradient echo sequence; Cor: coronal; Ax: axial; Fat sat: fat saturation; Gd: iv gadolinium-DTPA contrast; TR: repetition time; TE: echo time; ST: slice thickness; Gap: interslice gap; FOV: field of view; flip: flip angle; T: tesla.



Figure 1. Representative examples of images by different MRI techniques. Figure A–C. Technique A (Leeds, UK): Coronal images of 3rd–4th MCP joint by (A) T1 weighted spin-echo; (B) T2 weighted fat saturated turbo spin-echo; (C) T1 weighted fat saturated after iv Gd. By centers, scores of selected variables were as follows: Bone global scores: 3rd MCP: 2, 2, 1, 2, 1; 4th MCP: 0, 0, 0, 0, 1. Synovitis global scores: 3rd MCP: 3, 3, 3, 3, 2; 4th MCP: 0, 0, 0, 1, 0. Figure D–E. Technique B (San Francisco, USA): coronal images of 2nd and 3rd MCP joint by (D) T1 weighted spin-echo; (E) T2\* weighted 3-dimensional fast gradient echo. Bone global scores, 3rd MCP: 2, 2, 1, 1, 1. Synovitis global scores, 3rd MCP: 2, 3, 3, 3, 2. Joint space narrowing scores, 3rd MCP: 2, 0, 0, 0, 0. Figure F–I. Technique C (Hvidovre, DK): Coronal and axial T1 weighted spin images of the wrist before (F and H) and after (G and I) iv Gd. Bone global scores, carpus: 2, 2, 2, 2, 1. Synovitis global scores, carpus: 3, 3, 3, 2, 2.



percent involvement of subarticular bone, as described above, in each bone of the wrist (metacarpal bases 1–5, the carpal bones, and radius and ulna). *Statistical methods.* Descriptive statistics (mean, range, standard deviations) by group and per joint and/or joint area were calculated and single factor repeated measures analyses of variance were used to compare scores by different groups. For assessment of interreader agreement, kappa statistics were employed on all variables. Kappa values, which express agreement beyond chance, can be interpreted as follows: 0.0–0.20: poor, 0.20–0.40: fair, 0.40–0.60: moderate, 0.60–0.80: good, 0.80–1.00: excellent<sup>13</sup>. Kappa values only take into account presence or absence of absolute agreement and not the degree of disagreements among scores<sup>13</sup>. On selected variables, we calculated fixed effects intraclass correlation coefficients (ICC), which are identical to quadratic weighted kappas and which give a measure of the squared difference among the values on an ordinal scale<sup>14</sup>. The mathematics for calculation of multiobserver kappas are explained in<sup>15</sup>. The statistical programs used were Stata 6.0, Agree 1.4, ICC 1.03, and Statistica 5.0.

## RESULTS

*Substudy 1.* Five sets of 2nd–5th MCP joints of patients with RA were scored by the 5 centers. The entire range of the scale of the synovitis global score (possible score: 0–3) was used. Complete agreement was observed in 5 out of 20 cases, with all groups agreeing on the absence of synovitis. In 10 of 20 cases, the scores were divided by one grade and in 5 of 20 MCP joints by two grades. The kappa value was 0.36. Mean scores of the individual groups varied from 0.40 to 1.28 (mean 0.85) (Table 2). Analysis of variance was not performed because of the small number of readings.

Bone global scores of 0, 1, and 2 but not the maximum of

3 were assigned. Observers gave the same score in 8 out of 20 assessments, always with agreement on the score 0 (absence of lesions). Scores were divided by one grade in 10 of 20 joints and by two grades in 2 of 20 joints. The kappa was 0.27. Considerable intergroup differences in scores were found. The mean bone global score of one group was 0.65, while the mean scores of the remaining 4 groups varied from 0.20 to 0.35 (Table 2).

Finally, bone lesion quadrant scores (0–1) were assigned. Total agreement occurred in 62 out of 80 MCP joint quadrants. Of these, the centers agreed on the score 0 (i.e., absence of bone lesion) in 60 quadrants and on the score 1 (i.e., presence of bone lesion) in 2 quadrants. In 14 of 80 quadrants, 4 of 5 centers agreed on the score, while in 4 of 80 quadrants 2 centers scored differently from the 3 others. Overall agreements for each quadrant were 0.91 (proximal ulnar), 0.86 (proximal radial), 0.89 (distal radial), and 0.93 (distal ulnar). The corresponding kappas were 0.05, 0.55, 0.03, and 0.26.

In 4 out of 5 centers, inclusion of axial images, as a supplement to coronal images, led to a change of the scores estimated from coronal images alone in at least one of the 5 patients (range 1–4 patients, mean 2.3).

*Substudy 2.* A total of 19 sets of wrist joints and 19 sets of 2nd–5th MCP joints of patients with RA were scored by the 5 centers. Descriptive statistics of selected variables are given in Table 2. Statistically significant differences among

Table 2. Means and ranges of scores in wrist and MCP joints by different centers.

	AU	UK	DK	NZ	USA	ANOVA
<b>MCP joints</b>						
<b>Substudy 1</b>						
Synovitis global score (0–3)	1.1 (0.0–2.3)	1.3 (0.0–2.0)	0.9 (0.0–1.8)	0.7 (0.0–1.0)	0.4 (0.0–1.0)	Not calculated
Bone global score (0–3)	0.7 (0.5–1.0)	0.2 (0.0–0.5)	0.3 (0.0–0.8)	0.4 (0.0–0.8)	0.2 (0.0–0.5)	Not calculated
<b>Substudy 2</b>						
Synovitis global score (0–3)	1.4 (0–2.8)	1.7 (0.8–2.0)	1.9 (0.5–3.0)	1.3 (0.5–1.8)	1.4 (0.8–3.0)	< 0.001
Bone global score (0–3)	1.3 (0.8–2.3)	1.0 (0.0–2.5)	0.9 (0.0–2.5)	0.7 (0.0–1.3)	0.9 (0.0–2.0)	< 0.0001
Joint space narrowing (0–3)	1.1 (0.0–2.0)	0.6 (0.0–1.3)	—	0.1 (0.0–1.5)	0.5 (0.0–2.0)	< 0.0001
Bone erosion proximal + distal (0–10)	2.5 (0.0–6.0)	1.2 (0.0–3.0)	1.4 (0.0–4.5)	1.2 (0.0–4.0)	1.5 (0.0–5.3)	< 0.0001
Bone lesion proximal + distal (0–10)	1.9 (0.5–3.0)	1.3 (0.0–2.3)	1.5 (0.0–5.0)	1.3 (0.0–3.5)	1.6 (0.5–5.3)	< 0.05
<b>Wrist joints</b>						
<b>Synovitis global score (0–3)</b>						
Carpus	0.7 (0–2)	1.1 (0–3)	1.1 (0–3)	0.6 (0–3)	0.7 (0–3)	< 0.001
1st Carpometacarpal joint	2.0 (0–3)	1.7 (1–3)	2.0 (1–3)	1.7 (0–3)	1.5 (0–3)	NS (0.08)
Radioulnar joint	1.3 (0–3)	1.7 (0–3)	1.7 (0–3)	0.8 (0–3)	1.4 (0–3)	< 0.001
<b>Joint space narrowing (0–3)</b>						
Radiocarpal joint	0.7 (0–3)	1.0 (0–2)	—	1.4 (0–3)	0.5 (0–3)	NS (0.26)
Midcarpal joints	1.0 (0–2)	1.1 (0–3)	—	1.1 (0–3)	0.7 (0–3)	NS (0.25)
Carpometacarpal joints	0.6 (0–2)	0.4 (0–2)	—	0.1 (0–1)	0.3 (0–2)	< 0.05
<b>Bone global score (0–3)</b>						
Radius + Ulna	1.2 (0–3)	1.2 (0–2)	1.4 (0–3)	1.1 (0–2)	1.2 (0–3)	NS (0.31)
Carpus	1.7 (1–3)	1.6 (1–3)	1.8 (0–3)	1.5 (0–3)	1.3 (1–3)	< 0.01
Metacarpal bases	1.0 (0–2)	0.7 (0–2)	1.3 (0–2)	0.7 (0–2)	0.8 (0–2)	< 0.01

Values in the 2nd to 6th column are means; ranges are in parentheses. Values in the 7th (right) column are p values.

ANOVA: analysis of variance; NS: no statistically significant difference. Not calculated: the statistical test was not performed due to a low number of patients.

AU: Australia, UK: United Kingdom, DK: Denmark, NZ: New Zealand, USA: the United States.

centers were observed with respect to several variables. Marked differences were found particularly with respect to joint space narrowing, in which the mean scores varied by up to a factor 10 among groups (Table 2). Joint space narrowing was not scored by the Hvidovre center (all images) and the Auckland center (Hvidovre images), because the observers considered the images unsuitable for this purpose.

Kappa values of interreader agreement on all variables are given in Table 3. Kappa values for synovitis global score were 0.12–0.36, mean 0.24 (i.e., poor to fair, according to Altman<sup>13</sup>), for bone global scores 0.23–0.48, mean 0.36 (fair–moderate), for joint space narrowing 0.01–0.21, mean 0.08 (poor), for bone erosions 0.14–0.48, mean 0.36 (poor–moderate), and for bone lesions 0.04–0.62, mean 0.34 (poor–good). Thus, kappa values for joint space narrowing were poor, while kappas for the other variables varied from site to site, generally with bone lesion scores showing the highest level of agreement.

Kappa values for the images derived with each of the different MRI techniques are noted in Table 3. With respect to the majority of variables, kappas were highest on images obtained in 2 planes with iv gadolinium (Techniques A and C). For these images, ICC of global scores were also calculated (Table 3, values given in square brackets). ICC for synovitis global scores were 0.44–0.68, mean 0.58; for bone global scores 0.44–0.69, mean 0.62; and for joint space narrowing 0.01–0.77, mean 0.30. The inability of kappa values to give a reliable expression of the level of agreement is illustrated in Table 4.

In wrist joints and MCP joints, synovitis global and bone global were scored identical or within one grade by all 5 groups in 73% and 85% of joints, respectively. If only MRI scans that included 2 imaging planes and the use of iv gadolinium (Techniques A and C) were used, the corresponding agreement rates were 86% (synovitis) and 97% (bone).

## DISCUSSION

There is increasing evidence that MRI is capable of imaging both synovitis and bone damage in rheumatoid arthritis and of detecting change in damage more quickly than conventional radiography<sup>16–20</sup>. A number of methods for assessment of arthritic involvement of synovium, cartilage, and juxta-articular bone have been introduced, but these have never been validated outside the respective groups.

Our study is the first multicenter study to test the interreader agreement on MR images of RA joints. The study was performed by an OMERACT MRI study group with expertise in MRI in RA and in scoring methodology. Quantitative methods may require dedicated MRI equipment and/or dedicated image processing software and are often time consuming. Semiquantitative methods (scoring) are more easily applicable in clinical trials and practice and

were, accordingly, the subject of these first efforts. The current state of knowledge does not yet justify an attempt to provide, for general use, a definitive system for scoring MRI change in RA. Consequently, our aim was rather to assess whether the basic interpretation of RA joint pathology was comparable among experienced readers of MRI in RA from different countries and medical backgrounds. Second, we wanted to assess the intrinsic (without previous intergroup training or calibration) interreader agreement for assessing inflammatory and destructive changes in RA wrists and MCP joints.

The first exercise, substudy 1, mainly served as a pilot study to test and redesign score sheets, since only limited conclusions can be drawn on a sample limited to 5 patients. Only synovitis and bone lesions, defined as any observable abnormality, were scored. The trend was a fairly high agreement on absence of disease, while grading the abnormalities appeared more difficult. Most patients had relatively mild disease, particularly with respect to bone changes. A broader range of disease severities could have made it easier to define the extreme scores of the scale. Assessment of axial images, as a supplement to coronal images, changed some of the scores obtained from coronal images alone, suggesting a value of biplanar imaging.

In substudy 2 we increased the number of patients to be assessed, the number of abnormalities to be scored, and the range of disease severity. Assessment of wrist joints and scoring of bone erosions and joint space narrowing were included. On the other hand, scoring of MCP joint quadrants was replaced by scoring only proximal and distal halves in order to avoid the problems with scoring changes at borders between radial and ulnar quadrants that were observed in substudy 1.

Different methods were used to assess interobserver agreement. Kappa values, which express agreement beyond chance, can be interpreted as follows: 0.0–0.20: poor, 0.20–0.40: fair, 0.40–0.60: moderate, 0.60–0.80: good, 0.80–1.00: excellent<sup>13</sup>. Kappa values of interreader agreement were calculated for all variables (Table 3), resulting in rather low values, at best in the 0.20–0.50 (fair–moderate) range for bone erosions/lesions and synovitis. Kappa values for joint space narrowing were consistently poor (< 0.2). In fact, one center found assessment of joint space narrowing impossible on the images available, and refrained from scoring this variable.

It should be remembered that these kappa values only take into account presence or absence of absolute agreement and not the level of disagreements among scores (“unweighted” kappas). Thus, unweighted kappas are not optimal for assessing ordinal data (Table 4). Consequently, we calculated intraclass correlation coefficients on global scores on Technique A and C images. ICC are identical to quadratic weighted kappas and give a measure of the quadrated difference among the values on an ordinal scale

Table 3. Interreader agreement in substudy 2. Explanation of the 3 values on top of each other in each cell: 1st value (bold face): all images (19 sets of images); 2nd value: Technique A (Leeds) images (MCP joints) or Technique C (Hvidovre) images (wrist joints) only (10 sets of images); 3rd value: Technique B (San Francisco) images only (9 sets of images).

		Kappa Values [intraclass correlation coefficients]					Agreement within 1 grade, %
		MCP JOINTS					
		MCP 2	MCP 3	MCP 4	MCP 5	MCP Total	
Synovitis global score, 0–3		<b>0.25</b> 0.30 [0.66] 0.16	<b>0.32</b> 0.28 [0.68] 0.35	<b>0.25</b> 0.23 [0.51] 0.10	<b>0.12</b> 0.12 [0.44] 0.03	<b>75</b> 90 69	
Bone global score, 0–3		<b>0.33</b> 0.42 [0.69] 0.19	<b>0.46</b> 0.52 [0.68] 0.34	<b>0.38</b> 0.46 [0.59] 0.29	<b>0.23</b> 0.30 [0.44] 0.12	<b>82</b> 98 69	
Joint space narrowing*, 0–3		<b>0.05</b> –0.12 [0.01] 0.15	<b>0.09</b> 0.13 [0.46] 0.03	<b>0.01</b> 0.08 [0.12] –0.01	<b>0.02</b> –0.03 [0.07] 0.07	<b>63</b>	
Bone erosion, 0–5	Proximal	<b>0.15</b> 0.20 [0.66] 0.08	<b>0.44</b> 0.50 [0.80] 0.33	<b>0.31</b> 0.33 [0.55] 0.26	<b>0.19</b> 0.16 [0.39] 0.13		
	Distal	<b>0.16</b> 0.18 [0.30] 0.12	<b>0.35</b> 0.30 [0.58] 0.37	<b>0.18</b> 0.06 [0.02] 0.18	<b>0.21</b> 0.22 [0.40] 0.05		
Bone lesion, 0–5	Proximal	<b>0.31</b> 0.36 [0.63] 0.17	<b>0.39</b> 0.39 [0.70] 0.36	<b>0.31</b> 0.24 [0.39] 0.38	<b>0.29</b> 0.29 [0.39] 0.19		
	Distal	<b>0.40</b> 0.36 [0.66] 0.47	<b>0.29</b> 0.49 [0.54] –0.01	<b>0.12</b> 0.05 [0.07] 0.20	<b>0.61</b> 0.61 [0.74] 0.49		
		WRIST JOINTS					
Synovitis global score, 0–3		Carpus <b>0.15</b> 0.19 [0.50] 0.08	First CMC <b>0.36</b> 0.46 [0.61] 0.23	Radio-ulnar joint <b>0.24</b> 0.24 [0.64] 0.20		Wrist total <b>70</b> 80 59	
Bone global score, 0–3		Carpus <b>0.32</b> 0.37 [0.66] 0.25	Metacarpal bases <b>0.30</b> 0.32 [0.62] 0.27	Radius/ulna <b>0.48</b> 0.59 [0.66] 0.33		<b>89</b> 97 81	
Joint space narrowing*, 0–3		Radiocarpal joint <b>0.06</b> –0.16 [0.08] <b>0.08</b>	Midcarpal joint <b>0.21</b> 0.31 [0.77] 0.12	Carpometacarpal joint <b>0.14</b> 0.21 [0.60] 0.04		<b>80</b>	
Bone erosions, 0–5	Metacarp 1	<b>0.48</b> 0.31 0.62	Metacarp 2 <b>0.22</b> 0.19 0.25	Metacarp 3 <b>0.28</b> 0.28 0.17	Metacarp 4 <b>0.35</b> 0.35 0.35	Metacarp 5 <b>0.35</b> –0.04 0.50	
	Hamate	<b>0.14</b> 0.12 0.15	Capitate <b>0.31</b> 0.31 0.08	Trapezoid <b>0.14</b> 0.23 0.02	Trapezium <b>0.29</b> 0.25 0.11	Triquetrum <b>0.18</b> 0.23 0.10	
	Pisiform	<b>0.31</b> 0.32 0.26	Lunate <b>0.34</b> 0.43 0.18	Scaphoid <b>0.22</b> 0.27 0.12	Distal ulna <b>0.35</b> 0.39 0.25	Distal radius <b>0.31</b> 0.28 0.27	
Bone lesions, 0–5	Metacarp 1	<b>0.63</b> 0.51 0.72	Metacarp 2 <b>0.32</b> 0.38 0.26	Metacarp 3 <b>0.39</b> 0.43 0.33	Metacarp 4 <b>0.41</b> 0.31 0.50	Metacarp 5 <b>0.32</b> 0.29 0.42	
	Hamate	<b>0.39</b> 0.32 0.29	Capitate <b>0.37</b> 0.58 0.12	Trapezoid <b>0.35</b> 0.39 0.41	Trapezium <b>0.15</b> –0.05 0.04	Triquetrum <b>0.04</b> 0.16 0.09	
	Pisiform	<b>0.59</b> 0.62 0.46	Lunate <b>0.48</b> 0.48 0.42	Scaphoid <b>0.34</b> 0.37 0.26	Distal ulna <b>0.38</b> 0.44 0.28	Distal radius <b>0.39</b> 0.42 0.35	

All values in the 2nd–6th column are unweighted kappa values, except those in square brackets, which are intraclass correlation coefficients (ICC = quadratic weighted kappas). ICC were only calculated for selected variables on Techniques A and C images. The values in the 7th (i.e., right) column indicate the percentage of joints that all readers scored identically or maximally divided on 2 adjacent grades. CMC: carpometacarpal joint. \*Only 3–4 centers assessed joint space narrowing.

Table 4. Example of score distributions, kappa and ICC values.

Scores by centers Pat.	Centers					Number of centers at each score			
	AU	UK	DK	NX	USA	Score 0	Score 1	Score 2	Score 3
1	2	1	1	2	1		•••	••	
2	3	3	3	2	2			••	•••
3	2	2	1	1	1		•••	••	
4	0	1	1	1	1	•	••••		
5	2	1	1	2	2		••	•••	
6	2	2	2	2	2			•••••	
7	2	3	3	2	3			••	•••
8	2	2	1	2	1		••	•••	
9	2	2	2	2	1		•	••••	
10	2	1	2	2	2		•	••••	

Scores of synovitis global score on 10 sets of wrist images from Hvidovre, DK. Despite scores being identical or differing by only one point in the scoring system, the kappa value was 0.19, corresponding to “poor”<sup>13</sup>. The quadratic weighted kappa (interclass correlation coefficient, ICC) was 0.50, better reflecting that major disagreements were not found. AU: Australia, UK: United Kingdom, DK: Denmark, NZ: New Zealand, USA: the United States.

(in this case the scores of the different readers)<sup>14</sup>. As it is obviously important whether scores of the same joint by different readers are at adjacent steps or at the ends of the scale, this measure is better suited for analyzing the available data. ICC were roughly 0.15 to 0.30 higher than the corresponding unweighted kappa values, giving mean ICC of synovitis global scores and bone global scores for Technique A and C images of 0.58 and 0.62, respectively. With respect to joint space narrowing scores, both ICC and unweighted kappas were low, indicating that joint space narrowing could not be reliably scored. The fact that agreement within one grade was found in 80–90% of synovitis global scores and 97–98% of bone global scores on the Techniques A and C images (Table 3) shows that major disagreements were rare, particularly concerning bone damage.

As different MRI techniques provide varying qualities for visualization of different types of joint pathology, the selection of magnet type and MRI sequence is important. Bone marrow edema, which may be an important early sign of bone damage<sup>8,11,21</sup>, was easiest to recognize on T2 weighted images with fat suppression. Synovitis could be identified without gadolinium contrast on fat suppressed T2 weighted images, but it was the general impression of the observers that iv gadolinium markedly facilitated the assessment of synovitis. This was supported by better agreement rates. The fact that the readers were familiar with different MRI field strengths, spatial resolutions, and sequences for assessing RA joints may have contributed significantly to the variation of the scores. Training in reading different image types would perhaps have changed the results. Differences in the patients (e.g., predominantly early/mild versus predominantly severe/late changes) between the different centers may also have influenced the scores.

In our study, agreement rates were higher on image sets that included 2 imaging planes and pre- and post-

gadolinium contrast imaging, i.e., the Techniques A and C images, compared with the Technique B images. This, and the fact that assessing 2 planes sometimes changed scores assigned after assessment of only one plane (substudy 1), suggest that imaging in 2 planes may be desirable. It should be mentioned that the images were circulated as hard copies, hindering full appreciation of the 3D sequences in Technique B, which are displayed optimally using a video technique.

From a pathophysiological point of view, assessment of early cartilage damage by MRI would be very interesting. High resolution cartilage visualization by MRI is possible by means of dedicated research equipment<sup>22,23</sup>. However, the quality of wrist and finger joint MR images on most currently available clinical MR units is insufficient for this application, as were the MRI techniques used in the present study. The poor interreader agreement rates for joint space narrowing in our study reflect this. Consequently, we do not recommend that the cartilage on MCP joints or wrist be assessed with the MRI techniques used in this study.

Sharing some basic MRI sequences (a “core set”) would increase the homogeneity and comparability of MRI studies of RA joints. Accordingly, future MRI studies by our research group will explore the following set of MRI sequences: imaging in 2 planes with T1 weighted images before and after gadolinium contrast, plus a T2 weighted fat saturated sequence; or, if this is not available, a STIR sequence. These and other provisional working recommendations are described in detail by Conaghan, *et al*<sup>24</sup>.

The results of our study and of its analysis and discussion at OMERACT 5 indicate accord on the important pathology that can be detected by MRI in RA joints, although better agreement on image acquisition technique and lesion grading is required to allow its confident use in multicenter studies. Without previous reader training and calibration and without precise definitions of lesions to be scored, the

moderate agreement rates found in this study were a logical consequence of its design. However, we purposely avoided the attempt at this early stage to set up very precise definitions of the changes to be scored, as consensus could have been difficult to achieve and because it would have prevented the evaluation of the intrinsic differences in interpretation of inflammation and destruction among the different centers. In fact, the experiences acquired in these exercises have provided a useful basis for ongoing studies, as detailed by Conaghan, *et al*<sup>24</sup>.

In conclusion, these preliminary results suggest that the basic interpretation of MRI changes in RA wrist and MCP joints is relatively consistent among readers from different countries and medical backgrounds, but that further training, calibration, and standardization of imaging protocols and grading schemes will be necessary to achieve acceptable intergroup reproducibility in assessing synovitis and bone destruction in RA.

## REFERENCES

1. Reiser MF, Bongartz GP, Erlemann R, et al. Gadolinium-DTPA in rheumatoid arthritis and related diseases: First results with dynamic magnetic resonance imaging. *Skeletal Radiol* 1989;18:591-7.
2. Kursunoglu-Brahme S, Riccio T, Weisman MH, et al. Rheumatoid knee: Role of gadopentetate-enhanced MR imaging. *Radiology* 1990;176:831-5.
3. Gilkeson G, Polisson R, Sinclair H, et al. Early detection of carpal erosions in patients with rheumatoid arthritis: a pilot study of magnetic resonance imaging. *J Rheumatol* 1988;15:1361-6.
4. Corvetta A, Giovagnoni A, Baldelli S, et al. MR imaging of rheumatoid hand lesions: comparison with conventional radiology in 31 patients. *Clin Exp Rheumatol* 1992;10:217-22.
5. Jevtic V, Watt I, Rozman B, et al. Precontrast and postcontrast magnetic resonance imaging of hand joints in patients with rheumatoid arthritis. *Clin Radiol* 1993;48:176-81.
6. Rominger MB, Bernreuter WK, Kenney PJ, Morgan SL, Blackburn WD, Alarcon GS. MR imaging of the hands in early rheumatoid arthritis: preliminary results. *Radiographics* 1993;13:37-46.
7. Østergaard M, Gideon P, Sørensen K, et al. Scoring of synovial membrane hypertrophy and bone erosions by MR imaging in clinically active and inactive rheumatoid arthritis of the wrist. *Scand J Rheumatol* 1995;24:212-8.
8. McQueen FM, Stewart N, Crabbe J, et al. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals a high prevalence of erosion at four months after symptom onset. *Ann Rheum Dis* 1998;57:350-6.
9. Østergaard M. Magnetic resonance imaging in rheumatoid arthritis. Quantitative methods for assessment of the inflammatory process in peripheral joints [doctoral thesis]. Copenhagen: University of Copenhagen; 1999. *Dan Med Bull* 1999;46:313-44.
10. Klarlund M. Magnetic resonance imaging of wrist and finger joints in rheumatoid arthritis and early unclassified polyarthritis [PhD thesis]. Copenhagen: University of Copenhagen; 2000.
11. McGonagle D, Gibbon W, O'Connor P, et al. An anatomical explanation for good-prognosis rheumatoid arthritis. *Lancet* 1999;353:123-4.
12. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315-24.
13. Altman DG. *Practical statistics for medical research*. 1st ed. London: Chapman and Hall; 1991.
14. Armitage P, Berry G. *Statistical methods in medical research*. 3rd ed. Oxford: Blackwell; 1994.
15. Svanholm H, Starklint H, Gundersen HJ, et al. Reproducibility of histomorphologic diagnoses with special reference to kappa statistics. *APMIS* 1989;97:689-98.
16. Jevtic V, Watt I, Rozman B, et al. Contrast enhanced Gd-DTPA magnetic resonance imaging in the evaluation of rheumatoid arthritis during a clinical trial with DMARDs. A prospective two-year follow-up study on hand joints in 31 patients. *Clin Exp Rheumatol* 1997;15:151-6.
17. McQueen FM, Stewart N, Crabbe J, et al. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals progression of erosions despite clinical improvement. *Ann Rheum Dis* 1999;58:156-63.
18. McGonagle D, Conaghan P, O'Connor P, et al. The relationship between synovitis and bone changes in early untreated rheumatoid arthritis. A controlled magnetic resonance imaging study. *Arthritis Rheum* 1999;42:1706-11.
19. Østergaard M, Hansen M, Stoltenberg M, et al. Magnetic resonance imaging-determined synovial membrane volume as a marker of disease activity and a predictor of progressive joint destruction in the wrists of patients with rheumatoid arthritis. *Arthritis Rheum* 1999;42:918-29.
20. Klarlund M, Østergaard M, Jensen KE, et al. Magnetic resonance imaging, radiography and scintigraphy of the finger joints: one year follow up of patients with early arthritis. *Ann Rheum Dis* 2000;59:521-8.
21. Jevtic V, Watt I, Rozman B, Kos-Golja M, Demšar F, Jarh O. Distinctive radiological features of small hand joints in rheumatoid arthritis and seronegative spondyloarthritis by contrast-enhanced magnetic resonance imaging. *Skeletal Radiol* 1995;24:351-5.
22. Hodgson RJ, Barry MA, Carpenter TA, Hall LD, Hazleman BL, Tyler JA. Magnetic resonance imaging protocol optimization for evaluation of hyaline cartilage in the distal interphalangeal joint of fingers. *Invest Radiol* 1995;30:522-31.
23. Peterfy CG, van Dijke CF, Lu Y, et al. Quantification of the volume of articular cartilage in the metacarpophalangeal joints of the hand: accuracy and precision of three-dimensional MR imaging. *AJR Am J Roentgenol* 1995;37:1-5.
24. Conaghan P, Edmonds J, Emery P, et al. MRI in rheumatoid arthritis: a summary of OMERACT activities, current status and plans. *J Rheumatol* 2001;28:1158-61.