Development and Preliminary Validation of a Magnetic Resonance Imaging Joint Space Narrowing Score for Use in Rheumatoid Arthritis: Potential Adjunct to the OMERACT RA MRI Scoring System

MIKKEL ØSTERGAARD, PERNILLE BØYESEN, IRIS ESHED, FRÉDÉRIQUE GANDJBKHCH, SIRI LILLEGRAVEN, PAUL BIRD, VIOLAINE FOLTZ, ANNELIES BOOSEN, MARISSA LASSERE, KAY-GEERT HERMANN, ALLEN ANANDARAJAH, UFFE MØLLER DØHN, JANE FREESTON, CHARLES G. PETERFY, HARRY K. GENANT, ESPEN A. HAAVARDSHOLM, FIONA M. McQUEEN, and PHILIP G. CONAGHAN

ABSTRACT. Objectives. To develop and validate a magnetic resonance imaging (MRI) method of assessment of joint space narrowing (JSN) in rheumatoid arthritis (RA).

Methods. Phase A: JSN was scored 0–4 on MR images of 5 RA patients and 3 controls at 15 wrist sites and 2nd–5th metacarpophalangeal (MCP) joints by 8 readers (7 once, one twice), using a preliminary scoring system. Phase B: Image review, discussion, and consensus on JSN definition, and revised scoring system. Phase C: MR images of 15 RA patients and 4 controls were scored using revised system by 5 readers (4 once, one twice), and results compared with radiographs [Sharp-van der Heijde (SvdH) method].

Results. Phase A: Intraobserver agreement: intraclass correlation coefficient (ICC) = 0.99; smallest detectable difference (SDD, for mean of readings) = 2.8 JSN units (4.9% of observed maximal score). Interobserver agreement: ICC = 0.93; SDD = 6.4 JSN units (9.9%). Phase B: Agreement was reached on JSN definition (reduced joint space width compared to normal, as assessed in a slice perpendicular to the joint surface), and revised scoring system (0–4 at 17 wrist sites and 2nd–5th MCP: 0: none; 1: 1–33%; 2: 34–66%; 3: 67–99%; 4: ankylosis). Phase C: Intraobserver agreement: ICC = 0.90; SDD = 6.8 JSN units (11.0%). Interobserver agreement: ICC = 0.92 and SDD = 6.2 JSN units (8.7%). The correlation (ICC) with the SvdH radiographic JSN score of the wrist/hand was 0.77. Simplified approaches evaluating fewer joint spaces demonstrated similar reliability and correlation with radiographic scores.

Conclusion. An MRI scoring system of JSN in RA wrist and MCP joints was developed and showed construct validity and good intra- and interreader agreements. The system may, after further validation in longitudinal data sets, be useful as an outcome measure in RA. (J Rheumatol 2011;38:2045–50; doi:10.3899/jrheum.110422)
Magnetic resonance imaging (MRI) is now frequently used as an outcome measure in rheumatoid arthritis (RA) clinical trials. The OMERACT RA MRI score (RAMRIS), evaluating bone erosions, bone edema, and synovitis, is generally used in such trials. Evaluation of joint space narrowing (JSN) was omitted in the early phase of developing the RAMRIS during the late 1990s, because the quality of MR images at that time was insufficient to evaluate cartilage thickness. However, MR images with higher signal and resolution are now available. Cartilage damage is an important aspect of structural joint damage in RA, and a reliable MRI assessment system of JSN, reflecting cartilage loss, would further improve the usefulness of MRI in measuring outcome in RA clinical trials.

The aim of the present initiative was to develop and validate an OMERACT MRI assessment method of JSN in RA, which may be a useful adjunct to the current OMERACT RAMRIS.

MATERIALS AND METHODS
The project, which was undertaken by the OMERACT MRI in Inflammatory Arthritis Task Force, had 3 phases. Phase A (multireader Exercise 1). In a pilot study, coronal T1-weighted gradient echo MR images of unilateral wrist and metacarpophalangeal (MCP) joints from 5 patients with RA (4 women/1 man; median age 64 (range 30–68) yrs) and 3 controls (67% female; median age 33 (20–48) yrs) were scored 0–4 by a primary scoring system for JSN at 15 sites in the wrist and each of the 2nd–5th MCP joints, without reader pre-training or calibration, by 8 readers (7 readers read once, one reader twice), after circulation by DVD. Phase B. After Exercise 1, the group held a half-day meeting in Philadelphia, in October, 2009. Results were presented, and selected images reviewed and discussed. At the meeting and during subsequent Web-based communication and teleconferences, agreement by consensus was achieved on an MRI definition of JSN, a revised scoring system, and reader rules. Phase C (multireader Exercise 2). In this exercise, coronal T1-weighted 3D gradient echo MR images (voxel size 0.4 × 0.4 × 0.4 mm) of unilateral wrist and 2nd–5th MCP joints of a different cohort of 15 RA patients (12 women/3 men; median age 51 (33–78) yrs; disease duration 7 (4–22) yrs) and 4 controls (3 women/one man; median age 36 (34–57) yrs) were scored for JSN by 5 readers (4 readers read once, one reader twice) using the revised score system (21 sites scored 0–4, see Figure 1 for scoring system). Radiographs of the same hand (wrist: 6 sites, MCP joints: 5 sites; PIP: 4 sites) were assessed by one reader according to the Sharp-van der Heijde (SvdH) scoring method for JSN (0: normal; 1: focal or doubtful; 2: generalized, > 50% of original joint space lost; 3: generalized, < 50% of original joint space lost; 4: ankylosis) and bone erosion.

The reader who read MRI images twice was an experienced musculoskeletal radiologist, and MRI results for that reader (second read) were compared with radiographic JSN assessments, per joint space area and per patient. One RA patient did not have complete radiographs available and was excluded from comparisons between MRI and radiographs. Descriptive statistics, intraclass correlation coefficients (ICC; mixed effects, absolute agreement definition; average measure for intraobserver/interobserver agreements, single measure for intraobserver agreement/comparison between MRI and radiographs) and smallest detectable differences (SDD, for mean of readings) were calculated, using SPSS Statistics®, version 17.0.

RESULTS
Phase A: Multireader Exercise 1. The intraobserver and interobserver ICC were 0.99 and 0.93, respectively, while the corresponding SDD were 2.8 JSN units (9.4% of maximal observed score [observed max]) and SDD 6.4 JSN units (4.9% of observed max).

Phase B: Agreement on definition, scoring system, and reader rules. Based on experiences from the small data set in Phase A, image review and discussion, consensus was reached on JSN definition (reduced joint space width compared to normal, as assessed in a slice perpendicular to the joint surface), scoring system (0–4 at 21 sites (17 wrist sites and each of 2nd–5th MCP joints, total range 0–84): 0: no narrowing; 1: focal or mild (< 33%) narrowing; 2: moderate (34%–66%) narrowing; 3: moderate to severe (67%–99%) narrowing; 4: ankylosis), and reader rules (Figure 1).

Phase C: Multireader Exercise 2. MRI vs radiographs, per joint area. The presence of JSN in each individual joint space area on MRI, as assessed by the experienced radiologist, and on radiographs, is shown in Table 1. On MRI, 111 of 294 (38%) areas in the RA patients were registered as having JSN (JSN score ≥ 1 in that specific area; Figure 2). When areas assessed by both MRI and radiography were compared (and trapezium-scaphoid and trapezoid-scaphoid joint spaces considered as one, as in the SvdH method), 54 joint spaces had JSN by MRI and 17 by radiography. Only 2 (1%) areas had JSN by radiography but not by MRI, while 39 (28%) had JSN on MRI but not radiography. MRI and radiography agreed in 99 areas (71%). The median difference between MRI and radiography scores in individual areas was 0, and the numerical difference never exceeded 2.

In healthy controls, JSN was noted at 12 (14%) joint spaces in total on MRI (0 on radiographs). Seven (58%) of these 12 areas were registered in the oldest control person (age 57 yrs).

MRI vs radiography, per subject. MRI JSN scores by the radiologist were median 7 (range 1–62) in RA patients and 2.5 (range 0–7) in healthy controls, whereas corresponding values were 0 (0–22) and 0 (0–0), respectively, for radiographs. It should be remembered that the areas assessed by MRI and radiography were not identical. The MRI JSN sum scores, including various joint space combinations (see below) are provided in Table 2.

The MRI and radiography total JSN scores (MRI\textsubscript{Total} and Xray\textsubscript{SvdH-Total}) were highly correlated (single-measures ICC 0.77, p < 0.001). When MRI and radiographic scores from identical areas were compared, ICC were 0.42 (p < 0.01) for 2nd–5th

The presence of JSN in each individual joint space area on MRI, as assessed by the experienced radiologist, and on radiographs, is shown in Table 1. On MRI, 111 of 294 (38%) areas in the RA patients were registered as having JSN (JSN score ≥ 1 in that specific area; Figure 2). When areas assessed by both MRI and radiography were compared (and trapezium-scaphoid and trapezoid-scaphoid joint spaces considered as one, as in the SvdH method), 54 joint spaces had JSN by MRI and 17 by radiography. Only 2 (1%) areas had JSN by radiography but not by MRI, while 39 (28%) had JSN on MRI but not radiography. MRI and radiography agreed in 99 areas (71%). The median difference between MRI and radiography scores in individual areas was 0, and the numerical difference never exceeded 2.

In healthy controls, JSN was noted at 12 (14%) joint spaces in total on MRI (0 on radiographs). Seven (58%) of these 12 areas were registered in the oldest control person (age 57 yrs).

MRI vs radiography, per subject. MRI JSN scores by the radiologist were median 7 (range 1–62) in RA patients and 2.5 (range 0–7) in healthy controls, whereas corresponding values were 0 (0–22) and 0 (0–0), respectively, for radiographs. It should be remembered that the areas assessed by MRI and radiography were not identical. The MRI JSN sum scores, including various joint space combinations (see below) are provided in Table 2.

The MRI and radiography total JSN scores (MRI\textsubscript{Total} and Xray\textsubscript{SvdH-Total}) were highly correlated (single-measures ICC 0.77, p < 0.001). When MRI and radiographic scores from identical areas were compared, ICC were 0.42 (p < 0.01) for 2nd–5th
MCP joints, and 0.83 (p < 0.001) in wrists (MRI SvdH-wrist vs Xray SvdH-wrist).

MRI, intrareader reliability. The mean (median/range) of readings 1 and 2 by the experienced radiologist was 13 (9/0–61), and 15 (5/0–62), respectively. The SDD (for mean of readings) was 6.8 JSN units, corresponding to 11.0% of the observed maximal score. The single measure intraobserver ICC for MRItotal was 0.90 (Table 2).

MRI, interreader reliability. The mean scores for the 19 subjects ranged from 11–25 for the 5 readers, whereas the median (min/max) ranged from 3–23 (0–10/62–71). The SDD (for mean of readings) was 6.2 JSN units, corresponding to 8.7% of the observed maximal score. The interobserver ICC for MRItotal was 0.92 (Table 2). ICC for different reader pairs are provided in Table 3.

Simplified MRI scoring methods. The performance of various simplified scores was investigated. Table 2 provides intraobserver and interobserver ICC of separate MCP and wrist scorings and scorings of the joint spaces assessed by the SvdH and Sharp-Genant radiographic methods and by 2 further suggestions for simplified scores, assessing 14 and 7 joint spaces in the wrist, respectively. Table 2 explains which joint spaces are included in the different approaches. The score with only 7 assessed areas (MRIJSN7, Table 2) demonstrated high correlation (ICC 0.79) with the total hand radiograph score (Xray SvdH-Total), and very good intra- and interobserver ICC (> 0.90).

DISCUSSION

We describe the first steps in developing and validating an MRI tool for assessment of JSN in RA joints, performed by experts with a mixed radiology/rheumatology background. The construct validity of the developed MRI JSN score was documented by a high correlation with its well validated radiographic counterpart, the SvdH JSN score. Its reliability was supported by high intraobserver and interobserver agreements, as assessed by ICC. Substantial reader variation in absolute score suggests further improvement may be gained from more extensive reader training and calibration in future studies. Simplified approaches evaluating fewer joint spaces demonstrated similar reliability and correlation with radiographic scores, as did the total scores, and may constitute useful, quicker alternatives. However, further testing in longitudinal studies is required.

A major strength of MRI is its tomographic perspective, and the resulting lack of projectional superimposition of bones and joint spaces. First, this allows for more detailed

---

**Figure 1.** The OMERACT MRI joint space narrowing (JSN) scoring system: Score sheet, MRI definition, scoring system, and reader rules. MCP: metacarpophalangeal; CMC: carpometacarpal; TRM: trapezium; TRD: trapezoid; CAP: capitate; HAM: hamate; SCA: scaphoid; LUN: lunatum; TRI: triquetrum; RAD: radius; ULN: ulna; SvdH: Sharp-van der Heijde.
assessment of the areas included in the radiographic scoring systems. Further, it provides the opportunity to assess joint spaces that have been omitted from radiographic scores because of poor visualization. It remains to be determined whether the ability of MRI to assess more joint space areas in the wrist than radiography translates into a higher sensitivity to change.

A limitation of our study is its cross-sectional design. Registration of change over time is key in clinical trials and practice, and further studies are needed to explore the responsiveness and discriminatory capacity of the total MRI JSN score and its simplifications. Regarding feasibility, it should be mentioned that MRI is a more costly and time-consuming method for assessment of JSN than radiography. However, MRI simultaneously provides information about inflammation in synovium and bone that cannot be achieved by radiography, and the MRI sequence used for JSN assessment in this study does not prolong the MRI examination, as it is optimal for assessment of bone erosion.

Other MRI sequences, allowing better delineation of cartilage and separation of the cartilage on the adjoining bone surfaces, are available and may improve performance. Our approach in this study was to use sequences that can be acquired on almost all units, without the requirement of a field strength and field homogeneity that allow robust fat suppression in all parts of the joints. Further, we envisage that JSN assessment will usually be done in conjunction with other RAMRIS assessments (synovitis, bone edema, and erosion) so it is advantageous to use common sequences suitable for all pathologies, to avoid very long MRI examination times.

JSN was registered in some areas in the healthy control population, in accordance with previous studies. The radiologist found JSN in 12 areas in healthy controls, but this JSN never exceeded grade 1. Although the controls had no symptoms, it would be expected that mild JSN can occur, particularly with increasing age. In accordance with this, 7 of 12 areas with JSN in controls were noted in the oldest patient, a 57-year-old woman. Nevertheless, further study of healthy subjects, in different age groups and sex, is warranted.

ICC were calculated per reader pair since 2 readers are the...
The most likely number for clinical trials. The presented values are average measure ICC (Table 3) and reflect the reliability of readings for pairs of readers.

Previous studies have, in agreement with our results, reported that MRI and radiographic assessment of JSN correlate well. The correlation between MRI and radiographic
scores of JSN were higher in wrists than in MCP joints. This may be at least partly explained by the higher mobility between bones in the MCP joints, with the result that these joints, despite stipulating full extension, may be imaged in different degrees of flexion, compromising reliable assessment of JSN. This highlights the need for exact and meticulous positioning of finger joints in the extended position if JSN assessment is planned, preferably using a dedicated splint. It also encourages further work on refining the scoring procedure in joints where full extension is not possible, and/or have subluxation and/or severe destruction.

An MRI scoring system of JSN in RA wrist and MCP joints was developed and showed construct validity and good intra- and interreader reliability. The system may, after further validation in longitudinal data sets, be useful as an outcome measure to assess cartilage damage in RA, and may thereby further improve the usefulness of MRI in RA clinical trials.

REFERENCES


