2002-947-1

OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 5: An International Multicenter Reliability Study Using Computerized MRI Erosion Volume Measurements

PAUL BIRD, BO EJBJERG, FIONA McQUEEN, MIKKEL OSTERGAARD, MARISSA LASSERE, and JOHN EDMONDS

ABSTRACT. Scoring erosions on magnetic resonance imaging (MRI) is one method of estimating damage in patients with rheumatoid arthritis (RA), but it has limitations. The aim of this pilot study was to assess the feasibility and inter-reader reliability of computer assisted erosion volume estimation in patients with RA. Intra-reader and inter-occasion reliability was also assessed, and different slice thicknesses were compared in terms of erosion volume estimation. A 3 mm slice thickness 3D gradient-echo sequence followed by a 1 mm sequence was performed at baseline and repeated within 24 h with metacarpophalangeal (MCP) joints 2 to 5 of the dominant hand included in the field of view. Three readers were instructed to grade MCP 2 and 3 using the OMERACT grading system and then to measure the erosion volume of the same joints using OSIRIS software. The inter-reader reliability of the grading method and the volume method was calculated, as well as the inter-occasion reliability, by comparing results from each reader from baseline to the followup scan. One reader performed repeat volume measurements on 5 patients to assess the intra-reader reliability. Five patients were included in the study. Expressed in terms of intraclass correlation coefficients (ICC), the inter-reader and inter-occasion reliability of the volume method were comparable to the existing OMERACT scoring system, but large systematic differences in volume estimations were found between readers. The intra-reader reliability was excellent. Good correlation was demonstrated between the total erosion scores and the total erosion volumes. For both erosion volumes and erosion scores, 1 mm and 3 mm acquisitions produced variable results between readers, with no clear pattern of underestimation or overestimation for either slice thickness. The volume estimation method was more time consuming, taking roughly 5 times as long as the scoring method. Computerized MRI erosion volume measurements are feasible, with high intra-observer and inter-occasion reliabilities. Despite high ICC, the inter-observer reliability is not sufficient for multicenter use without prior reader training and calibration. The optimal slice thickness was not determined. (J Rheumatol 2003;30:1380–4)

Key Indexing Terms:
MAGNETIC RESONANCE IMAGING EROSION VOLUMES RHEUMATOID ARTHRITIS

From the Department of Rheumatology, St. George Hospital, Sydney, Australia; University of NSW, Sydney, Australia; the Danish Research Center of Magnetic Resonance and Departments of Rheumatology at the Copenhagen University Hospitals at Hvidovre, Herlev and Rigshospitalet, Copenhagen, Denmark; and Department of Rheumatology, Auckland Hospital, Auckland, New Zealand.

P. Bird, BMed(Hons), Grad Cert MRI, FRACP, Research Fellow, Department of Rheumatology, St. George Hospital; B. Ejbjerg, MD, Research Fellow, Danish Research Centre of Magnetic Resonance, Copenhagen University; F. McQueen, MD, FRACP, Senior Lecturer in Rheumatology, Auckland Hospital; M. Østergaard, MD, PhD, DMSc, Professor in Rheumatology/Arthritis, Danish Research Centre of Magnetic Resonance and Departments of Rheumatology at the Copenhagen University Hospitals at Hvidovre, Herlev and Rigshospitalet; M. Lassere, MBBS(Hons), Grad Dip Epi, PhD, FRACP, FAPHM, Senior Lecturer in Medicine, Staff Specialist Rheumatologist; J. Edmonds, MBBS, MA, FRACP, Director and Professor of Rheumatology, Department of Rheumatology, St. George Hospital.

Address reprint requests to Dr. P. Bird, Rheumatology Department, St. George Hospital, Belgrave St., Kogarah NSW 2217, Australia.
E-mail: pcbird@bigpond.com.au

Magnetic resonance imaging (MRI) has enormous potential as an outcome measure in rheumatoid arthritis (RA), offering a multiplane alternative to radiographs in the assessment of joint damage and activity. With respect to joint damage, MRI provides a remarkable sensitivity to bony erosions, detecting erosions earlier than radiographs[1]; and, of equal importance, short term erosion progression on MRI is correlated with longterm radiographic progression[2,3].

Given the large amount of information that MRI provides, one of the challenges is to find the most effective way of harnessing this to provide a meaningful measure that could be applied in clinical trials or in daily practice. Scoring is one method of achieving this goal, and the OMERACT MRI group has been working to develop a RA damage and activity score with adequate standards of validity, reliability, and feasibility to serve as an outcome

measure[4,5]. Scoring, however, has limitations[6], and direct measurement of erosion size has been proposed as an alternative method of quantifying damage[7]. Additionally, one of the important parameters in MRI acquisition is the slice thickness. Thinner slices (e.g., 1 mm) would be expected to be more sensitive in detecting small erosions and less sensitive to partial volume artefacts and therefore may provide more accurate volumes, but the tradeoff is an increase in image analysis time. This study provided an opportunity to assess erosion volume size and to consider the issue of timing, using acquisitions with 2 different slice thicknesses.

The primary aims of this study were to assess the feasibility and inter-reader reliability of computer assisted erosion volume measurement in patients with RA and to compare the results with the existing OMERACT scoring system. The secondary aims were to assess the intra-reader reliability and the inter-occasion reliability of the scoring and volume method, and to compare 3 mm slice thickness acquisitions with 1 mm slice acquisitions in terms of erosion volume size and image analysis time.

Materials and Methods

Study design. The study was structured to assess the inter-reader reliability, inter-occasion reliability, and intra-reader reliability. MRI examinations, under conditions as close as possible to identical, were performed at baseline and again within 24 hours. Each of the MRI examinations was scored for erosion size using the OMERACT MRI scoring system and then for erosion volume by the 3 readers (PB, BE, FM). Readers were not blinded, and reading of MRI studies was in sequence. One reader (BE) performed repeat volume measurements on all 5 patients for acquisition 1 (3 mm and 1 mm slice thickness).

The inter-reader reliability for the MRI measurements was calculated by comparing the initial MR reading with the second MR reading for each MRI examination. The inter-occasion reliability was measured by comparing the readings from the initial MRI with the readings from the MRI performed within 24 hours. The intra-reader reliability for the MRI measurements was calculated by comparing the initial MR reading with the second MR reading for one reader.

The reader results for scoring and volume estimation for the 3 mm and 1 mm acquisitions were compared.

Patients. Five subjects with RA were selected — 3 from Sydney and 2 from Copenhagen.

MRI. A Siemens Magnetom 1.5 T unit was used for the Sydney acquisitions and a Siemens 1.0 T unit was used for the Copenhagen acquisitions. 3D gradient-echo images of the dominant metacarpophalangeal (MCP) joints were performed with the following standardized acquisition parameters: TR/TE 30/12, NEX 1, matrix 256 × 256, FOV 100 mm. Slice thickness was 3 mm for the initial study (voxel size 0.46 mm^2) and then 1 mm (voxel size 0.15 mm^2) for the second study. The scans were repeated within 24 hours with exactly the same parameters. Total imaging time was 7 min for each scan.

Image distribution. MR studies were transferred to compact disk and distributed to 3 readers in 3 centers — Sydney (PB), Auckland (FM), and Copenhagen (BE). Included with the image were specific instructions for the OSIRIS software.

Erosion volumes. Images were transferred to a personal computer for the erosion volume calculations. The axial and coronal images were viewed initially, and an erosion was considered confirmed if the bone defect with sharp margins was present in both planes and breached the bone cortex in at least one plane. The volume calculations were then performed in each of the T1 weighted coronal slices using OSIRIS software[8]. Each erosion was outlined manually in each coronal slice, the erosion area was calculated by the computer software from multiple slices, and this was multiplied by the slice thickness to provide the erosion volume using the following standard formula:

$$Vol_{eros} = \Sigma\ (Ar_{eros} \times ST)$$

where ST is the slice thickness and Ar_{eros} represents the area of the erosion (Figures 1, 2, and 3). Erosion volumes were calculated for the second and third MCP joints. These erosion volumes were summed to provide a total erosion volume for each patient.

Erosion scores. Erosion scores were performed on the initial 3 mm acquisitions by all 3 readers using the OMERACT 5 RA-MRI score (RAMRIS) criteria[9]. The volume results were compared to the erosion scores obtained by the same observer.

Timing. During the scoring and volume estimations in this study the reader recorded the time taken for each image. For the scoring method this interval included arrangement of the images on a radiograph box for reading through to the end of scoring. For the volume estimations, timing encompassed the interval from opening the image on the computer screen to the completion of the volume analyses.

Statistical analysis. Statistical analysis was undertaken using Statistical Program for the Social Sciences Version 10[10]. Intraclass correlation coefficient (ICC) values were calculated for the inter-observer and inter-occasion volumes and scores using a 2 way mixed model with absolute agreement (i.e., fixed effects ICC) and a 95% confidence interval. Pearson's correlation coefficient was utilized to express the relationship between the erosion volumes and grading score.

Results

The total erosion volume per patient ranged from 0 to 0.57 cm^3. The ICC values for the volume method were acceptable for all MRI acquisitions (ICC 0.73–0.87), and these results were comparable to the scoring system (ICC 0.60) (Table 1). However, large systematic differences between volumes obtained by different observers were identified (Table 2). The inter-occasion ICC for the 3 mm and 1 mm slice acquisitions were excellent (ICC 0.86–0.99) and similar to the results obtained using the OMERACT scoring system (ICC 0.94–0.98) (Table 3). The intra-reader ICC for one reader indicated a high level of agreement (ICC 0.93, 0.99).

There was a strong positive correlation (r = 0.81–0.96) between the total erosion volumes and the total erosion scores for all acquisitions. No clear pattern emerged from erosion volume measurements of the 3 mm and 1 mm acquisitions to suggest that either slice thickness provided a consistently larger or smaller erosion volume (Table 3).

The median time per patient for the computerized volume method was 18 min (range 15–40) for the 3 mm acquisitions and 25 min (range 20–45) for the 1 mm acquisitions. The median analysis time per patient for the erosion scoring was considerably shorter — 4 min (range 2–7) for the 3 mm slices and 6 min (range 5–10) for the 1 mm acquisitions.

Discussion

Our study confirms that the quantitative measurement of

Figure 1. Erosions involving 2nd and 3rd proximal MCP joints are identified on coronal image.

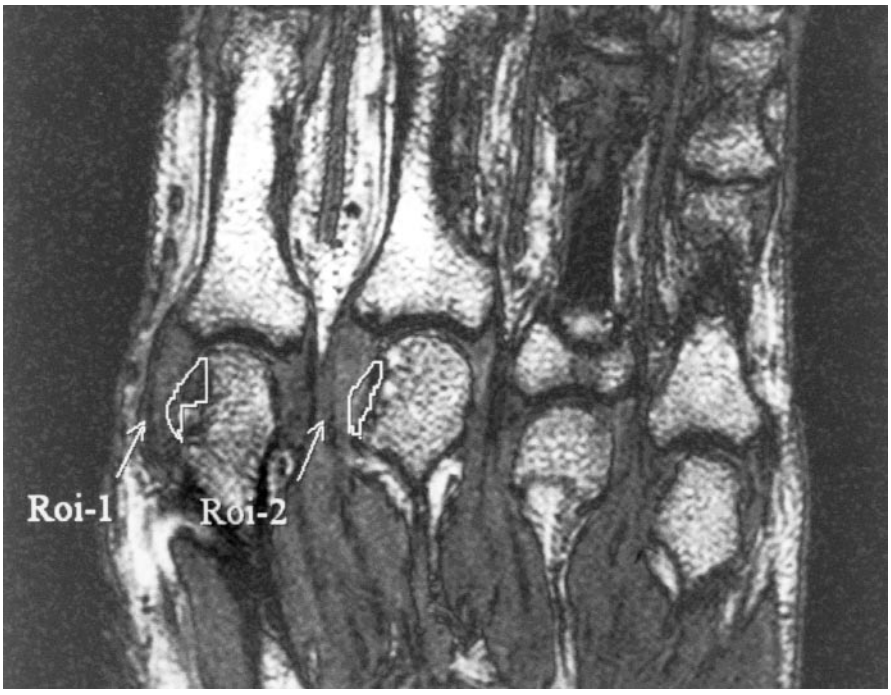

Figure 2. Erosions outlined using OSIRIS imaging software. ROI: region of interest.

MRI erosion volumes in patients with RA is feasible, with high intra-reader and inter-occasion reliabilities. The inter-reader ICC indicated a reasonable level of agreement, comparable to the scoring system, but it should be noted that there were large sytematic differences in volumes between readers. Although this pilot study examined cross sectional data, the underlying hypothesis, as yet untested, is that in principle, computerized erosion volume measurements may be more responsive to change in longitudinal studies.

Feasibility. The software program used for this study was

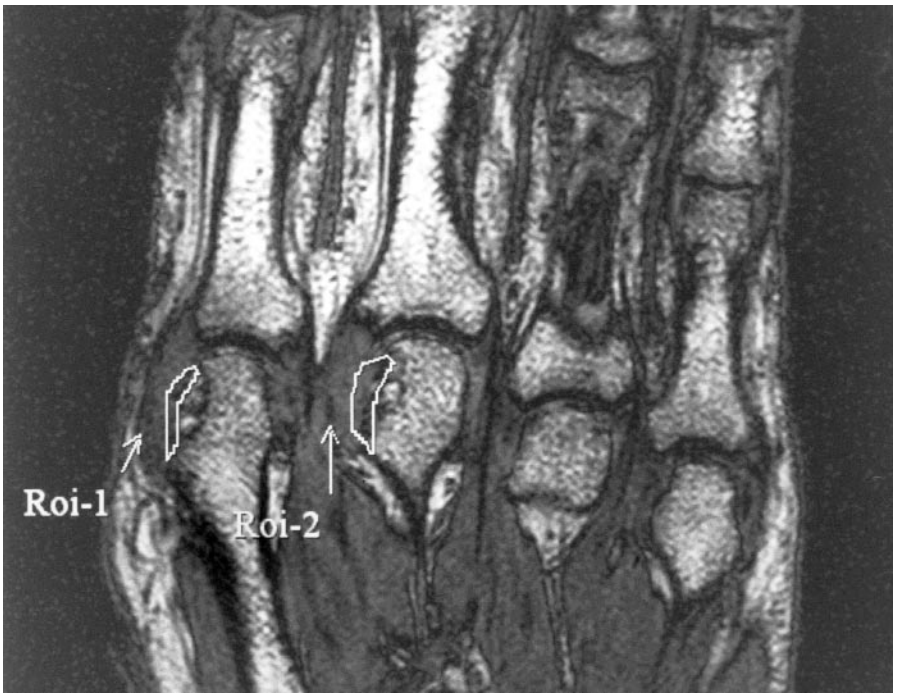

Figure 3. Erosions outlined in subsequent slices until the erosion is no longer visible. ROI: region of interest.

Table 1. Inter-reader reliability — total erosion volumes and total erosion score.

	ICC	95% CI
Erosion volume		
3 mm acquisition 1	0.73	0.16–0.96
3 mm acquisition 2	0.79	0.34–0.97
1 mm acquisition 1	0.73	0.26–0.96
1 mm acquisition 2	0.87	0.53–0.98
Erosion score		
3 mm acquisition 1	0.60	0.09–0.94

Table 2. Raw volume estimations comparing the 3 mm and the 1 mm acquisitions. Initial MRI series (volumes in mm^3).

	3 mm	1 mm	Difference
Sydney			
1	521	408	113
2	530	525	5
3	0	0	0
4	176	113	63
5	512	565	–53
Auckland			
1	462	573	–111
2	349	333	16
3	0	0	0
4	48	162	–114
5	409	284	125
Copenhagen			
1	207	320	–113
2	243	191	52
3	0	0	0
4	12	19	–7
5	303	474	–171
Mean differences:	Sydney 25	Auckland –16	Copenhagen –47

available free of charge, and was compatible with a standard personal computer. The method used is therefore generalizable and available to any group with a personal computer and access to the internet.

The time required to analyze the images is an important part of feasibility assessment. The computer volume method was more time consuming than the scoring method, and in its present form is difficult to recommend as a feasible measure in large clinical studies. However, as computer methods advance and automatic segmentation methods become available, the time taken for analysis will decrease. The manual segmentation process presented here attempts to set a foundation for future measurement and, as such, may provide a standard whereby future segmentation processes may be assessed.

Discrimination and reliability. Given that only one of the readers (PB) had experience in using the program and the other 2 readers (BE, FM) had no training with the software prior to undertaking the study, the inter-reader ICC results are encouraging. It should be noted, however, that there was

Table 3. Interoccasion reliability for the total erosion volumes.

	ICC	95% CI
3 mm acquisition 1 vs 3 mm acquisition 2		
Sydney	0.95	0.61–0.99
Auckland	0.91	0.51–0.99
Copenhagen	0.86	0.32–0.98
1 mm acquisition 1 vs 1 mm acquisition 2		
Sydney	0.94	0.56–0.99
Auckland	0.91	0.35–0.99
Copenhagen	0.99	0.95–0.98

a large systematic difference between volumes measured by different observers, probably mainly related to difficulties when determining the superficial border of each erosion. This difference was far less pronounced with the intra-reader volumes, suggesting that volumes obtained by the same observer based on the same or repeated MRI sessions have an acceptable reproducibility, but the inter-reader agreement is not sufficient at this time for a multi-reader setting. A larger study is required to address these issues, with reader training and calibration as essential prerequisites.

Inter-occasion variability is a composite measure representing 2 possible sources of variation in this study — variability in MRI acquisition or inter-reader variation. The erosion volume estimations were stable over a 24 hour period and, therefore, for the erosion volume measurements, occasion variability was not significant if patient positioning and acquisition specifications were defined.

Validity. Optimal slice thickness and correlation with scoring system. The positive correlation between the total scores and the erosion volumes is important because it supports construct validity of the semiautomated erosion volume method. With such small numbers it would be presumptuous to conclude that the 2 methods are equivalent, but the results suggest that a larger study would be worthwhile.

The question of the importance of slice thickness could not be definitely answered by this study. As expected, image analysis took longer when using the 1 mm acquisitions for both the volume method and the scoring method. Total erosion volume size was not consistently larger or smaller for either of the slice thickness acquisitions and from these results, we cannot conclude that 3 mm slices can replace high resolution 1 mm slices. The question remains unanswered and should be addressed in a larger study.

Conclusion

This study confirms that computerized erosion volume measurements are feasible, with high intra-reader and inter-occasion reliabilities. The inter-reader reliability in readers of varied experience was, in terms of ICC, comparable to the existing scoring system, but large systematic inter-observer variations were found, probably caused by problems when determining the superficial border of the erosions. Reader training in using the software program and calibration of readers will probably improve the inter-reader reliability and should be a prerequisite in planning further studies, which should also address the unresolved question regarding optimal slice thickness.

Whether the extra work involved in measuring rather than scoring erosions is currently justified is best addressed in a longitudinal intervention study, where its benefits as a continuous rather than a categorical measure may become more apparent, particularly when the degree of change is small.

The study supports the notion that direct measurement of erosions could function as an alternative to the scoring method, with the important caveat that neither approach represents the last word in the ways in which MRI can be used in rheumatoid arthritis assessment.

REFERENCES

1. McQueen FM, Stewart N, Crabbe J, et al. Magnetic resonance imaging of the wrist in early rheumatoid arthritis reveals a high prevalence of erosions at four months after symptom onset. Ann Rheum Dis 1998;57:350-6.
2. Ostergaard M, Hansen M, Stoltenberg M, et al. Magnetic resonance imaging as a predictor of longterm radiographic joint damage in rheumatoid arthritis [abstract]. Arthritis Rheumatism 2001;44 Suppl:S222.
3. McQueen FM, Benton N, Crabbe J, et al. What is the fate of erosions in early rheumatoid arthritis? Tracking individual lesions using x rays and magnetic resonance imaging over the first two years of disease. Ann Rheum Dis 2001;60:859-68.
4. Ostergaard 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.
5. McQueen F, Lassere M, Edmonds J, et al. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Summary of OMERACT 6 MR imaging module. J Rheumatol 2003;30:1387-92.
6. Lassere M. Pooled meta-analysis of radiographic progression: comparison of Sharp and Larsen methods. J Rheumatol 2000;27:269-75.
7. Edmonds JP, Lassere MN. Imaging damage: scoring versus measuring. J Rheumatol 2001;28:1749-51.
8. University Hospital of Geneva, Radiology Department. OSIRIS Imaging Software, 2000-2001. [cited January 14, 2003]. Available from: http://www.expasy.ch/www/UIN/html1/projects/osiris/osiris.html
9. Conaghan P, Edmonds J, Emery P, et al. Magnetic resonance imaging in rheumatoid arthritis: summary of OMERACT activities, current status, and plans. J Rheumatol 2001;28:1158-62.
10. Statistical Package for the Social Sciences (SPSS) for Windows version 10. Chicago, IL: SPSS Inc.; 1991-1999. INSO Corporation.