OMERACT 7 Special Interest Group

The Evidence for Magnetic Resonance Imaging as an Outcome Measure in Proof-of-Concept Rheumatoid Arthritis Studies

PHILIP G. CONAGHAN, FIONA M. MCQUEEN, CHARLES G. PETERFY, MARISSA N.D. LASSERE, BO EJBJERG, PAUL BIRD, PHILIP J. O’CONNOR, ESPEN HAAVARDSHOLM, JOHN P. EDMONDS, PAUL EMMERY, HARRY K. GENANT, and MIKKEL ØSTERGAARD

**ABSTRACT.** Magnetic resonance imaging (MRI) has now been used extensively in cross-sectional and observational studies as well as in controlled clinical trials to assess disease activity and joint damage in rheumatoid arthritis (RA). MRI measurements or scores for erosions, bone edema, and synovitis have been developed and validated by several groups. The OMERACT criteria require that outcome measures demonstrate adequate validity, discriminative power, and feasibility if they are to be useful in clinical trials. Specific performance targets for these criteria depend on the scientific, regulatory, logistical, and financial context of the study in question. We review the extent to which MRI assessments of joint erosion, bone edema, and synovitis fulfill these criteria, particularly as they relate to proof-of-concept RA clinical trials. (J Rheumatol 2005;32:2465–9)

**Key Indexing Terms:** MAGNETIC RESONANCE IMAGING, RHEUMATOID ARTHRITIS, VALIDITY

From the Academic Unit of Musculoskeletal Disease, University of Leeds, Leeds, UK; Department of Rheumatology, Auckland University, Auckland, New Zealand; Synarc Inc., San Francisco, CA, USA; Department of Rheumatology, St. George Hospital, University of NSW, Sydney, Australia; Copenhagen University Hospital at Hvidovre, Copenhagen, Denmark; Department of Rheumatology, University of NSW, Sydney, Australia; Department of Radiology, Leeds General Infirmary, Leeds, UK; Diakonhjemmet Hospital, University of Oslo, Oslo, Norway; Department of Rheumatology, University of NSW, Sydney, Australia; University of California, San Francisco, CA, USA; and Copenhagen University Hospitals at Herlev and Hvidovre, Copenhagen, Denmark.

P.G. Conaghan, MB, BS, PhD, FRACP, FRCP; Professor of Musculoskeletal Medicine, Academic Unit of Musculoskeletal Disease, University of Leeds; F.M. McQueen, MD, FRACP, Senior Lecturer in Rheumatology, Department of Rheumatology, Auckland University; C.G. Peterfy, MD, PhD, Chief Medical Officer, Synarc Inc.; M.N.D. Lassere, MB, BS, Grad Dip Epi, PhD, FRACP; FAFPHM, Associate Professor in Medicine, Department of Rheumatology, University of NSW; B. Ejbjerg, MD, PhD, Research Fellow, Copenhagen University Hospital at Hvidovre; P. Bird, BMed (Hons), FRACP, Consultant Rheumatologist, Department of Rheumatology, University of NSW; P.J. O’Connor, MB, BS, MRCP, FRCR, Consultant Skeletal Radiologist, Department of Radiology, Leeds General Infirmary; E.A. Haavardsholm, MD, Research Fellow, Diakonhjemmet Hospital, University of Oslo; J.P. Edmonds, MB, BS, MA, FRACP, Director and Professor of Rheumatology, Department of Rheumatology, University of NSW; P. Emery, MA, MD, FRCP, ARC Professor in Rheumatology, Academic Unit of Musculoskeletal Disease, University of Leeds; H.K. Genant, MD, FACR, FRCR, Professor of Radiology, University of California; M. Østergaard, MD, PhD, DMSc, Professor in Rheumatology/Arthritis, Copenhagen University Hospitals at Herlev and Hvidovre.

P.G. Conaghan and F.M. McQueen contributed equally to this work.

Address reprint requests to Dr. P.G. Conaghan, Academic Unit of Musculoskeletal Disease, Leeds General Infirmary, Great George Street, Leeds LS1 3EX, United Kingdom.
E-mail: philip.conaghan@leeds.nhs.uk

Magnetic resonance imaging (MRI) is now widely used in both research and clinical settings to image joints of patients with rheumatoid arthritis (RA). It is a multiplanar modality that offers excellent imaging of many tissues, including synovial membrane, tendons, and bone. This provides the clinician with information about joint inflammation and damage, even in very early disease when conventional radiography (CR) is typically uninformative. In the clinical trial setting, MRI potentially has many advantages over CR in measuring responses to therapeutic agents. It is more sensitive in detecting erosions and can document changes in bony damage over a shorter period of time. Thus trials using bone erosions as an endpoint could be run over 3 to 6 months rather than the typical 12–24 months seen currently. CR does not allow imaging of synovitis or tendinitis, both of which can be seen and measured using MR and are likely to be influenced by new therapies. Further, recent data suggest that both MR synovitis and bone marrow edema have prognostic implications in RA, and could be used in preselecting patients for trials of new and expensive agents that need to be targeted to those with the most aggressive disease.

In recent years, MRI measurements or scores for erosions, bone edema, and synovitis have been developed and validated by several groups. The aim of this review was to apply the OMERACT filter, which requires an outcome measure to exhibit “truth, discrimination, and feasibility,” to the current literature in order to determine whether MRI fulfills these criteria in the assessment of RA. The review
did not attempt to list all the publications reporting the measurement properties of MRI. It is important to note that many of the studies reviewed here focused on the small joints of the hand and wrist, and many have employed conventional high-field 1.5 T magnets. As well, the utility of MRI in clinical practice (where the performance metrics for MRI may be very different to that required for clinical trials) or in large, randomized trials has not been extensively reported, so this review of MRI usefulness should be interpreted in the context of proof-of-concept RA trials.

Erosions

Until recently, definitions of MRI bone damage were not identical from study to study. The OMERACT consensus group has addressed this issue and defined a MRI erosion as a “sharply marginated bone lesion with correct juxtaarticular localization and typical signal characteristics visible in 2 planes with a cortical break seen in at least one plane”7. To satisfy the “truth” or validity component of the OMERACT filter, MRI erosions need to reflect underlying bony pathology as closely as possible and should also be consistent with other imaging modalities. As yet, data are scarce regarding histopathological confirmation of MR erosions, apart from the work of Ostendorf, et al9, who found that surface bone defects seen arthroscopically at metacarpophalangeal (MCP) joints of RA patients coincided with erosions scored on matching MR scans. MR has been shown to be more sensitive than CR for erosion detection in many studies9,10, and the better comparator is computerized tomography (CT) that incorporates multiplanar imaging (equivalent to MR) with the clear depiction of cortical bone that is characteristic of plain radiography. Bedair, et al showed that 74% of wrist erosions identified by CT were detected on MRI11, while a recent study by Perry, et al found concordance between CT and MR erosions in 87% of cases12.

MR erosions have also been compared with their ultrasound (US) equivalents13-15. In the study reported by Wakefield, et al14, MRI, radiographic, and US erosions were compared at the radial aspect of the second MCP joint (where US has access equivalent to MRI). US revealed 13 lesions, of which 10 were seen at identical sites on MRI. One MRI erosion was seen as 2 separate lesions on US, and 2 more lesions were revealed on US alone. Radiography detected only one erosion that was also picked up on MR and US. In another study, biopsy of a small number of these US erosions demonstrated necrotic tissue consistent with erosion pathology16.

A critical question of predictive validity relates to whether MR erosions seen in early RA will develop into radiographic erosions over time. McQueen, et al showed that only 25% of lesions at the wrist were detectable radiographically 2 years later, but the authors noted that CR is a poor imaging modality for detecting individual erosions at this site, which has a complex 3-dimensional structure17. It should be noted that radiography is insensitive to bone erosions not only because of projectional superimposition, which obscures erosions that are en face to the x-ray beam, but also because almost all of the radiographic lucency of an erosion is attributable to cortical bone loss and not trabecular bone loss. Trabecular loss is virtually invisible on radiographs. This fundamental process makes the intramedullary component of bone erosions, which can be extensive, radiographically occult. In another study of wrist RA, Østergaard, et al showed that 78% of new radiographic erosions were detectable by MRI 1–5 years earlier18. This group estimated the median time interval between detection of an erosion using MR and its appearance on plain radiographs at 2 years18. An important comparison with healthy controls was performed by Ejbjerg, et al, who found MR “erosions” in only 0.2–0.4% of normal carpal and metacarpal bones19, indicating a low false-positive rate. It should be noted that in most of the intermodality comparison studies, both cross-sectional and longitudinal, exact geographic mapping of erosions at a particular intraarticular site has not been performed.

The second component of the OMERACT filter is discrimination. This incorporates qualities of reliability and sensitivity to change over time. Quantification of erosions may involve simple numeric counting, as used in many of the studies listed above, or some estimation of erosion volume. The OMERACT MRI-RA group developed a semiquantitative score (the RAMRIS system) and tested the reliability of this score for MR erosions. Intraclass correlation coefficients (ICC) of 0.5–0.8 for interobserver reliability were recorded for a group of 6 noncalibrated observers with different medical backgrounds, from multiple international centers, and without any substantial project-specific training20. However, studies using only 2 trained readers have demonstrated ICC of 0.75–0.9521,22. These levels of reliability are equivalent to those published for radiographic erosion scoring methods22. A EULAR-OMERACT Atlas has been developed to improve standardization of MR erosion scoring and this would be expected to improve reliability23.

New methods are being developed to quantify MR erosions using computerized techniques24-26. Bird, et al have used an outlining method to measure erosion volume at the wrist24 and MCP joints25. The former study revealed a strong correlation between erosion volume and the RAMRIS erosion score at the wrist. At MCP joints, interobserver reliability ICC of 0.73–0.87 were recorded, but there was significant systematic variation between readers relating to difficulty in estimating the erosion border, and further work is needed to determine reliability of this method25. Others have also quantified eroded bone volume using MRI, and found strong correlations with visual scoring26,27.

An outcome measure for scoring erosions must show sensitivity to change over time to be useful in clinical trials. Conaghan, et al used MRI in their study comparing
methotrexate with combination methotrexate/intraarticular steroid (IAST) therapy in patients with very early RA. They showed significant slowing of erosion progression in the combination group, who developed a mean of only one new erosion over 3 months, compared with 7 in the methotrexate-alone group. Interestingly, CR did not detect any of these erosions, indicating its relative insensitivity as an outcome measure in this context. Østergaard, et al compared MR with CR in RA patients treated with anakinra, and found that MR erosion progression over 3 months correlated with the change in the radiographic score over 9 months, providing further evidence for the usefulness of MRI in short-term studies. Moreover, MRI could detect erosion progression in a similar number of patients using a smaller field of view. However, in a recent longer-term (2 year followup) study, CR of both hands and wrists was more responsive to change in erosion score than MRI of the dominant MCP joints. However, unlike the anakinra study, this latter study did not use MRI to assess the wrist. When an observational study of 34 RA patients on maintenance therapy compared MRI of one hand and wrist and CR of both hands, wrists, and one foot, MRI showed statistically significant progression of erosion scores within 3 months, whereas CR showed no change over 6 months. Clearly, further studies are needed to determine which joint regions provide the optimum coverage.

**Bone Edema**

Bone edema has been shown by many groups to be a prominent feature of RA in very early as well as established disease. Bone edema is easily depicted with MRI, but cannot be visualized by radiography or US. Bone edema has been defined by the OMERACT group as “a lesion within the trabecular bone, with ill-defined margins and signal characteristics consistent with increased water content”. As yet no histopathological correlate has been defined, and whether it represents an inflammatory infiltrate within bone (osteitis) remains a matter of debate. Nevertheless, bone edema has been shown to correlate with measures of disease activity, including synovitis and clinical scores. Its validity and relevance to outcome relates to evidence that it is a pre-erosive lesion. Moreover, the bone edema score in early RA predicts both radiographic damage and functional outcome, underlining its importance as a prognostic indicator. The reliability of semiquantitative scores for bone edema has been poorer than for erosions, with reported ICC ranging from 0.5 to 0.86, again with small numbers of readers demonstrating better agreement. This may relate to difficulty in defining the borders of involved bone as well as technical challenges with failed fat-suppression. Another challenge is dealing with dependency of bone edema score on bone erosion score. As bone erodes, the volume of bone that could be involved with edema decreases proportionately. In the current draft of RAMRIS, edema score is based on the proportion of original articular bone involved rather than the proportion of residual bone involved. Therefore, the maximum possible score for edema decreases as erosion score increases. Basing edema score on the amount of residual bone involved would eliminate this effect, but at the expense of giving progressively greater weight to edema presenting in severely eroded bones, which the group felt to be an undesirable tradeoff. As bone edema seems to straddle roles between disease activity measure and forerunner of structural damage in RA, its most appropriate application in clinical trials has yet to be defined.

**Synovitis**

There is now an extensive literature describing the assessment of synovitis using MRI and its application as a disease activity measure in clinical trials. Various MR parameters have been used as markers of synovitis, including synovial membrane volume, measured by semiautomated outlining methods, synovial membrane thickness from axial scans, signal intensity after contrast enhancement, and a combination of these features in global scoring methods. Validity has been confirmed by comparison with histopathology and also direct visualization of synovial membrane at miniarthroscopy. Although low-grade post-contrast enhancement of synovium has been reported in healthy controls, the higher grades of synovitis are confined to inflammatory disease states. The OMERACT RAMRIS uses the following definition of synovitis: “an area in the synovial compartment that shows above normal post-gadolinium enhancement of a thickness greater than the width of the normal synovium”.

MRI is often put forward as the gold standard for assessment of synovitis against which other techniques must be measured. Comparison of MR synovitis with power Doppler ultrasonography showed very high levels of agreement (> 95%) in one study, while a strong correlation was demonstrated between the uptake of labelled tracers on positron emission tomographic scanning and MR synovial volume. Interestingly, while several groups have shown strong correlations with joint tenderness and swelling scores, there are also reports of MRI detecting subclinical synovitis, especially in early RA. Further evidence for the validity of MRI synovitis comes from longitudinal studies. It has been demonstrated that new erosions developed only at MCP joints where there was preceding MRI synovitis, that area-under-the-curve synovitis measurements predicted erosive progression, and that a high baseline score for wrist MRI erosions and synovitis best predicted longterm radiological erosive damage.

Reliability of a global MRI synovitis measurement has been demonstrated by the OMERACT group with ICC of 0.68–0.89. Two-reader studies have again shown higher ICC ranging from 0.74 to 0.90. Sensitivity to change in a clinical trial setting was indicated in the methotrexate-IAST...
study described above, in which synovial membrane thickness decreased by almost one-half over 3 months in joints injected with active drug compared with the placebo group, in which it remained unchanged. However, there are as yet scarce published data comparing changes in synovial volume and thickness with the smallest detectable difference for these measures. Caveats also apply to the interpretation of MRI synovitis data using post-contrast signal intensity, as this can vary significantly according to the region of synovium assessed.

Feasibility Issues

Feasibility is the third aspect that the OMERACT filter requires of an ideal outcome measure. It must be emphasized that feasibility is a particularly contextual parameter that depends on the scientific, logistical, and financial circumstances of the study at hand. While the amount of information available from MRI scans can seem daunting to the untrained observer, it is anticipated that application of the EULAR-OMERACT atlas will make the process of scoring RA bone erosion, bone edema, and synovitis feasible in a broader spectrum of research contexts. However, studies are needed to determine which regions should be scanned for optimal assessment of joint damage and disease activity. In some situations, the number of sequences may be reduced if only limited MR outcome data are required. Cost can be a barrier, but should be judged against the shorter time necessary to discern changes in disease status for trials using MRI outcome measures. The influence of smaller office-based units, developed at significantly lower cost and offering outcome measures, will make the process of scoring RA bone erosion, bone edema, and synovitis feasible in a broader spectrum of research contexts. However, studies are needed to determine which regions should be scanned for optimal assessment of joint damage and disease activity. In some situations, the number of sequences may be reduced if only limited MR outcome data are required.

In summary, there is strong evidence supporting the contention that MRI measurement of synovitis and bone erosions is a valid and reliable tool for assessing disease progression and treatment response in RA wrist and hand joints. The defined role and measurement properties of bone edema remain more problematic. The routine incorporation of MRI into clinical trials would allow further comparisons with CR is for the routine evaluation of rheumatoid disease status and as a means to measure the effectiveness of therapeutic interventions.

REFERENCES