Imaging in Rheumatoid Arthritis: Results of Group Discussions

ESMERALDA T.H. MOLENAAR, MAARTEN BOERS, DÉSIRÉE M.F.M. van der HEIJDE, GRACIELA ALARCÓN, BARRY BRESNIHAN, MARIO CARDIEL, JOHN EDMONDS, DAVID FELSON, DANIEL E. FURST, JOHN KIRWAN, MARISSA LASSERE, HAROLD PAULUS, ROLF RAU, PIET L.C.M. van RIEL, DAVID SCOTT, LEE SIMON, and VIBEKE STRAND

ABSTRACT. None of the current scoring methods for radiological damage in rheumatoid arthritis (RA) is ideal. The objective for RA imaging at OMERACT IV was to start discussion about the problems and applicability of the current scoring methods for radiological damage and to start discussion on the challenge of new imaging techniques. The RA imaging module comprised preconference reading material, plenary sessions, small group discussions, and a plenary report of the group sessions, combined with interactive voting. The OMERACT filter guided the discussions. Priorities for further research in imaging studies were: (1) pathologies versus features on radiographs; (2) relation with longterm outcome; and (3) definition of minimum clinically important difference. (J Rheumatol 1999;26:749–51)

Key Indexing Terms:
RHEUMATOID ARTHRITIS   RADIOGRAPHS   SCORING METHODS
LARSEN METHOD   SHARP METHOD   MAGNETIC RESONANCE IMAGING

In rheumatoid arthritis (RA), damage on radiographs is considered an important outcome measure. Radiological dam-

From the Department of Rheumatology, VU University Hospital, Amsterdam; Department of Rheumatology, University Hospital Maastricht, Maastricht, The Netherlands; Department of Medicine, University of Alabama at Birmingham, Birmingham, USA; Department of Rheumatology, St Vincents Hospital, Dublin, Ireland; Immunology and Rheumatology, Instituto Nacional de la Nutrición, Mexico City, Mexico; Department of Rheumatology, St. George Hospital, Sydney, Australia; Arthritis Center, Boston University, Boston; Virginia Mason Research Center, Seattle, USA; University of Bristol, Bristol, UK; University of California at Los Angeles, Los Angeles, USA; Evangelisches Fachkrankenhaus Ratingen, Ratingen, Germany; Department of Rheumatology, University Hospital Nijmegen, Nijmegen, The Netherlands; Division of Immunology, Stanford University, Stanford, USA.

E.T.H. Molenaar, MD, Department of Rheumatology; M. Boers, MSc, MD, PhD, Professor of Clinical Epidemiology, Department of Clinical Epidemiology, VU University Hospital, Amsterdam; D.M.F.M. van der Heijde, MD, PhD, Associate Professor of Rheumatology, Department of Rheumatology, University Hospital Maastricht; G.S. Alarcón, MD, MPH, Professor of Medicine, Department of Medicine, University of Alabama at Birmingham; B. Bresnihan, MD, FRCP, Professor of Rheumatology, Department of Rheumatology, St. Vincents Hospital; M. Cardiel, MD, Immunology and Rheumatology, Instituto Nacional de la Nutrición; J. Edmonds, MB, BS, FRACP, Professor of Rheumatology, Department of Rheumatology, St. George Hospital; D. Felson, MD, MPH, Arthritis Center; D.E. Furst, MD, Virginia Mason Research Center; J. Kirwan, MD, Head, Rheumatology Unit, University of Bristol; M. Lassere, MB BS, FRACP, Department of Rheumatology, St. George Hospital; H. Paulus, MD, Professor of Medicine, University of California at Los Angeles; R. Rau, MD, PhD, Professor of Rheumatology, Evangelisches Fachkrankenhaus Ratingen; P.L.C.M. van Riel, MD, PhD, Professor of Rheumatology, Department of Rheumatology, University Hospital Nijmegen; D. Scott, BSc, MD, FRACP, Reader in Rheumatology; L. Simon, MD, Deaconess Medicine, Boston, USA; V. Strand, MD, Division of Immunology, Stanford University.

Address reprint requests to Dr. E.T.H. Molenaar, Department of Rheumatology, VU University Hospital, PO Box 7057, 1007 MB Amsterdam, The Netherlands. E-mail: eth.molenaar@azvu.nl

age can be quantified by several scoring methods. None of these methods is ideal; each has its strengths and limitations. The RA imaging module at the OMERACT IV conference considered several aspects of imaging. The objectives for the RA imaging module were to discuss the problems and applicability of the current scoring methods for radiological damage and to consider what might be proposed as a smallest detectable difference in progression in damage, based in the first instance on an estimation of the measurement error. In addition this module also started discussion of the challenge posed by new imaging techniques. Such techniques [e.g., magnetic resonance imaging (MRI), ultrasonography, radionuclide imaging, dual energy x-ray absorption (DEXA)] are currently being applied to assess RA. These techniques give new and sometimes spectacular images of joints. The RA imaging module discussed MRI as a primary example and ultrasound as a secondary example of new imaging methods.

In general, measurement is performed for one of 3 purposes: classification, prognostication, and measurement of change over time. In recent years several studies have presented MRI findings in patients with RA, as illustrated by an informal survey based on a literature search. Most of the 10 retrieved studies were cross sectional: they compared MRI findings of patients with early RA with those of osteoarthritis patients, patients with other inflammatory joint diseases, and healthy controls. MRI findings were also related to tissue findings from synovial biopsies in the same patients. The abnormalities include findings related to synovitis, damage, or changes that may be related to both. These studies are targeted toward classification, possibly with a view to prognosis. Only one study was targeted toward measurement of change over time. Over all, this survey suggested

that formal research into the applicability of MRI in RA has just started. The OMERACT IV RA imaging module therefore also discussed ways to improve the quality of such studies.

METHODS

The imaging module consisted of preconference reading material, plenary sessions, small group sessions, and a plenary report of the group sessions combined with interactive voting. The preconference reading material comprised a review of the OMERACT process[1], a paper on statistical methods for assessing agreement between 2 methods of clinical measurement[2], and abstracts of the introductory plenary lectures.

These lectures introduced the issues to be discussed in the group sessions and are summarized elsewhere in these proceedings[3-5]. Participants then formed 6 groups of about 16 persons. Each group discussed the proposed smallest detectable difference (SDD) in progression of damage presented in the plenary session[6]. Each group then discussed a specific subject: Groups 1 and 2 examined the problems and applicability of the current scoring methods for radiological damage on plain radiographs of hands and feet; Groups 3 and 4 considered whether there is a need for a new scoring method for plain radiography, and if so, what aspects should be incorporated. The discussions also elicited ideas for research on the development and validation of a new scoring method. The specific task for Groups 5 and 6 was to devise recommendations on how to validate new imaging methods.

The OMERACT method to reach consensus has been described[7]. A group leader who had been briefed by the module committee led the group discussions. The discussions were based on the nominal group technique[8]. This is a technique to generate ideas, freely discuss each item, and make decisions where appropriate.

The OMERACT filter[1] guided discussion of the items. This filter was framed after considering purposes and applicability of measures and measurement in medicine. The OMERACT filter for assessing measures is composed of 3 concepts: truth, discrimination, and feasibility. Each represents a question to be answered while considering the measure in its intended setting.

Truth. Is the measure truthful, unbiased, and relevant?

Discrimination. Does the measure discriminate between situations of interest (classification, prognosis, and measurement of change over time)?

Feasibility. Can the measure be applied easily, given constraints of time, costs, and interpretability?

RESULTS

Four of the groups were charged with discussing plain radiography either by formulating an extensive list of problems with current methods of scoring, or by envisaging new scoring methods, their requirements, and the way to validate such systems.

The results of these discussions were similar despite the different approaches taken by the groups. In relation to the truth aspect of the OMERACT filter, it was felt important to report joint space narrowing (JSN) and erosions separately in any scoring method for radiological damage. These abnormalities should be scored in both hands and feet. The usefulness of other abnormalities such as osteopenia, cysts, subluxation, and fusion, as well as healing phenomena, should be explored. However, it was also recognized that the relevance of the abnormalities we score depends on our understanding of the underlying processes. Much more study is needed to clarify the relationship between damage

and outcomes important to the patient, suggesting that radiography should be a routine part of most longitudinal studies. The issue of patient preference in damage scoring was briefly considered, but not resolved (i.e., whether damage in joints that are important to the patient should be weighted more heavily than damage in other joints).

Considering the discrimination aspect of the filter, most of the issues brought up in the plenary session were placed into the research agenda. These included the order of reading, scaling problems (e.g., ordinal scales for continuous damage, ceiling effects), and weighting problems (e.g., the weight of damage in the feet). It was felt necessary to continue the comparison of the discriminatory capacity of the Sharp/van der Heijde and Larsen/Scott as well as other methods. Considering the smallest detectable difference[6], it was felt that this was not the same as the minimum clinically important difference, and that radiological measurement error (RME) was a more appropriate term. It was thought that RME could be influenced by disease related factors (e.g., disease duration) and technical factors (e.g., radiographic technique). It was recommended to perform this RME analysis in each clinical trial with radiographs. Considering RME in relation to clinically important difference, it was felt that more research was needed into the usual progression rates and arrest in progression (these data could come from longitudinal studies).

Considering the feasibility aspect, most participants experienced first-hand that the Larsen method was easier to master and apply than the Sharp method[9]. Although little time was spent on the discussion of this very important aspect, it was recognized that we must also consider parsimony when looking at new and better ways to study joint damage. Parsimony can be explained as measuring what needs to be measured efficiently and without redundancy.

Groups 5 and 6 discussed validation of new imaging methods. As an example of a new imaging method (and due to time constraints), the groups discussed only the applicability of MRI. This technique might be useful in all 3 areas: classification, prognosis, and evaluation of change over time. Especially with regard to classification, MRI might be useful in detecting early damage (erosions), distinguishing between inflammatory and noninflammatory disease, and distinguishing between RA and other arthritides. MRI might be a tool to anatomically classify RA, and might even be able to distinguish different types of cell groups involved in inflammation. MRI might also prove useful in prognosticating and for followup over time. It was recommended that studies of MRI should be explicit in what purpose(s) are being addressed in view of the multiplicity of objectives and their differing study design and analysis requirements. Considering the OMERACT filter, participants recognized problems with MRI in all aspects. Regarding *truth*, it is unknown what the lesions seen on MRI signify. With respect to *discrimination*, little has been done to study reli-

ability of the scoring of abnormalities. MRI might make actual quantification possible instead of scoring, but this needs to be developed further. Currently, MRI is certainly not a feasible technique in most settings, although technology is rapidly evolving. The cost aspect of this technique might also limit its widespread use.

The groups identified as high priority for research: to confirm differences between RA and other diseases; to identify abnormalities with prognostic significance in early RA; and to identify abnormalities that are sensitive to change. Finally, standardization of technique and scoring and reliability studies need to be done in conjunction with exploratory studies. Although there was no time to discuss other imaging techniques, the groups recognized that many of the issues raised in the discussions were generic and thus applicable to any new technique.

CONCLUSION

We suggest some key recommendations and priorities for further research in imaging studies.

Truth. There is a need to relate pathology to features on radiographs. The relationship with longterm functional outcome requires more clarification and is likely to be elucidated by longterm cohort studies. In addition, studies should be explicit about the purpose of measurement: classification, evaluation over time, prognostication; it may be that different techniques are required for each of these purposes.

Discrimination. Greater standardization of imaging techniques and measurement methods between centers and across studies is desirable and the definition of clinically important differences needs more work through further research on radiological measurement error. Other suggested approaches and expected progression rates need to be defined and validated.

Feasibility. Future studies on new imaging techniques and scoring methods should be developed with attention to ease of use, cost, and interpretability. A key word is parsimony.

As a result of these discussions and the recognition of the potential for further developments in this area, participants chose to establish an OMERACT Imaging Task Force. The task force's mission is to facilitate research by: exploring existing radiograph databases (contact person: John Kirwan), and contacting MRI (contact person: John Edmonds) and ultrasound professionals (contact person: Barry Bresnihan). The task force met for the first time at the American College of Rheumatology 1998 annual scientific meeting in San Diego (contact: Désirée van der Heijde, dhe@ms-azm-3.azm.nl).

REFERENCES

1. Boers M, Brooks P, Strand V, Tugwell P. The OMERACT filter for outcome measures in rheumatology [editorial]. J Rheumatol 1998;25:198-9.
2. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1:307-10.
3. Bresnihan B. Pathogenesis of joint damage in rheumatoid arthritis. J Rheumatol 1999;26:717-9.
4. Kirwan JR. Conceptual issues in scoring radiographic progression in rheumatoid arthritis. J Rheumatol 1999;26:720-5.
5. Van der Heijde DMFM, Boers M, Lassere M. Methodological issues in radiographic scoring methods in rheumatoid arthritis. J Rheumatol 1999;26:726-30.
6. Lassere M, Boers M, van der Heijde DMFM, et al. Smallest detectable difference in radiological progression. J Rheumatol 1999;26:731-9.
7. Fried BJ, Boers M, Baker PRA for the OMERACT Committee. A method for achieving consensus on rheumatoid arthritis outcome measures: the OMERACT Conference process. J Rheumatol 1993;20:548-61.
8. Delbecq AL, Van de Ven AH, Gustafson DH. Group techniques for program planning. A Guide to nominal group and delphi processes. Glenview: Scott, Foresman; 1975.
9. Molenaar ETH, Edmonds J, Boers M, van der Heijde DMFM, Lassere M. A practical exercise in reading RA radiographs by the Larsen and Sharp methods. J Rheumatol 1999;26:746-8.