Is There a Preferred Method for Scoring Activity of the Spine by Magnetic Resonance Imaging in Ankylosing Spondylitis?

DÉSIRÉE van der HEIJDE, ROBERT LANDEWÉ, KAY-GEERT HERMANN, MARTIN RUDWALEIT, MIKKEL ØSTERGAARD, ANS OOSTVEEN, PHIL O'CONNOR, WALTER P. MAKSYMOWYCH, ROBERT G. LAMBERT, CÉDRIC LUKAS, ANNE GRETHE JURIK, MAARTEN BOERS, XENOFON BARALIAKOS, JÜRGEN BRAUN, for the ASAS/OMERACT MRI in AS Working Group

ABSTRACT. This report summarizes the discussion during a module update at OMERACT 8 on scoring methods for activity in the spine on magnetic resonance imaging. The conclusion was that the 3 available scoring methods are all very good with respect to discrimination and feasibility: the Ankylosing Spondylitis spine MRI score for activity (ASspiMRI-a), the Berlin method (a modification of the ASspiMRI-a), and the Spondyloarthritis Research Consortium of Canada Magnetic Resonance Imaging Index for Assessment of Spinal Inflammation in AS (SPARCC). All 3 methods were judged to be similar with respect to responsiveness and discrimination, although the differences in between-reader intraclass correlation coefficients (ICC) were judged to be relevant (the SPARCC method provided consistently higher ICC). The Berlin and SPARCC methods were preferred most frequently. The development of a new method combining the best elements of all methods is an additional possibility. (J Rheumatol 2007;34:871–3)

Key Indexing Terms:
MAGNETIC RESONANCE IMAGING ANKYLOSING SPONDYLITIS
SPINE IMAGING METHODOLOGY

From the Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht and CAPHRI Research Institute, University Maastricht, Maastricht, The Netherlands; Department of Radiology, Charité Medical School, Berlin, Germany; Department of Rheumatology, Charité – Campus Benjamin Franklin, Berlin, Germany; University of Copenhagen, Hvidovre and Herlev Hospitals, Copenhagen, Denmark; Department of Rheumatology, Twenteborg Hospital, Almelo, The Netherlands; Department of Radiology, Leeds General Infirmary, Leeds, UK; Department of Medicine and Department of Radiology, University of Alberta, Edmonton, Canada; Department of Radiology, Aarhus University Hospital, Aarhus, Denmark; Department of Clinical Epidemiology and Biostatistics, Free University Medical Center, Amsterdam, The Netherlands; and Rheumazentrum Ruhrgebiet, Herne, Germany.

D.M.F.M. van der Heijde, MD, PhD, Professor of Rheumatology; R.B. Landewé, MD, PhD, Associate Professor of Rheumatology; C. Lukas, MD, Rheumatologist, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht and CAPHRI Research Institute, University Maastricht; K.G.A. Hermann, MD, PhD, Radiologist, Department of Radiology, Charité Medical School; M. Rudwaleit, MD, PhD, Rheumatologist, Department of Rheumatology, Charité – Campus Benjamin Franklin; M. Østergaard, MD, PhD, DMSc, Professor of Rheumatology, University of Copenhagen; A. Oostveen, MD, PhD, Rheumatologist, Department of Rheumatology, Twenteborg Hospital; P. O'Connor, MD, Radiologist, Department of Radiology, Leeds General Infirmary; W.P. Maksymowych, FRCPC, Professor of Medicine, Department of Medicine, University of Alberta; R.G. Lambert, MD, Professor of Radiology, Department of Radiology, University of Alberta; A.G. Jurik, MD, DMSc, Radiologist, Department of Radiology, Aarhus University Hospital; M. Boers, MD, PhD, Professor of Clinical Epidemiology, Department of Clinical Epidemiology and Biostatistics, Free University Medical Center; X. Baraliakos, MD; J. Braun, MD, PhD, Professor of Rheumatology, Rheumazentrum Ruhrgebiet.

Address reprint requests to Dr. D. van der Heijde, Department of Rheumatology, University Hospital Maastricht, PO Box 5800, 6202 AZ Maastricht, The Netherlands. E-mail: d.vanderheijde.kpnplanet.nl

Assessment of disease activity in patients with ankylosing spondylitis (AS) is mainly based on patient reported outcomes. The assessment of damage in AS is based on radiographs. Magnetic resonance imaging (MRI) provides the advantage of direct visualization of activity and damage. Quantification of activity on the MRI may under certain conditions serve as a useful, more objective assessment of disease activity. To be able to validate whether the activity seen on MRI is indeed a reflection of disease activity a scoring system is needed. The most important sites in AS are the sacroiliac (SI) joints and the spine. During OMERACT 7, scoring methods for defining activity in SI joints were compared. It was decided that most emphasis should be placed now on the validation of scoring methods of MRI activity in the spine. The ASsessment in Ankylosing Spondylitis(ASAS)/OMERACT MRI in AS Working Group presented a module update at OMERACT 8. In preparation, a large multireader study was performed, comparing 3 available methods to assess activity: the Ankylosing Spondylitis spine MRI score for activity (ASspiMRI-a), the Berlin method (a modification of the ASspiMRI-a), and the Spondyloarthritis Research Consortium of Canada Magnetic Resonance Imaging Index for Assessment of Spinal Inflammation in AS (SPARCC)[1-3]. The results of this multireader study are published in this issue of *The Journal*, and were presented to the participants at OMERACT 8 to serve as a scientific basis for comparisons of performance of the methods according to the OMERACT filter[4].

In summary, feasibility of all methods was good and large-

ly comparable. Interreader reliability was more difficult to judge. While the intraclass correlation coefficients (ICC) were consistently better for the SPARCC method as compared to the ASspiMRI-a and Berlin methods, using the smallest detectable difference as the measure for agreement yielded the opposite result. Sensitivity to change over a 24-week period and discrimination between a tumor necrosis factor-blocker and placebo were comparable, and excellent for all 3 methods.

A discussion followed the presentation of the data, and concluded with formal voting. During the closing session a summary of the data was presented with a second voting in a larger audience. The answer options for the voting questions always included all possible combinations of scoring methods (Table 1). For example, the 10 answer categories for the question "Which method(s) do you consider feasible for use in RCTs?" were: none, ASspiMRI-a only, Berlin only, SPARCC only, ASspiMRI-a and Berlin, ASspiMRI-a and SPARCC, Berlin and SPARCC, all, insufficient data, don't know. However, in the presentation of the results a vote for the category "Berlin and SPARCC" would imply a vote for both Berlin and SPARCC, resulting in a higher number of votes than voters and a different total number per question. In the module update, each question was answered on average by 56 persons (range 42–65). For the various aspects of the OMERACT filter considered separately, the participants felt that all 3 methods fulfilled the requirements sufficiently, with a tendency to more votes for Berlin and SPARCC as compared to ASspiMRI-a, and a majority indicated that the differences in

Table 1. Questions and responses from the first vote during the module update and from the second vote during the plenary wrap-up session.

Voting Question	Response Range	First Vote	Second Vote
Which method(s) do you consider feasible for use in RCTs?	None	1	—
	ASspiMRI-a	25	
	Berlin	33	
	SPARCC	30	
	Insuff. data	2	
	Don't know	2	
Do you consider the observed differences in time to score between the methods relevant?	Yes	25	45
	No	28	51
	Don't know	2	4
Which method(s) do you consider sufficiently reliable for use in RCTs?	None	4	—
	ASspiMRI-a	29	
	Berlin	37	
	SPARCC	34	
	Insuff. data	4	
	Don't know	5	
Do you consider the observed differences in ICCs between the methods relevant?	Yes	30	55
	No	16	49
	Don't know	9	16
Which method(s) do you consider sufficiently discriminatory for use in RCTs?	None	2	—
	ASspiMRI-a	35	
	Berlin	49	
	SPARCC	45	
	Insuff. data	4	
	Don't know	7	
Which method(s) do you consider sufficiently sensitive to change for use in RCTs?	None	2	—
	ASspiMRI-a	35	
	Berlin	35	
	SPARCC	41	
	Insuff. data	6	
	Don't know	8	
Taking all aspects of the OMERACT filter into account, which method(s) are suitable for use in RCTs?	None	2	3
	ASspiMRI-a	40	62
	Berlin	50	69
	SPARCC	49	76
	Don't know	5	17
Taking all aspects of the OMERACT filter into account, which method(s) do you prioritize for use in RCTs?	ASspiMRI-a	16	37
	Berlin	35	56
	SPARCC	35	72
	Don't know	7	17

ASspiMRI-a: Ankylosing Spondylitis spine MRI score for activity; Berlin: a modification of the ASspiMRI-a; SPARCC: Spondyloarthritis Research Consortium of Canada Magnetic Resonance Imaging Index for Assessment of Spinal Inflammation in AS.

ICC were relevant. This was also expressed in the overall application of the OMERACT filter. With regard to prioritization, there was an equal preference for the Berlin and the SPARCC methods. The second voting in the general assembly confirmed the impression of the first voting, with the highest preference figures for SPARCC, followed by Berlin and ASspiMRI-a.

The positive conclusions from the voting results were that there are 3 methods available on MRI for assessing activity in the spine that have passed important requirements of the OMERACT filter. However, the voting results did not justify the selection of only one preferred method, which led to major discussions. Apart from the validity aspects tested here, there are other major issues on which we have sparse data so far: How truthful are the methods? Truth takes into account questions such as, "Is the method credible?," "Are all important features part of the method?," "Is it a good representation of disease activity?," "Can the method predict structural damage?" etc. Now that the methods have proven to pass the OMERACT filter, data on these aspects can be collected. Credibility aspects can be judged now to some extent. There are some major differences in how the scores are pulled together and what aspects are included in the scores. The only difference between ASspiMRI-a and Berlin, for instance, is the inclusion of erosions in the scores. In the ASspiMRI-a score, active lesions that include an erosion get a higher score (4–6) compared to active lesions without an erosion (0–3), whereas the Berlin method does not take erosions into account (0–3). A number of people consider erosions a sign of damage and are therefore in favor of not including erosions in the activity score. In addition, there is a concern regarding assigning higher scores to erosions than inflammation in an instrument that is designed to measure inflammation. For all other aspects the ASspiMRI-a and Berlin methods can be discussed together.

Although all 3 approaches require assessment of the entire spine, ASspiMRI-a/Berlin methods score all 23 levels, whereas the SPARCC method scores only the 6 most severely affected vertebral unit (VU). Another distinction is the scoring per VU. The ASspiMRI-a/Berlin methods score the entire VU with a range from 0 to 6 or 0 to 3, and the SPARCC method divides the VU in quadrants and scores lesions in all 3 dimensions by scoring consecutive sagittal images, but does not acknowledge a grading in activity within the individual quadrants.

Moreover, the SPARCC method does give additional points to "depth of inflammation" and "intensity of the lesion." The potential advantage of scoring all VU is that it is a better reflection of the truth (the entire spine), and that there may be less risk of a ceiling effect, but it might also introduce extra noise because of inclusion of doubtful lesions, and it mandates the scoring of regions of artefact that occur commonly in MRI examinations. The SPARCC score may or may not suffer from a ceiling effect due to the limited number of levels scored, but this was not evident in the OMERACT scoring exercise, which included MRI scans from patients with very active disease recruited to a trial of infliximab therapy. The Berlin method may or may not suffer from a ceiling effect due to the limited scoring permitted at each level, and again there are no data to indicate either way. At this stage, we have no way of knowing if any of the systems are at serious risk of this effect, or whether the consequences would be more detrimental to one system than another. An argument against using the 6 VU with the highest activity, and adding points for extension (quadrants and depth) and intensity, is that this score may overestimate the true activity in a given patient. Scoring the entire spine might especially be important in longterm followup, as the lesions with the most severe activity could differ over time, causing difficulties in use of the SPARCC score.

The major conclusion from the module update at OMERACT was that there are 3 scoring methods available for scoring activity in the spine on MRI that pass the OMERACT filter. The Berlin and SPARCC methods were preferred most frequently. All 3 methods were judged to be similar with respect to responsiveness and discrimination, although the higher ICC for the SPARCC method were judged to be relevant. More data must be collected on the truth aspects. An important issue for the research agenda could be to challenge all 3 methods in a setting with less difference between treatment and no-treatment groups. Another option is to develop a new scoring method that combines the best elements of the currently available methods.

REFERENCES

1. Braun J, Baraliakos X, Golder W, et al. Magnetic resonance imaging examinations of the spine in patients with ankylosing spondylitis, before and after successful therapy with infliximab: evaluation of a new scoring system. Arthritis Rheum 2003;48:1126-36.
2. Rudwaleit M, Schwarzlose S, Listing J, et al. Is there a place for magnetic resonance imaging (MRI) in predicting a major clinical response (BASDAI 50%) to TNF alpha blockers in ankylosing spondylitis? [abstract]. Arthritis Rheum 2004;50 Suppl:S211.
3. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. Arthritis Rheum 2005;53:502-9.
4. Lukas C, Braun J, van der Heijde D, et al, for the ASAS/OMERACT MRI in AS working group. Scoring inflammatory activity of the spine by magnetic resonance imaging in ankylosing spondylitis: A multi-reader experiment. J Rheumatol 2007;34:862-70.