

## EXTENDED REPORT

# Assessment of acute spinal inflammation in patients with ankylosing spondylitis by magnetic resonance imaging: a comparison between contrast enhanced T<sub>1</sub> and short tau inversion recovery (STIR) sequences

X Baraliakos, K-G A Hermann, R Landewé, J Listing, W Golder, J Brandt, M Rudwaleit, M Bollow, J Sieper, D van der Heijde, J Braun



*Ann Rheum Dis* 2005;64:1141–1144. doi: 10.1136/ard.2004.031609

See end of article for authors' affiliations

Correspondence to: Professor J Braun, Rheumazentrum Ruhrgebiet, Landgrafenstr 15, 44652 Herne, Germany; j.braun@rheumazentrum-ruhrgebiet.de

Accepted 2 January 2005  
Published Online First  
13 January 2005

**Objectives:** To compare the performance of two different MRI sequences—T<sub>1</sub> weighted, fat saturated, spin echo after application of contrast medium, and short τ inversion recovery (STIR) sequences—to detect spinal inflammation in patients with ankylosing spondylitis (AS).

**Methods:** Both MRI sequences were performed in 38 patients with active AS and compared using the MRI activity scoring system, ASspiMRI-a. One vertebral unit (VU) was defined as the region between two virtual lines drawn through the middle of each vertebral body.

**Results:** Intraclass correlation coefficients were excellent—0.91 and 0.86 for the Gd-DTPA and STIR sequences, respectively. The overall correlation of the single MRI scores for both sequences was also good ( $r=0.84$ ,  $p=0.01$ ). The intrarater variance was 6.71 and 9.41 and the interrater variance was 13.16 and 19.04 for the Gd-DTPA and STIR sequences, respectively. The smallest detectable distance was 4.7 and 5.6 for the Gd-DTPA and STIR sequences, respectively. The concordance rate for both sequences was 83.5% (range 80.5–87.7% in the three spinal segments). Inflammatory spinal lesions were found in 10.1% of the VUs in the STIR sequence but not in the T<sub>1</sub>/Gd-DTPA sequence, while the T<sub>1</sub>/Gd-DTPA sequence showed inflammatory lesions in 6.4% of the VUs that were found normal by STIR.

**Conclusions:** Both MRI techniques can evaluate active spinal lesions in patients with AS. More spinal lesions are detected by the STIR sequence, but the reliability between readings and readers is better for the Gd-DTPA sequence. The ASspiMRI-a is a reliable instrument for evaluating acute spinal changes in AS.

Ankylosing spondylitis (AS) is a chronic inflammatory rheumatic disease that mainly affects the spine. Conventional plain radiography of the spine and pelvis is the current standard for imaging in AS, as it can visualise chronic changes such as syndesmophytes.<sup>1</sup> Magnetic resonance imaging (MRI) has been shown to detect acute spinal lesions and to assess the change of such lesions over time in patients treated with the anti-tumour necrosis factor antibody, infliximab.<sup>2</sup>

Active spinal lesions can be assessed by MRI by using T<sub>1</sub> weighted, fat saturated sequences after application of contrast agents such as gadolinium-diethylenetriamine-pentaacetic acid (Gd-DTPA). Enhancement of the contrast agent is believed to indicate continuing inflammation. It is unclear whether another MRI technique, the short τ inversion recovery (STIR) sequence, which it is known can visualise normal bone marrow and bone marrow oedema,<sup>3</sup> performs similarly well in this regard. The first scoring system for evaluation of MRI sequences in AS, the ASspiMRI-a, which has recently been proposed and evaluated by our group,<sup>2</sup> includes both techniques. STIR is easier and faster to perform, and less costly than techniques depending on the use of contrast agents, but the Gd-DTPA technique is believed to be more specific in depiction of inflammatory spinal lesions. So, the question of relative performance of both techniques is clinically relevant.

The primary aim of this study was to compare the performances of T<sub>1</sub> weighted, fat saturated post-Gd-DTPA and STIR MRI sequences by using the recently proposed scoring method to assess spinal inflammation in patients with AS.

## PATIENTS AND METHODS

### Patients' characteristics

Thirty eight patients with AS, who had to fulfil the modified NY classification criteria for AS,<sup>4</sup> were randomly selected. Twenty five (66%) of the 38 patients with AS were male, with a mean age of 40.9 years (range 32–54), and 35 (92%) of the patients were HLA-B27 positive. The mean (SD) C reactive protein was 222 (219) mg/l and the mean (SD) erythrocyte sedimentation rate 31.2 (23.0)/1st h. The patients had active disease with a mean (SD) Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) of 6.4 (1.4) and a mean (SD) Bath Ankylosing Spondylitis Functional Index (BASFI) of 5.5 (2.1). Conventional radiographs of the pelvis were available in all patients and, in some, also of the spine.

### Magnetic resonance imaging

MRI investigations were executed with a 1.5 T unit (Magnetom vision, Siemens, Erlangen, Germany), using a spine array coil or a body array coil, or both. The MRI techniques applied to assess spinal inflammation in patients with AS were performed as described previously.<sup>5</sup> The sagittal section orientation was chosen and the following sequences were used:

- T<sub>1</sub> weighted spin echo sequences (repetition time/echo time 500/12 ms, slice thickness 3 mm, four acquisitions, .....

**Abbreviations:** AS, ankylosing spondylitis; Gd-DTPA, gadolinium-diethylenetriamine-pentaacetic acid; ICC, intraclass correlation coefficient; MRI, magnetic resonance imaging; SDD, smallest detectable difference; STIR, short τ inversion recovery; VU, vertebral unit

**Table 1** Inter- and intrarater variance of the scores with 95% confidence intervals (CI), intraclass correlation coefficients (ICC), and smallest detectable differences (SDDs)

	Interrater variance	95% CI	Intrarater variance	95% CI	Variance between patients	ICC	SDD
<i>ASspiMRI-a (Gd-DTPA)</i>							
CS	1.47	1.09 to 2.11	1.07	0.79 to 1.53	4.85	0.776	1.9
TS	8.96	6.61 to 12.83	4.20	3.10 to 6.01	58.67	0.882	3.7
LS	1.84	1.35 to 2.65	0.83	0.61 to 1.20	15.16	0.905	1.7
Spine (all 3 Segments)	<b>13.16</b>	<b>9.63 to 19.05</b>	<b>6.71</b>	<b>4.91 to 9.71</b>	<b>120.68</b>	<b>0.914</b>	<b>4.7</b>
<i>ASspiMRI-a (STIR)</i>							
CS	1.62	1.19 to 2.31	1.08	0.80 to 1.54	5.88	0.795	1.9
TS	8.56	6.31 to 12.30	5.14	3.78 to 7.38	62.34	0.875	4.1
LS	2.62	1.93 to 3.78	1.32	0.97 to 1.89	4.93	0.651	2.1
Spine (all 3 Segments)	<b>19.04</b>	<b>13.97 to 27.49</b>	<b>9.41</b>	<b>6.90 to 13.58</b>	<b>109.29</b>	<b>0.858</b>	<b>5.6</b>

CS, cervical spine; TS, thoracic spine; LS, lumbar spine.

field of view 20 cm×40 cm, matrix 128×512 pixels) before, and

- The same sequence with fat saturation after application of Gd-DTPA (Schering AG, Berlin, Germany) at 0.1 mmol/kg body weight.

No dynamic imaging was performed. C2 and L5 were taken as orientation points and the spine was examined in two parts, always starting with the upper part. After rapid adjustment of the table into the appropriate position the lower part of the spine was examined.

- Similarly, fat saturated STIR sequences (repetition time/inversion time/echo time 4000/150/60 ms, slice thickness 3 mm, five acquisitions, field of view 25 cm×40 cm, matrix 121×256 pixels) were performed.
- T<sub>2</sub> weighted images were also available and were taken into account in doubtful cases of differentiation between chronic and acute lesions.

**Scoring of the MRI sequences**

After all MR images had been blinded for patient identity, an independent person randomly selected the order of the films, which were then evaluated twice by two readers (JB, WG). Each evaluation included first the STIR and, secondly, the Gd-DTPA MR images of each patient. Thus, each image was

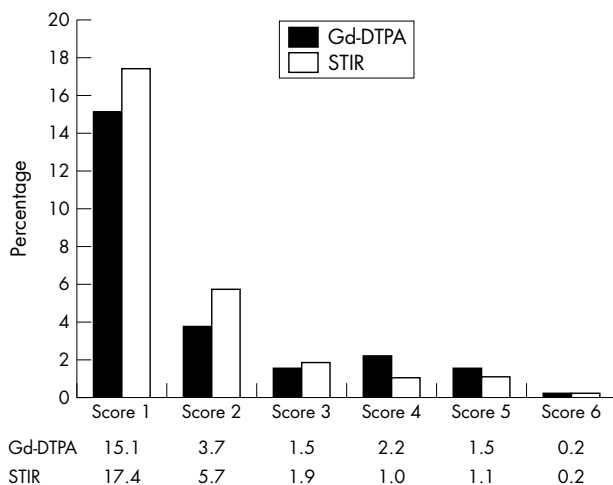
evaluated four times. The recently proposed MRI scoring system, ASspiMRI-a, which has been evaluated for assessment of acute inflammatory and possibly simultaneous erosive spinal lesions<sup>2-6</sup> was used to analyse the MR images of both sequences on the basis of a vertebral unit (VU), which was defined as the region between two virtual lines drawn through the middle of each vertebral body.<sup>2</sup>

Separate scores were used to test the intra- and interrater variability. Definite involvement of a VU by inflammation was defined as a score >1 as proposed elsewhere.<sup>7</sup> All scorings were done on the basis of single VUs. Concordance and discordance rates were calculated by identifying positive definite inflammatory involvement in both MRI sequences and by always comparing the same VUs in each MRI sequence.

**Statistical analysis**

The reliability of the entire score was evaluated by estimating the variability between the two readers, as well as the variability within the readers. A nested variance analysis approach was used for the calculation of both types of variance—the interrater variance and the intrarater variance. The intrarater variance was estimated in an analysis of a variance type I model with the patients as the first factor and the reader as the second random factor. Similarly, the interrater variance was estimated in a nested model with patients as first factor and readings as second factor.

The intrarater variance was used to calculate the smallest detectable difference (SDD) between two readings of one



**Figure 1** Distribution of the percentage of the six positive MRI scores of the ASspiMRI-a on the basis of all VUs assessed in the whole spine for both sequences. The data suggest that 75.8% and 72.7% of the scores were negative (score = 0) for T<sub>1</sub> Gd-DTPA and STIR, respectively.

**Table 2** Disease affection for all evaluated patients and VUs for each reader and for each reading, as assessed by the two different MRI techniques by using the ASspiMRI-a

	Affected patients (%)	Affected VUs (%)
<i>Gd-DTPA</i>		
Reading 1	92.1	27.4
Reading 2	86.8	27.2
<i>STIR</i>		
Reading 1	89.5	35.7
Reading 2	84.2	32.9
<i>Gd-DTPA</i>		
Reading 1	89.5	23.2
Reading 2	81.6	27.1
<i>STIR</i>		
Reading 1	89.5	30.9
Reading 2	89.5	27.9

reader for one patient. By means of the normal approximation, the SDD was calculated by 1.812 times the square root of the interrater variance. This ensures an 80% probability that an observed difference is larger or smaller than the measurement error.

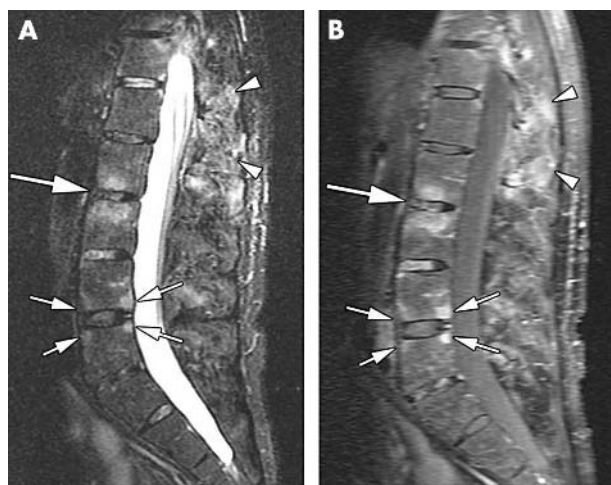
## RESULTS

### Inter- and intrareader reliability of the ASspiMRI for both MRI sequences

Table 1 shows that the intra- and interreader variance was low (<10% of all variance) for both sequences, resulting in high intraclass correlation coefficient (ICC) values. By far the greatest proportion of variance could be attributed to true variance among patients. As a consequence, we obtained high ICC values for scoring the entire spine by both sequences. The ICC values were clearly lower for the three spinal segments separately, with consistently lower scores for the STIR sequence than for the Gd-DTPA sequence (table 1). The analysis of variance results stratified for the spinal site showed that the thoracic spine, as compared with the lumbar and cervical spine, generated by far the highest amount of between-patient (true) variance, and also the highest amount of intra- and interreader variance. The stratified analyses of the three spinal segments also showed poorer ICCs than the analysis of the entire spine, with in general, and as expected, somewhat higher levels of interrater than intrarater variances. As a consequence of low levels of intrarater variance, low SDDs were found. The smallest detectable difference was calculated to be 4.7 for the Gd-DTPA sequence and 5.6 for the STIR sequence (table 1).

### Comparison of the Gd-DTPA and the STIR sequence using the MRI activity score ASspiMRI-a

The overall correlation between the ASspiMRI-a scores obtained with the two MRI techniques, Gd-DTPA and STIR, was rather good ( $r = 0.84$ ;  $p = 0.01$ ). The distribution of the scorings showed a clear preponderance for the lower scorings (0 and 1–3) rather than the higher scorings (4–6), but no important differences between the two MRI sequences. Scorings of 1–3 were found in 20.3% of the VUs in the Gd-DTPA sequence and in 25% of the VUs in the STIR sequence, while scorings of 4–6 were found in 3.9% in the Gd-DTPA sequence and in 2.3% in the STIR sequence (fig 1).



**Figure 2** (A) Sagittal STIR sequence; (B) sagittal T<sub>1</sub> weighted sequence with fat saturation after administration of Gd-DTPA. Florid anterior and posterior spondylitis at the L4/5 level (small arrows) and spondylodiscitis at the L2/3 level (large arrow) as depicted by the two techniques. Additionally, there is severe enthesitis of the interspinous ligaments in the T12–L2 region (arrowheads).

**Table 3** Concordance and discordance between the two MRI techniques for each reader and for each reading. Evaluation was performed by comparing the same VUs of each patient by both MRI techniques for inflammation, as assessed by the ASspiMRI-a

	Concordant VU pairs (%)	Discordant VU pairs (%)
Reader 1		
Reading 1	80.1	19.9
Reading 2	84.5	15.5
Reader 2		
Reading 1	81.1	18.9
Reading 2	89.0	11.0

Overall, the level of involvement was high, with a range of 23.2%–35.7% of all VUs, and with 81.6%–92.1% of all patients showing at least one inflammatory lesion (table 2). The percentage of involvement was similar among different readers and among different readings.

Concordance of the scorings, defined as the percentage VUs with spinal inflammation (score  $\geq 1$ ) compared with the percentage without inflammation (score 0), in both MRI sequences and for both readers/readings separately were found in 83% of the VUs scored (fig 2), with minor variation across the three segments (80.5% concordance in the VUs of the cervical spine, 83.3% in the VUs of the thoracic spine, and 87.7% in the VUs of the lumbar spine). The level of discordance was higher when using the STIR sequence than when using the Gd-DTPA sequence: STIR showed inflammation in 10.1% of the VUs that were found to be normal with Gd-DTPA. In contrast, STIR showed no inflammatory lesions in 6.4% of the VUs in which Gd-DTPA identified inflammation. Table 3 shows an analysis of concordant VU pairs for each reader and for each reading separately, both at the patient level and at the level of single VUs. The percentage of concordant observations was similar among both readers, but appeared to increase slightly in the second reading as compared with the first one in both readers.

More inflammatory spinal lesions were seen by the STIR sequence than by the Gd-DTPA sequence: inflammation was present in 30.6% of the VUs, as assessed by STIR, compared with 26.8% of the same VUs when assessed by Gd-DTPA for the entire spine ( $p = 0.001$ ). A detailed evaluation of the three spinal segments showed inflammation in 20.7% and 16% of the VUs in the cervical spine for the STIR and the Gd-DTPA sequence, respectively ( $p < 0.05$ ), in 38.7% and 34.5% for the thoracic spine ( $p < 0.001$ ), and in 23% and 20.3% for the lumbar spine for the STIR and Gd-DTPA sequences, respectively ( $p < 0.05$ ).

## DISCUSSION

MRI techniques are rapidly gaining importance in the evaluation of acute spinal inflammation in AS. In this study we evaluated the two most important techniques and their prominent findings: enhancement of Gd-DTPA seen in T<sub>1</sub> sequences with fat saturation technique after application of the contrast agent and/or the bone marrow oedema seen by the STIR technique with intrinsic fat suppression. For evaluation of these MRI changes we used the recently proposed scoring system ASspiMRI-a, which was developed by our group.<sup>2</sup> For both sequences we compared the intra- and interreader reliability and the sensitivity of the sequences to detect inflammatory lesions in the entire spine.

The results of our study confirmed a high level of intra- and interreader reliability for both sequences, when

evaluated by the ASspiMRI-a. The ICCs obtained for scores of the whole spine were good; those of the separate parts of the spine were poorer, except for the thoracic spine. The latter observation is due to the fact that this site is the source of a higher level of variability in scorings, and emphasises the importance of scoring the entire spine instead of only scoring part of the spine. This is also in agreement with the concept that a combination of a higher number of components leads to higher reliability. Although intra- and interreader reliability of STIR appeared to be somewhat worse than that of Gd-DTPA, the differences in ICC were small, and probably not of considerable importance. This study also shows that the two MRI techniques analysed by using the ASspiMRI-a scoring system have face validity.<sup>8</sup> MRI is the most reasonable way to assess spinal inflammation. Furthermore, the correlation between the ASspiMRI-a and C reactive protein suggest that the criterion validity is good (data not shown).

Active lesions in the spine of patients with AS can be detected by both STIR and T<sub>1</sub> Gd-DTPA MRI sequences. Both readers consistently saw more active lesions with STIR than with Gd-DTPA, which suggests that STIR is somewhat more sensitive to signals representing inflammatory activity than Gd-DTPA. Obviously, it is not known whether this higher sensitivity reflects a true or false signal. The analysis of concordant and discordant observations showed that both techniques provide complementary information: some "STIR negative" patients/VUs appeared to be "Gd-DTPA positive", and vice versa. However, STIR positive observations were more often Gd-DTPA negative than the Gd-DTPA positive observations were STIR negative. The latter observation suggests—but does not prove—that there may be some overreporting when using the STIR technique, and that the Gd-DTPA technique is more selective and specific. An additional explanation for this phenomenon may also be that STIR was read first, and Gd-DTPA afterwards, which may indicate a potential recall bias. In clinical practice Gd-DTPA is often used to confirm abnormalities detected by STIR.

An interesting finding that only partially relates to the topic of this manuscript is the observation that concordance rates increased in both readers in their second reading. This may point to a training effect, and adds to the conclusion that STIR and Gd-DTPA do not really differ in their ability to pick up a true inflammatory signal.

Finally, we should mention that the content of the ASspiMRI-a was not the topic of this study. The inclusion of erosions as an additional outcome measure in an activity score has been debated at OMERACT 2004. The prevalence of a score >3 (which indicates inflammation plus erosion) has been in the range of 10% in the studies performed so far (unpublished data). We are currently analysing how much information and sensitivity to change is lost when the ASspiMRI-a is performed without counting erosions.

In summary, both STIR and Gd-DTPA sequences can detect spinal inflammation in patients with AS. It needs to be emphasised that the conclusions of this methodological paper apply to the situation in which the ASspiMRI-a scoring method should be used: clinical trials and observational studies (groups of patients). Clearly, picking up signals in an individual patient (for diagnostic purposes and for differential diagnosis) is a different issue. Thus, our conclusion is that, for the purpose of clinical studies and group comparisons, STIR and Gd-DTPA perform equally well, and that feasibility should determine whether both or just one technique should be used. At present, it cannot be finally decided whether one technique provides "truer" findings of spinal inflammation than the other because there is no "gold standard". Possibly, some information is gained in individual patients when both sequences are available. In clinical practice, the STIR technique is likely to be preferred for feasibility reasons (costs, time).

#### Authors' affiliations

**X Baraliakos, J Brandt, J Braun**, Rheumazentrum Ruhrgebiet, Herne, Ruhr University Bochum, Germany

**K-G A Hermann**, Department of Radiology, Charité Medical School, Berlin, Germany

**R Landewé, D van der Heijde**, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, The Netherlands

**J Listing**, German Rheumatism Research Centre, Berlin, Germany

**W Golder**, Department of Radiology, DRK Kliniken Westend, Berlin, Germany

**M Rudwaleit, J Sieper**, Department of Rheumatology, Charité Medical School, Campus Benjamin Franklin, Berlin, Germany

**M Bollow**, Department of Radiology, Augusta-Krankenanstalt, Bochum, Germany

#### REFERENCES

- 1 **Spoorenberg A**, de Vlam K, van der Linden S, Dougados M, Mielants H, van de Tempel H, et al. Radiological scoring methods in ankylosing spondylitis. Reliability and change over 1 and 2 years. *J Rheumatol* 2004;**3**:125–32.
- 2 **Braun JX**, Baraliakos W, Golder J, Brandt M, Rudwaleit J, Listing M, 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.
- 3 **McGonagle DW**, Gibbon P, O'Connor M, Green C, Pease, Emery P. Characteristic magnetic resonance imaging enthesal changes of knee synovitis in spondylarthropathy. *Arthritis Rheum* 1998;**41**:694–700.
- 4 **van der Linden S**, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum* 1984;**27**:361–8.
- 5 **Braun J**, Bollow M, Sieper J. Radiologic diagnosis and pathology of the spondyloarthropathies. *Rheum Dis Clin North Am* 1998;**24**:697–735.
- 6 **Braun JD**, van der Heijde. Imaging and scoring in ankylosing spondylitis. *Best Pract Res Clin Rheumatol* 2002;**16**:573–604.
- 7 **Baraliakos X**, Landewe R, Hermann KG, Listing J, Golder W, Brandt J, et al. Inflammation in ankylosing spondylitis: a systematic description of the extension and frequency of acute spinal changes using magnetic resonance imaging. *Ann Rheum Dis* 2004;**63**:730–4.
- 8 **Boers M**, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998;**25**:198–9.