Assessment of acute spinal inflammation in patients with ankylosing spondylitis by magnetic resonance imaging (MRI) – a comparison between contrast enhanced T1 and short-tau inversion recovery (STIR) sequences

Xenofon Baraliakos, Kay-Geert A Hermann, RBM Landewé, Joachim Listing, Werner Golder, Jan Brandt, Martin Rudwaleit, Matthias Bollow, Joachim Sieper, DMFM van der Heijde, and Jürgen Braun

DISCLAIMER

The initial version of *ARD Online First* articles are papers in manuscript form that have been accepted and published in *ARD Online* but they have not been copy edited and not yet appeared in a printed issue of the journal. Copy editing may lead to differences between the *Online First* version and the final version including in the title; there may also be differences in the quality of the graphics. Edited, typeset versions of the articles may be published as they become available before final print publication.

Should you wish to comment on this article please do so via our eLetter facility on *ARD Online* (http://ard.bmjjournals.com/cgi/eletter-submit/ard.2004.031609v1)

DATE OF PUBLICATION

ARD Online First articles are citable and establish publication priority. The publication date of an *Online First* article appears at the top of this page followed by the article's unique Digital Object Identifier (DOI). These articles are considered published and metadata has been deposited with PubMed/Medline.

HOW TO CITE THIS ARTICLE

Baraliakos X, Hermann KGA, Landewé RBM, *et al* Assessment of acute spinal inflammation in patients with ankylosing spondylitis by magnetic resonance imaging (MRI) – a comparison between contrast enhanced T1 and short-tau inversion recovery (STIR) sequences *Ann Rheum Dis* Published Online First [date of publication]*. doi: 10.1136/ard.2004.031609

*Replace with date shown at the top of this page - remove brackets and asterisk

Online First articles are posted weekly at http://ard.bmjjournals.com/onlinefirst.shtml

Assessment of acute spinal inflammation in patients with ankylosing spondylitis by magnetic resonance imaging (MRI) – a comparison between contrast enhanced T1 and short-tau inversion recovery (STIR) sequences

Baraliakos X[1], Hermann K-G A[2], Landewé R[3], Listing J[4,] Golder W[5], Brandt J[1], Rudwaleit M[6], Bollow M[7], Sieper J[6], van der Heijde D[3], Braun J[1]

Rheumazentrum Ruhrgebiet, Herne[1], Department of Radiology, Charité Medical School, Berlin[2], German Rheumatism Research Centre, Berlin[4], Department of Radiology, DRK Kliniken Westend, Berlin[5], Department of Rheumatology, Charité Medical School, Campus Benjamin Franklin, Berlin[6], Department of Radiology, Augusta-Krankenanstalt, Bochum[7], all Germany, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, The Netherlands[3]

Address for correspondence:
Prof. Dr. J. Braun
Rheumazentrum Ruhrgebiet
Landgrafenstr. 15
44652 Herne
Germany
Tel.: 0049 2325 592 130
Fax: 0049 2325 592 136
Email: j.braun@rheumazentrum-ruhrgebiet.de

Key words: ankylosing spondylitis - MRI - inflammation - STIR - T1/Gd-DTPA - ASspiMRI Score

Background. Spinal inflammation is a hallmark of ankylosing spondylitis (AS). Magnetic resonance imaging (MRI) has been found useful to detect acute spinal lesions in such patients.

Objectives. To compare the performance of two different MRI sequences: T1-weighted fat saturated spin echo after application of contrast medium and short-tau inversion recovery (STIR) sequences to detect spinal inflammation in AS patients.

Methods. MRI was performed in 38 AS patients (64.1% male, mean age 40.9 years, 91% HLA B27-positive) with active disease. T1-weighted spin echo sequences with fat saturation after application of gadolinium-diethylenetriamine-pentaacetic acid (Gd-DTPA) and short-tau inversion recovery (STIR) sequences were evaluated and compared by 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 (ICC) were excellent with 0.91 and 0.86 for the Gd-DTPA and the STIR sequence, respectively. Similarly, the overall correlation of the single MRI scores for both sequences was 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 the STIR sequence, respectively. The SDD was calculated to be 4.7 and 5.6 for the Gd-DTPA and the STIR-sequence, 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 T1/Gd-DTPA sequence, while the T1/Gd-DTPA sequence showed inflammatory lesions in 6.4% of the VUs that were found normal by STIR.

Conclusions. Both MRI techniques are able to evaluate active spinal lesions in AS patients. There are somewhat more spinal lesions detected by the STIR sequence, but the reliability between readings and readers may be somewhat better for the Gd-DTPA sequence. The ASspiMRI-a was confirmed to be a reliable instrument to evaluate acute spinal changes in AS.

Introduction

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

Active spinal lesions can be assessed by MRI by using T1-weighted fat saturated sequences after application of contrast agents such as Gadolinium-diethylenetriamine-pentaacetic acid (Gd-DTPA). The presence of an enhancement of the contrast agent is believed to indicate ongoing inflammation. It is unclear whether another MRI technique, the short tau inversion recovery (STIR) sequence, which is known to be able to visualize normal bone marrow and bone marrow edema [3], performs similarly well in this regard. The first scoring system for evaluation of MRI sequences in AS, the ASspiMRI-a, which has been recently proposed and evaluated by our group [2], 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 T1-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

Patient's characteristics

Thirty-eight AS patients were randomly selected and had to fulfill the modified NY classification criteria for AS [4]. Twenty-five of the 38 AS patients were male (64.1%) with a mean age 40.9 years (range 32-54 years) and 91 % of the patients HLA-B27 positive. The mean CRP was 22.2 ± 21.9 mg/dl and the mean ESR 31.2 ± 23.0/1 hour. The patients had active disease with a mean BASDAI of 6.4 ± 1.4 and a mean BASFI of 5.5 ± 2.1 (Tab. 1). Conventional x-rays of the pelvis were available in all patients and, partly, also of the spine.

Magnetic resonance imaging (MRI)

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

1. T1-weighted spin-echo (SE) sequences (repetition time (TR)/echo time (TE) 500/12 ms, slice thickness 3 mm, 4 acquisitions, field of view (FOV) 20 cm * 40 cm, matrix 128 x 512 pixels) before, and

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

No dynamic imaging was performed. Taking C2 and L5 as orientation points the spine was examined in 2 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.

3. Similarly, fat-saturated short tau inversion recovery (STIR) sequences (TR/inversion time (TI)/TE 4,000/150/60 ms, slice thickness 3 mm, 5 acquisitions, FoV 25 cm * 40 cm, matrix 121 x 256 pixels) were performed.

4. T2-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 (Fig. 1), which has been evaluated for assessment of acute inflammatory and possibly simultaneous erosive spinal lesions [2, 6] was used to analyze the MR images of both sequences on a basis of a vertebral unit (VU), which was defined as the region between two virtual lines drawn through the middle of each vertebral body (Fig. 2).

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 [7]. All scorings were done on the basis of single VUs. Concordance and rates were calculated by identifying positive definite inflammatory involvement in both MRI sequences and by comparing always the same VUs in either MRI sequences.

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 variances, the interrater variance and the intrarater variance. The intrarater-variance was estimated in an ANOVA 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- and readings as second factor.

The intrarater variance was used to calculate the smallest detectable difference (SDD) between two readings of one 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 intra-reader reliability of the ASspiMRI for both MRI sequences

As shown in Table 2, that the intra- and inter-reader variance was found to be low (less than 10% of all variance) for both sequences, resulting in high 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 any of both sequences. The ICC values were clearly lower for the three spinal segments separately, with consistently lower scores for the STIR sequence as compared to the Gd-DTPA sequence (Table 2). The ANOVA results stratified for spinal site showed that the thoracic spine, as compared to the lumbar and cervical spine, generated by far the highest amount of between-patient (true) variance, and also highest amount of intra- and inter-reader variance. The stratified analyses of the three spinal segments also showed poorer ICCs as compared to the analysis of the entire spine, with in general, and expectedly, somewhat higher levels of interrater as compared to intrarater variances. As a consequence of low levels of intrarater variance, low smallest detectable differences 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 2).

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

The overall correlation between the ASspiMRI scores obtained with 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) compared to 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 STIR sequence and in 25% of the VUs in the Gd-DTPA sequence, while scorings of 4-6 were found in 3.9% in the STIR sequence and in 2.3% in the Gd-DTPA sequence (Fig. 3).

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 3). The percentage of involvement was rather similar among different readers and among different readings.

Concordance of the scorings, defined as % VUs with spinal inflammation (score ≥ 1) compared to % without inflammation (score = 0), in both MRI sequences and for both readers/readings separately were found in 83% of the VUs scored, with minor variation across the three segments (80.5% concordance in the VUs of the cervical spine (CS), 83.3% in the VUs of the thoracic spine (TS) and 87.7% in the VUs of the lumbar spine (LS)). The level of discordance was higher when using the STIR- as compared to 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. The analysis of concordant VU pairs for each reader and for each reading separately is shown in Table 4, both on the patient level and on 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 to the first one in both readers.

More inflammatory spinal lesions were seen by STIR- as compared to Gd-DTPA sequence: inflammation was present in 30.6% of the VUs, as assessed by STIR, compared to 26.8% of the same VUs when assessed by Gd-DTPA for the entire spine ($p = 0.001$). The detailed evaluation of the three spinal segments showed inflammation in 20.7% and 16% of the VUs in the CS for the STIR and the Gd-DTPA sequence, respectively ($p < 0.05$), in 38.7% and 34.5% for the TS ($p < 0.001$) and in 23% and in 20.3% for the LS and for the STIR and the Gd-DTPA sequence, 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 T1 sequences with fat saturation technique after application of the contrast agent and/or the bone marrow edema seen by the STIR technique with intrinsic fat suppression. For the evaluation of these MRI changes we used the recently proposed scoring system ASspiMRI-a, which was developed by our group [2]. We compared both sequences in terms of intra- and interreader reliability and we assessed the sensitivity of both 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. 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 stresses 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 inter-reader 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 [8]. To assess spinal inflammation, MRI is the most reasonable way to gain that information. Furthermore, the correlation between the ASspiMRI-a and CRP argue for good criterion validity (data not shown).

Active lesions in the spine of AS patients can be detected by both STIR and T1 Gd-DTPA MRI sequences. Both readers consistently saw more active lesions on STIR as compared to Gd-DTPA, which suggests that STIR is somewhat more sensitive to signals resembling inflammatory activity than Gd-DTPA. Obviously, it is not known whether this higher sensitivity reflects true of false signal. The analysis of concordant and discordant observations revealed 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 that Gd-DTPA positive observations were STIR-negative. The latter observation suggests – but does not prove – that there is some over-reporting when using the STIR technique, and that using Gd-DTPA 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 are not really different in terms of picking up true inflammatory signal.

Finally, we like to mention that the content of the ASspiMRI-a was not the topic of this study. The inclusion of erosions as an additional outcome parameter 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 analyzing 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, are capable to detect spinal inflammation in AS patients. It needs to be stressed that the conclusions of this methodological paper apply to the situation in which the ASspiMRI scoring method should be used: clinical trials and observational studies (groups of patients). Clearly picking up signal 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 about equally well, and that feasibility issues determine whether both or just one should be applied. As it stands now, it cannot be finally decided whether one technique provides more 'true' findings of spinal inflammation than the other because there is no gold standard. It seems possible that there is a gain in information in individual patients when both sequences are available. In clinical practice, the STIR technique is likely to be preferred for feasibility reasons (costs, time).

References:

1. Spoorenberg, A., de Vlam, K., van der Linden, S., Dougados, M., Mielants, H., van de Tempel, H., van der Heijde, D., *Radiological scoring methods in ankylosing spondylitis. Reliability and change over 1 and 2 years.* J Rheumatol, 2004. **31**(1): p. 125-32.

2. Braun, J., Baraliakos, X., Golder, W., Brandt, J., Rudwaleit, M., Listing, J., Bollow, M., Sieper, J., van Der Heijde, D., *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**(4): p. 1126-36.

3. McGonagle, D., Gibbon, W., O'Connor, P., Green, M., Pease, C. Emery, P., *Characteristic magnetic resonance imaging entheseal changes of knee synovitis in spondylarthropathy.* Arthritis Rheum, 1998. **41**(4): p. 694-700.

4. van der Linden, S., Valkenburg, H.A., Cats, A., *Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria.* Arthritis Rheum, 1984. **27**(4): p. 361-8.

5. Braun, J., Bollow, M., Sieper, J., *Radiologic diagnosis and pathology of the spondyloarthropathies.* Rheum Dis Clin North Am, 1998. **24**(4): p. 697-735.

6. Braun, J. and van der Heijde, D., *Imaging and scoring in ankylosing spondylitis.* Best Pract Res Clin Rheumatol, 2002. **16**(4): p. 573-604.

7. Baraliakos, X., Landewé, R., Hermann, K.G., Listing, J., Golder, W., Brandt, J., Rudwaleit, M., Sieper, J., van der Heijde, D., Braun, J., *Inflammation in ankylosing spondylitis - a systematic description of the extension and frequency of acute spinal changes using magnetic resonance imaging (MRI).* Ann Rheum Dis, 2004(accepted for publication).

8. Boers, M., Brooks, P., Strand, C.V., Tugwell, P., *The OMERACT filter for Outcome Measures in Rheumatology.* J Rheumatol, 1998. **25**(2): p. 198-9.

Tables

	Mean	Minimum	Maximum	Std. Deviation
Age (years)	40.9	32	54	7.95
Disease duration (years)	14.9	2	34	8.59
BASDAI Score	6.4	4.0	8.6	1.39
BASFI Score	5.5	0.9	8.6	2.09
BASMI Score	3.7	1	9	2.15
CRP (mg/l)	22.2	1	89	21.85
ESR (mm/h)	31.2	3	78	22.97

Table 1: Demographic data of the 38 AS patients (BASDAI = Bath Ankylosing Spondylitis Disease Activity Index, BASFI = Bath Ankylosing Spondylitis Functional Index, BASMI = Bath Ankylosing Spondylitis Metrology Index, CRP = C-Reactive Protein, ERS = Erythrocyte Sedimentation Rate)

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

Table 2: Inter- and Intrarater variance of the scores with 95% confidence intervals (CI), intra class correlation coefficients (ICC), and smallest detectable differences (SDD)

			Affected patients	Affected VUs
Reader 1	Gd-DTPA	Reading 1	92.1%	27.4%
		Reading 2	86.8%	27.2%
	STIR	Reading 1	89.5%	35.7%
		Reading 2	84.2%	32.9%
Reader 2	Gd-DTPA	Reading 1	89.5%	23.2%
		Reading 2	81.6%	27.1%
	STIR	Reading 1	89.5%	30.9%
		Reading 2	89.5%	27.9%

Table 3. Disease affection for all evaluated patients and VUs per reader and for each reading, as assessed by the two different MRI techniques by using 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%

Table 4. Concordance and discordance between the two MRI techniques per reader and for each reading. Evaluation was performed by comparing the same VUs of each patient in either MRI technique regarding inflammation, as assessed by the ASspiMRI-a.

Figure legends

Figure 1. The new scoring system ASspiMRI-a for evaluation of acute spinal lesions in patients with ankylosing spondylitis as assessed by Gd-DTPA and STIR MRI

Figure 2. Definition of the Vertebral Unit (VU) for using the ASspiMRI score in the evaluation of MR images in the spine of AS patients.

Figure 3. 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 implicate that 75.7% and 72.7% of the scores were negative (score = 0) for T1 Gd-DTPA and STIR, respectively.

Fig. 4 a. and b.
a. sagittal STIR sequence; **b.** sagittal T1-weighted sequence with fat saturation after administration of Gd-DTPA. Florid anterior and posterior spondylitis at L4/5 level (small arrows) and spondylodiscitis at L2/3 level (large arrow) as depicted by the two techniques. Additionally there is severe enthesitis of the interspinal ligaments in the T12-L2 region (arrowheads).

Figure 1

__ASspiMRI-a__

0 =normal, no lesions

1 =mild enhancement and bone marrow edema, covering $\leq$25% of a VU

2 =moderate bone marrow edema, covering $\leq$ 50% of a VU

3 =severe bone marrow edema, covering >50% of a VU

4 =bone marrow edema <u>and</u> erosion covering $\leq$ 25% of a VU

5 =bone marrow edema <u>and</u> erosion covering $\leq$ 50% of a VU

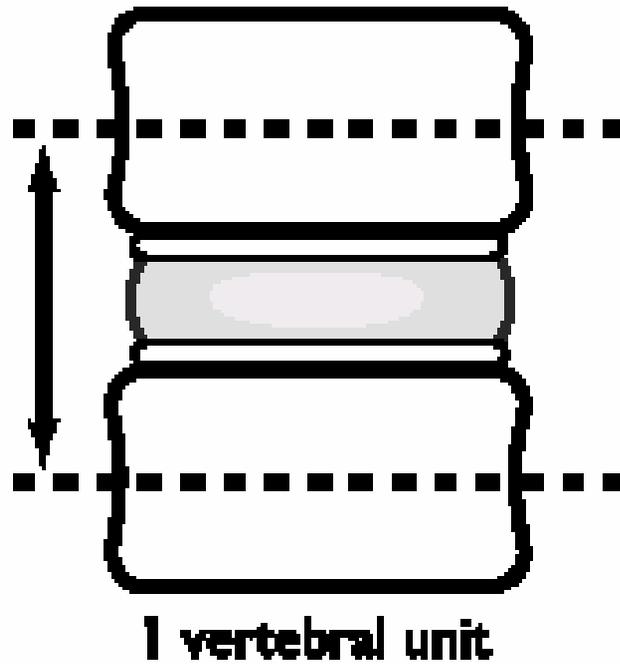6 =bone marrow edema <u>and</u> erosion covering >50% % of a VU

Figure 2
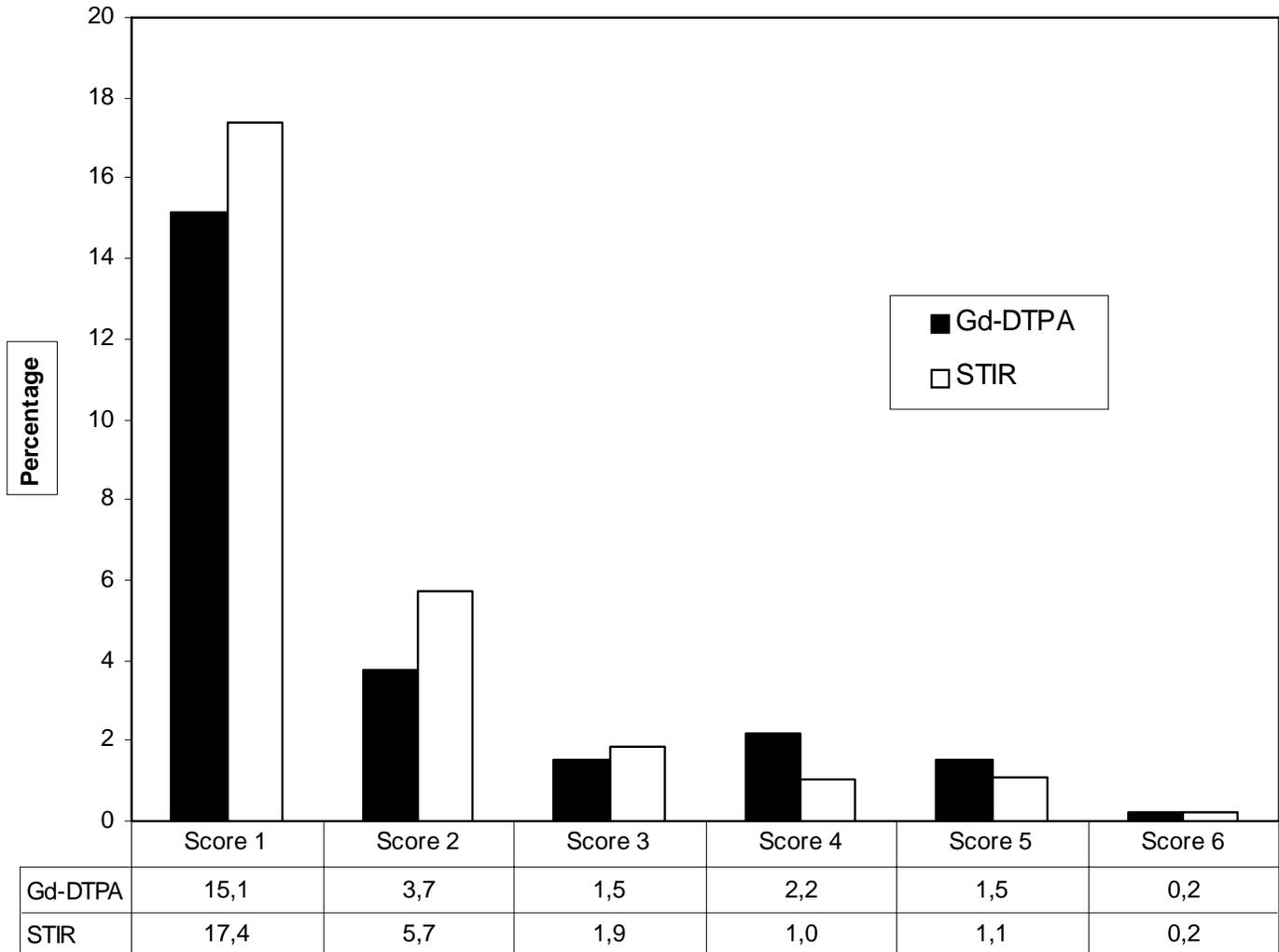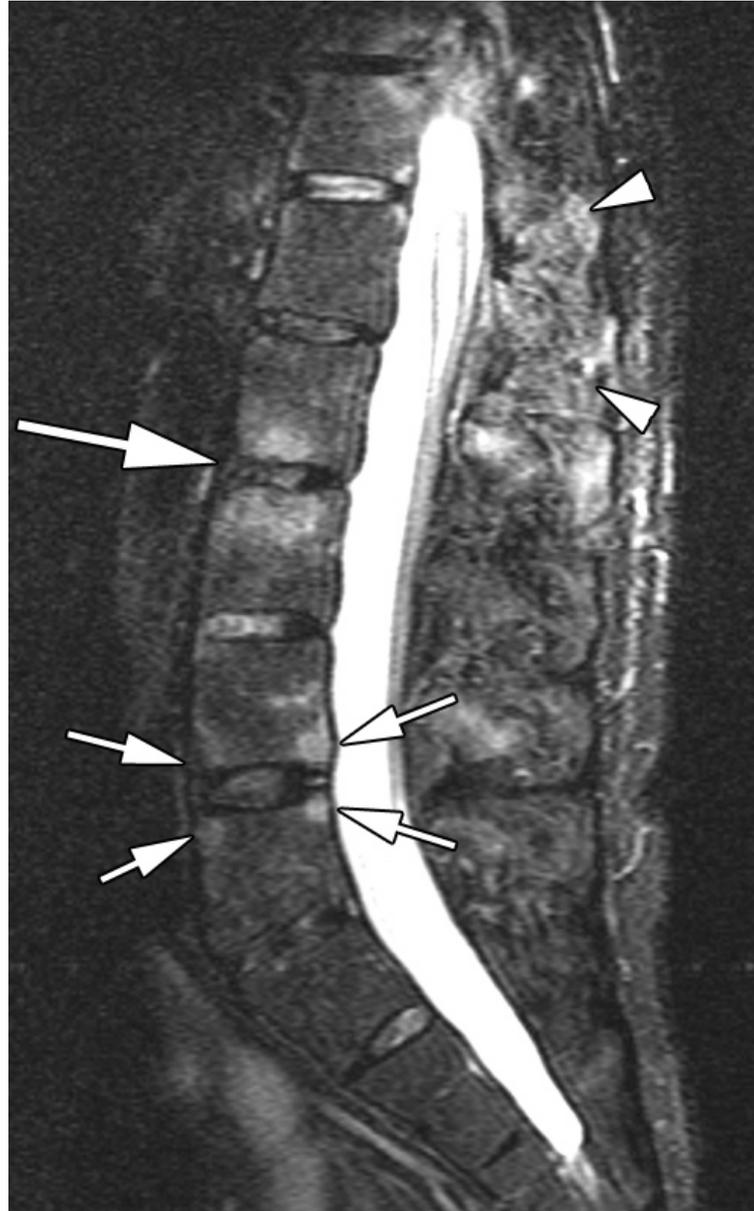

1 vertebral unit

Figure 3

	Score 1	Score 2	Score 3	Score 4	Score 5	Score 6
Gd-DTPA	15,1	3,7	1,5	2,2	1,5	0,2
STIR	17,4	5,7	1,9	1,0	1,1	0,2

Figure 4a

Figure 4b