

EXTENDED REPORT

Analysing chronic spinal changes in ankylosing spondylitis: a systematic comparison of conventional x rays with magnetic resonance imaging using established and new scoring systems

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



Ann Rheum Dis 2004;63:1046–1055. doi: 10.1136/ard.2003.019968

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 22 February 2004
Published online first 5 April 2004

Objectives: To compare conventional radiography and magnetic resonance imaging (MRI) for detection of chronic changes in the spine of patients with ankylosing spondylitis (AS).

Methods: Assessment of chronic lesions in conventional x rays and T₁ weighted MRI turbo spin echo sequences was performed with the established x ray scores BASRI and SASSS, the new Berlin score, and the MRI scoring system ASspiMRI-c. All images were read twice and "blindly" by two readers. One vertebral unit (VU) was defined as the region between two virtual lines drawn through the middle of each vertebra. Definite involvement was defined as a score ≥ 2 in a spinal segment.

Results: Thirty nine patients with AS were examined (25 (64%) male, mean age 40.9 years, 33/36 (92%) HLA-B27 positive). The Berlin score correlated with the BASRI ($r=0.73$, $p=0.01$). The ASspiMRI-c correlated well with the BASRI and the Berlin score ($r=0.66$ and $r=0.51$, respectively, $p=0.01$). The Berlin x ray score showed that 12/35 (34.3%), 13/35 (37.1%), and 12/28 (31.6%) patients had definite involvement of the cervical spine (CS), thoracic spine (TS), and lumbar spine (LS), respectively. The ASspiMRI-c showed that 10/36 (27.8%), 21/36 (58.3%), and 9/35 (25.7%) patients had definite involvement of the CS, TS, and LS, respectively. Syndesmophytes were found in 14.4% of all VUs with 90% agreement between the SASSS and Berlin score.

Conclusions: T₁ weighted MRI can detect chronic lesions in AS. The two new scoring systems proved valid in comparison with established scoring systems and based on aspects of the OMERACT filter. The thoracic spine is most commonly affected in AS. This part of the spine is best assessed by MRI.

Ankylosing spondylitis (AS) is a common chronic inflammatory rheumatic disease that affects young male and female patients.¹ The mean age at onset is 26 years. The leading clinical symptom is inflammatory back pain. The disease starts in the sacroiliac joints and spreads to the spine in the majority of patients.² The axial inflammation comprises sacroiliitis, spondylitis, spondylodiscitis, and spondylarthritis.³ Another major characteristic and pathognomonic sign of AS is new bone formation. Osteoproliferative processes occur often at previously inflamed areas and are detected by imaging techniques as syndesmophytes, calcification, and ankylosis of spinal joints, entheses, and ligaments. These structures are subject to different imaging procedures for assessments of the diagnosis and the course of the disease.⁴ Conventional radiography which has been the "gold standard" in the imaging of AS for the past decades has been included in the internationally well accepted ASAS core set of assessments in AS.⁵

Pelvic x ray measurements are critical for a diagnosis of AS⁶ and spinal x ray measurements have been used to quantify spinal lesions. Scoring systems such as the Stokes AS spinal score (SASSS⁷), the modified SASSS,⁸ and the Bath Ankylosing Spondylitis Radiological Index (BASRI⁹) have been evaluated. In a recent study the SASSS and the BASRI were found to be reproducible, but both had a rather low sensitivity to change.¹⁰ Furthermore, the modified SASSS was found to be the most reliable in comparison with the original SASSS and the BASRI for scoring chronic spinal lesions in AS.¹¹ All three x ray scoring systems assess only parts of the

spine—the SASSS only the lumbar spine, the modified SASSS the cervical and the lumbar spine, and the BASRI the lumbar and the cervical spine and the sacroiliac joints.

Magnetic resonance imaging (MRI) can visualise both acute and chronic inflammation in the sacroiliac joints,^{12–14} in peripheral joints,¹⁵ entheses and the spine^{3 15 16} of patients with AS and other spondyloarthritides. Regression of active spinal lesions after treatment with anti-tumour necrosis factor (TNF) agents has been described by several groups.^{17–19} We have recently proposed a new scoring system to evaluate MR images in the acute and chronic stages of AS.²⁰ The activity score proposed in that system was shown to be reliable and sensitive to change already after 3 months of anti-TNF therapy.²⁰

The assessment of chronic lesions by MRI has not been systematically evaluated to date.

Using the most important x ray scoring systems (modified SASSS, BASRI) to quantify spinal changes of patients with AS, we compared the ability of these scores to detect chronic spinal lesions and compared them with T₁ weighted MR images. In addition to that, and on the basis of the T₁ MRI chronicity score ASspiMRI-c,²⁰ we developed a new scoring

Abbreviations: AS, ankylosing spondylitis; BASFI, Bath Ankylosing Spondylitis Functional Index; BASMI, Bath Ankylosing Spondylitis Metronomy Index; BASRI, Bath Ankylosing Spondylitis Radiological Index; CS, cervical spine; LS, lumbar spine; MRI, magnetic resonance imaging; SASSS, Stokes AS spinal score; SDD, smallest detectable distance; TNF, tumour necrosis factor; TS, thoracic spine; VU, vertebral unit

Table 1 Demographic data of the 39 patients with AS

	Mean	Minimum	Maximum	SD
Age (years)	40.9	32	54	7.95
Disease duration (years)	14.87	2	34	8.59
BASDAI score	6.42	3.98	8.6	1.39
BASFI score	5.45	0.9	8.6	2.09
BASMI score	3.70	1	9	2.15
CRP (mg/l)	22.2	1	89	21.9
ESR (mm/1st h)	31.2	3	78	23.0

system for evaluating native *x* ray findings of the spine of patients with AS, the “Berlin score”. Furthermore, we intended to assess the relative importance of involvement of the spinal segments in AS because this has not been systematically done previously.

PATIENTS AND METHODS

Patients and study protocol

All 39 patients fulfilled the modified New York diagnostic criteria for AS⁶ and were randomly selected. Twenty five of the 39 patients with AS were male (64%), the mean age was 40.9 years (range 32–54 years) and 33/36 (92%) were HLA-B27 positive. The patients had active disease with a mean (SD) BASDAI of 6.4 (1.4) and a Bath Ankylosing Spondylitis Functional Index (BASFI²¹) of 5.5 (2.1). The mean (SD) CRP was 22.2 (21.9) mg/l and the mean ESR 31.2 (23.0) mm/1st h, respectively (table 1).

Conventional *x* rays

All patients with AS either brought *x* ray pictures not older than 3 months with them or they underwent anteroposterior and lateral *x* ray examinations of the spine and an anteroposterior view of the pelvis.

Magnetic resonance imaging

The MRI investigations of the spine of the patients with AS were executed with a 1.5 Tesla Magnetom vision (Siemens, Erlangen, Germany) using a spine coil and/or a body-array coil. The MRI techniques used to investigate sacroiliac and spinal inflammation in patients with AS have been recently described.^{3 15 16} T₁ weighted spin echo sequences (repetition time/echo time 500/14–20 ms, slice thickness 3–4 mm, two acquisitions) were performed in sagittal views. T₁ weighted images were used to detect chronic changes in the bone structure of the vertebra. The spine was examined in two parts, taking C2 and L5 as orientation points, always starting with the upper part. T₂ weighted images were available for comparisons only in doubtful cases, to differentiate between chronic and acute spinal lesions.

Berlin *x* ray score for AS

- 0 = Normal, no lesions
- 1 = Suspicious sclerosis
- 2 = Minor erosions and/or squared vertebral bodies
- 3 = Small single syndesmophytes and/or more severe erosions
- 4 = Two or more syndesmophytes/spondylitis/spondylodiscitis
- 5 = Vertebral bridging
- 6 = Vertebral fusion

Figure 1 The Berlin *x* ray score, a new scoring system to evaluate conventional *x* ray findings of the spine of patients with AS

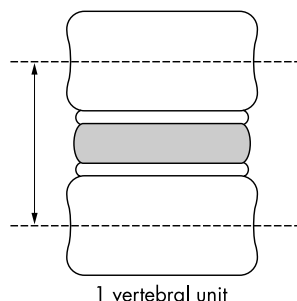


Figure 2 Definition of a VU for using the Berlin *x* ray score and the ASspiMRI-c for T₁ MRI in evaluation of the spine of patients with AS.

ASspiMRI-c

- 0 = Normal, no lesions
- 1 = Minor sclerosis/suspicion of relevant changes
- 2 = Sclerosis/vertebral squaring/ton shaped vertebrae/possible syndesmophyte
- 3 = One or two syndesmophytes/minor erosion
- 4 = More than two syndesmophytes/severe erosions
- 5 = Vertebral bridging
- 6 = Vertebral fusion

Figure 3 Grading system for assessing chronic spinal lesions in patients with AS by T₁ MRI.

Evaluation of the scoring systems

All images were first “blinded” and then evaluated twice by two readers (JB, WG). All readings were performed twice. Thus, each image was evaluated four times in total. The results were added and then divided by four, in order to calculate the means. Because some radiographs had been produced at other sites, some were of limited quality. Some patients did not want to be exposed to radiation again. Because we scored only *x* ray pictures with acceptable image quality, the number of patients in the different analyses may vary. Therefore, for the comparison between the scoring systems, the values are given as means or percentages.

The scores used to analyse the conventional *x* ray findings of the spine in patients with AS were the BASRI,⁹ the modified SASSS,⁸ and the new Berlin *x* ray score (figs 1, 7, 9, 10).

The BASRI was assessed as described,⁹ but no scoring of the sacroiliac joints was performed because we wanted to concentrate on the spinal lesions. For some of the analyses the scores for the cervical spine and the lumbar spine were analysed separately. The hips were also not scored in this study.

The SASSS was used in the modified form as proposed by Creemers *et al.*,⁸ where the anterior vertebral edges of the lumbar and the cervical spine are rated with a score of

Table 2 Inter- and intrarater variance of all three scoring systems (ASspiMRI-c, Berlin x ray score, and SASSS) used to evaluate chronic spinal changes in AS

	Interrater Variance	Intrarater Variance	Variance between patients
<i>ASspiMRI-c</i>			
CS	8.49	4.34	47.80
TS	27.50	11.2	96.12
LS	5.49	2.99	19.79
Spine*	70.54	36.74	341.94
<i>Berlin score</i>			
CS	6.41	3.12	75.12
TS	17.19	5.78	72.95
LS	1.72	1.49	45.47
Spine*	22.59	17.61	541.70
<i>SASSS</i>			
CS	5.67	3.09	74.34
TS	23.93	10.77	104.18
LS	2.84	7.16	69.08
Spine*	75.34	34.2	702.77

*All three segments.
CS, cervical spine; TS, thoracic spine; LS, lumbar spine.

between 0 and 3. This is different from the original SASSS,⁷ where the anterior and the posterior border of only the lumbar spine are scored. In a recent comparison with the BASRI and with the original version of the SASSS, the modified SASSS was found to be the most appropriate method for scoring conventional x rays of the spine of patients with AS.¹¹

In a further analysis of this study we applied the modified SASSS also to the thoracic spine, in order to look for possible differences between the three spinal segments in AS.

The Berlin x ray score (fig 1) was developed by our group in analogy to the T₁ weighted MRI score ASspiMRI-c.²⁰ By this x ray scoring system that uses lateral and anteroposterior views, chronic changes in all three spinal segments are evaluated by assessing 21 (6 in the cervical spine (CS) from C2/3 to C7/T1; 10 in the thoracic spine (TS) from T3/4 to T12/L1; and 5 in the lumbar spine (LS) from L1/2 to L5/S1) vertebral units (VU), which are defined as the region between

two virtual lines drawn through the middle of each vertebra (fig 2). Each VU is scored by a value between 0 and 6, with 0 indicating a normal finding and 6 spinal fusion (fig 1). Grade 1 indicates suspicious sclerosis and grade 2 minor erosions and/or squared vertebral bodies. Grade 3 indicates the presence of small single syndesmophytes and/or more severe erosions. Grade 4 refers to two or more syndesmophytes and/or spondylitis or spondylodiscitis. Bridging of two vertebral bodies leads to a score of 5 and total fusion to a score of 6. Thus, the maximum score for the Berlin x ray score was 6×21 = 126, analogous to the maximal score of the ASspiMRI-c (6×23 = 138), which has a very similar grading from 0 to 6 (fig 3,²⁰).

In this study only the lateral views of all three spinal segments were assessed. The T₁ weighted MR images were analysed by the ASspiMRI-c score as recently proposed (figs 3, 8, 9, 10²⁰) to detect chronic lesions. Because of the analogy between the two scores, there can be a direct comparison between the x ray findings assessed with the Berlin x ray score and T₁ weighted MR images assessed with the ASspiMRI-c.

Statistical analysis

The reliability of the scoring systems was evaluated by estimating the variability between the two raters as well as the variability within every reader. This method was used instead of the calculation of κ values as those are not suitable for ordinal scales.²² 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 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 and readings as second factor.

The intrarater variance was used to calculate the smallest detectable difference (SDD) between two readings of one rater for one patient. By means of the normal approximation, the SDD was calculated by 1.812 times the square root of the intrarater variance. This ensures an 80% statistical confidence that an observed difference is larger or smaller than the measurement error, a random difference. The paired

Table 3 Correlations for each single segment and for the whole spine in scoring systems of the spine of patients with AS

	Berlin score	ASspiMRI-c	Mod. SASSS	BASRI
<i>Cervical spine</i>				
Berlin score	1	0.496**	0.797**	0.871**
ASspiMRI-c	0.496**	1	0.011	0.676**
Mod. SASSS	0.797**	0.011	1	0.781**
BASRI	0.871**	0.676**	0.781**	1
<i>Thoracic spine</i>				
Berlin score	1	0.653**	0.857**	–
ASspiMRI-c	0.653**	1	0.326	–
Mod. SASSS	0.857**	0.326	1	–
<i>Lumbar spine</i>				
Berlin score	1	0.624**	0.944**	0.692**
ASspiMRI-c	0.624**	1	0.513*	0.646**
Mod. SASSS	0.944**	0.513**	1	0.561**
BASRI	0.692**	0.646**	0.561**	1
<i>Spine‡</i>				
Berlin score	1	0.514**	0.613**	0.731**
ASspiMRI-c	0.514**	1	0.015	0.663**
Mod. SASSS	0.613**	0.15	1	0.525**
BASRI	0.731**	0.663**	0.525**	1

*Significant correlation for p<0.05; †significant correlation for p=0.01; ‡all three segments.

Table 4 Means and standard deviations for each spinal segment and for the whole spine, as evaluated by each scoring system for chronic spinal lesions in patients with AS

	Mean	SD
<i>Cervical spine</i>		
ASspiMRI-c	4.98	6.1
Berlin score	9.65	10.44
SASSS	6.56	9.25
BASRI	1.84	1.18
<i>Thoracic spine</i>		
ASspiMRI-c	12	11.15
Berlin score	13.48	9.46
SASSS	15	10.58
BASRI	-	-
<i>Lumbar spine</i>		
ASspiMRI-c	4.5	5.17
Berlin score	6.5	7.42
SASSS	5.07	5.49
BASRI	1.93	1.08
<i>Spine (CS, LS)</i>		
ASspiMRI-c	4.48	1.18
Berlin score	8.01	9.31
SASSS	5.85	8.01
BASRI	3.77	1.01
<i>Spine (CS, TS, LS)</i>		
ASspiMRI-c	6.99	1.2
Berlin score	9.84	9.69
SASSS	8.79	9.87

Wilcoxon rank sum test was used to compare the readings between the scores. Correlation coefficients were calculated by Pearson's method.

RESULTS

Evaluation of the scoring systems

Reliability

The reliability of the ASspiMRI-c was comparable to that of the SASSS and the Berlin score. The values for the TS were acceptable but generally worse than for the other segments. Table 2 shows all the variances found.

The SDD was calculated to be 8.6 for the sum of the spinal evaluation with the Berlin score and 11 for the ASspiMRI-c.

Comparison of the x ray scoring systems

When all three segments of the spine were evaluated together, the Berlin score correlated well with the validated scoring systems BASRI ($r = 0.73, p = 0.01$) and the modified SASSS ($r = 0.61, p = 0.01$). The correlation between BASRI and the modified SASSS was moderate ($r = 0.53; p = 0.01$).

When the spinal segments were evaluated individually, the correlation between the scores showed similar trends as for the whole spine.

In the assessment of the CS, the Berlin score showed a very good correlation with the BASRI ($r = 0.87, p = 0.01$) and the modified SASSS ($r = 0.80, p = 0.01$). The BASRI correlated well with the modified SASSS ($r = 0.78, p = 0.01$).

For the TS, the Berlin score correlated very well with the modified SASSS, ($r = 0.86, p = 0.01$).

For the LS the correlation between the Berlin score and the modified SASSS ($r = 0.94, p = 0.01$) was better than the correlation between the Berlin score and the BASRI ($r = 0.69, p = 0.01$). The correlation between BASRI and the modified SASSS was moderate with $r = 0.56 (p = 0.01)$ (table 3).

Comparison between the x ray and MRI scoring systems

In the evaluation of the whole spine, the ASspiMRI-c showed a similarly good correlation with the BASRI ($r = 0.66, p = 0.01$) and a moderate correlation with the Berlin score

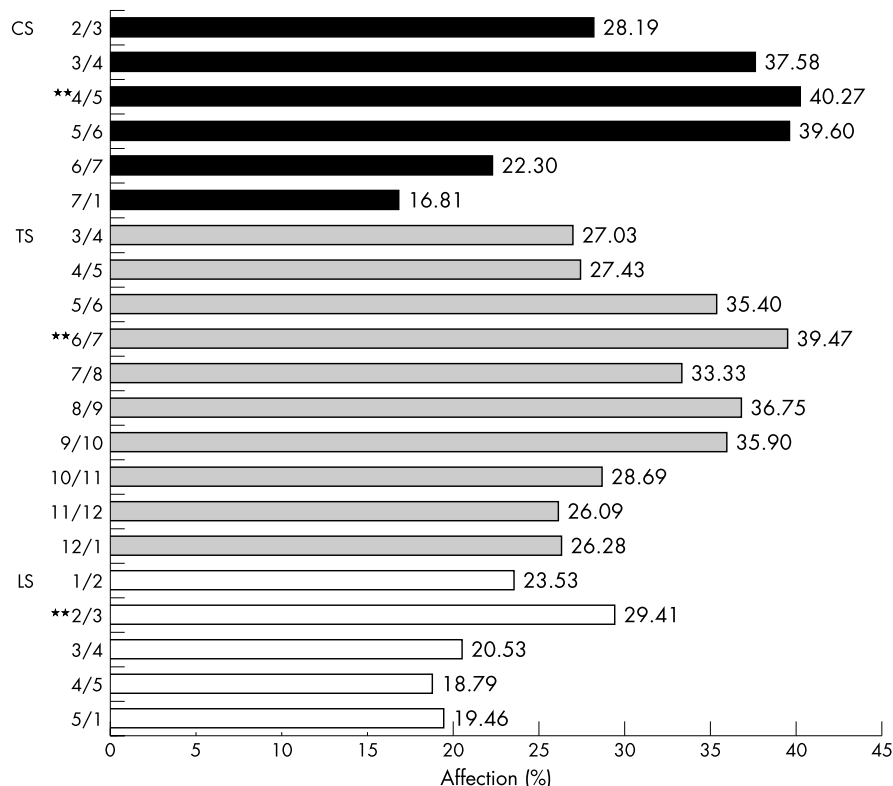


Figure 4 Relative affection of each VU, as seen by x ray examination using the lateral view Berlin x ray score for evaluation of chronic spinal changes in patients with AS. Values are shown as the percentage of affection. **VU with the most frequent affection in each spinal segment.

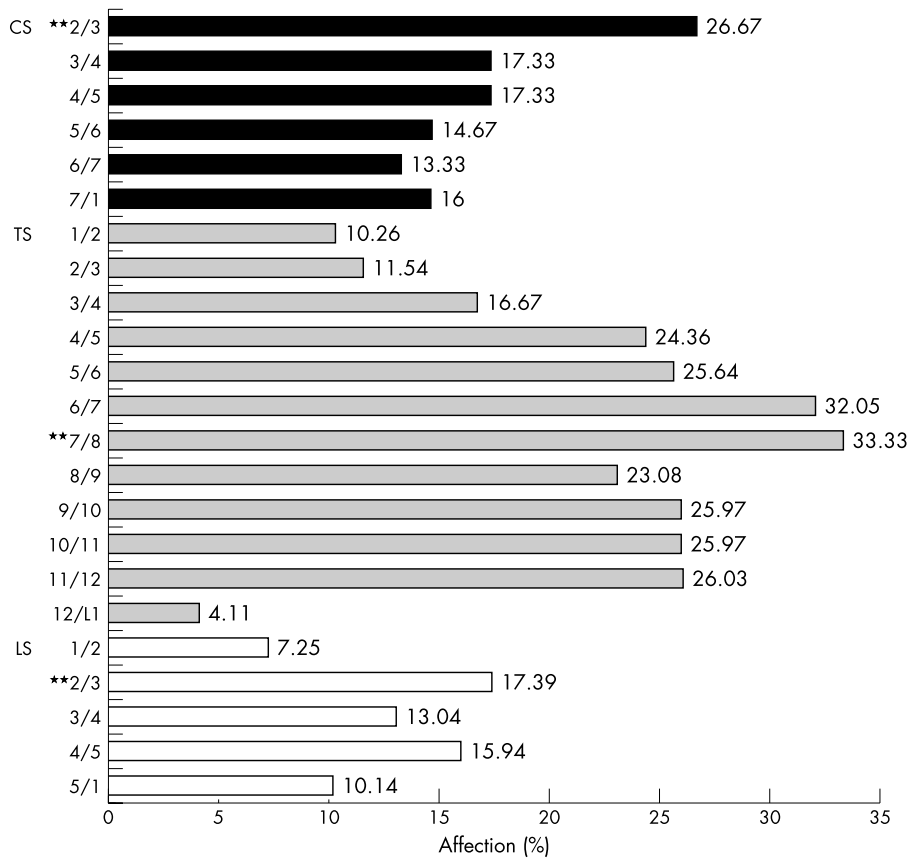


Figure 5 Relative affection of each VU, as seen by x ray examination using the ASspiMRI-c score for evaluation of chronic spinal changes in patients with AS. Values are shown as the percentage of affection. **VU with the most frequent affection in each spinal segment.

($r = 0.51$, $p = 0.01$), whereas it correlated poorly with the modified SASSS ($r = 0.15$; $p > 0.05$) (table 3).

In the assessment of the CS, the ASspiMRI-c correlated well with the BASRI ($r = 0.68$, $p = 0.01$), whereas the correlation with the Berlin score was moderate ($r = 0.50$, $p = 0.01$) and poor with the modified SASSS ($r = 0.11$, $p > 0.05$).

In the evaluation of the TS, the ASspiMRI-c correlated well with the Berlin score ($r = 0.65$, $p = 0.01$) and poorly with the modified SASSS ($r = 0.33$, $p > 0.05$).

The correlations for the LS were better. The ASspiMRI-c correlated well with the BASRI ($r = 0.65$, $p = 0.01$) and moderately with the modified SASSS ($r = 0.51$, $p = 0.01$). The correlation between the ASspiMRI-c and the Berlin score was also rather good ($r = 0.62$, $p = 0.01$) (table 3).

When only the syndesmophytes were considered as a chronic lesion, which occurred in 14.4% of all VUs scored, 90% agreement was found between the SASSS and the Berlin score at the total spine. The agreement between the anteroposterior and lateral views of the TS and the LS was only 84% with an additional 14.7% syndesmophytes seen anteroposteriorly but not laterally, indicating some loss of information when only scoring the lateral views.

There was 84% and 80% agreement between the Berlin score and the ASspiMRI-c score for the total spine and the TS, respectively. The ASspiMRI-c and the Berlin score both detected syndesmophytes in 11.1% in accordance, but there was disagreement in 8% of the VUs in each direction (syndesmophytes detected by one scoring method but not the other).

Analysis of the scoring systems

Total scores and means

For the whole spine, in relation to maximal scores of 138 (ASspiMRI-c) and 126 (Berlin score), only 21 (15.2%) and 29

(23%) mean scoring points for the ASspiMRI-c and the Berlin score, respectively, were counted. Table 4 shows all the mean scores and standard deviations (SD) for the different scoring systems for each spinal segment and for the whole spine.

Detailed analysis

When the data were calculated on the basis of *mean involvement per single VU*, the mean score (SD) for the three spinal segments was calculated as follows:

- For the CS it was 0.83 (1.2) when using the ASspiMRI-c and 1.6 (1.9) for the Berlin score
- For the TS, the mean score per single VU was 1 (1.2) for the ASspiMRI-c and 1.4 (1.2) for the Berlin score
- For the LS, the mean scores per single VU were 0.8 (1.2) and 1.3 (1.7) for the ASspiMRI-c and the Berlin score, respectively
- Finally, the mean scores per single VU in the whole spine were 0.9 (1.2) for the ASspiMRI-c and 1.4 (1.6) for the Berlin score.

With exception of the TS ($p = 0.1$), the differences between the scorings were significant ($p < 0.001$).

In comparison, when using the modified SASSS, the mean (SD) score for each single spinal rim was: in the CS 0.6 (0.9), in the TS 0.7 (0.8), in the LS 0.5 (0.8), and in the whole spine 0.7 (0.9) scoring points.

Figures 4–6 show the detailed analysis of the relative frequency of definite involvement by a disease associated lesion of each single VU. When using the Berlin score (fig 4), the most frequently affected VU in the CS was C4/5 (40.3%) and this was also the case when the whole spine was analysed. In the TS the VU T6/7 was most frequently affected

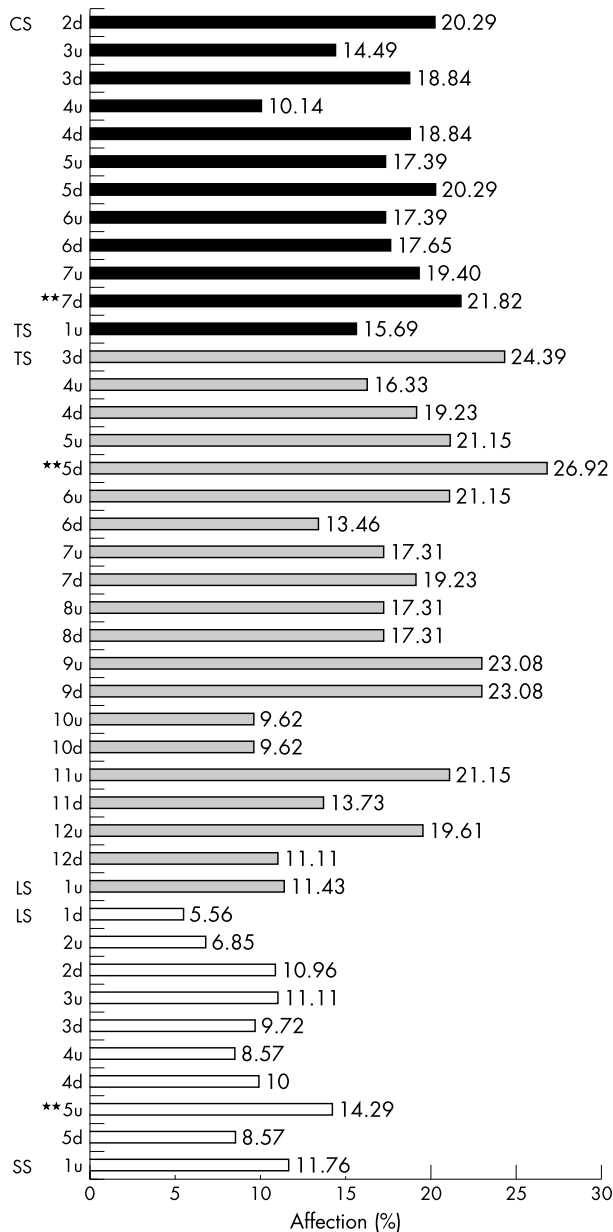


Figure 6 Relative affection of each VU, as seen by x ray examination using the SASS score for evaluation of chronic spinal changes in patients with AS. Values are shown as the percentage of affection. u, upper edge of the VU; d, lower edge of the VU; SS, sacral spine. **VU with the most frequent affection in each spinal segment.

(39.5%). In the LS the VU most frequently affected was L2/3 (29.4%).

When using the ASspiMRI-c (fig 5), the VU most frequently affected in the CS was C2/3 (26.7%) and in the LS L2/3 (33.3%). In the TS it was T7/8 (33.3%), and this was also the most frequently affected VU in the whole spine.

Finally, when using the modified SASSS the vertebral edge affected most frequently in the whole spine and in the TS was the lower edge of T5 (26.9%). In the CS it was the lower edge of C7 (21.8%) and in the LS the upper edge of L5 (14.3%) (fig 6).

Evaluation of the intensity and localisation of spinal involvement in AS

The distribution of the spinal changes due to AS was evaluated by comparing the relative frequency of involvement of

Table 5 Frequency of each grading score (0–6) for x ray and T₁ MR images as evaluated with the Berlin x ray score and the MRI score ASspiMRI-c in the spine of patients with AS

	CS (%)	TS (%)	LS (%)	Spine* (%)
<i>Berlin score</i>				
0	29.96	22.71	46.71	33.13
1	38.44	45.5	30.92	38.29
2	13.80	17.16	8.55	13.17
3	3.66	8.66	3.16	5.16
4	4.83	3.78	2.76	3.79
5	3.90	1.85	5.53	3.76
6	5.41	0.34	2.37	2.71
<i>ASspiMRI-c</i>				
0	46.62	39.22	53.64	46.49
1	27.78	30.83	24.39	27.67
2	11.84	14.40	9.39	11.88
3	7.61	8.87	8.33	8.27
4	1.33	4.64	3.03	3.00
5	3.86	1.79	1.21	2.29
6	0.97	0.24	0	0.40

*All three segments.

the three spinal segments on three levels: (a) in the individual patient; (b) at each spinal segment; and (c) for the mean involvement of one VU per spinal segment. For both, the ASspiMRI-c and the Berlin scores, lesion scores of 0 and 1 indicate no changes and suspicious changes, 2 and 3 indicate minor to mild sclerosis and beginning erosions, and 4–6 indicate severe changes, from growing syndesmophytes to segmental fusion. Therefore, the presence of a scoring value ≥ 2 was taken as a definite sign of affection by the disease.

When this approach was used, and for the *individual analysis of each patient* in the Berlin score, 34.3% of the patients showed a score ≥ 2 in at least one VU in the CS, 37.1% in the TS, and 31.6% in the LS. In comparison, the ASspiMRI-c performed differently, especially in the TS: definite involvement of the TS, as indicated by a scoring value ≥ 2 in at least

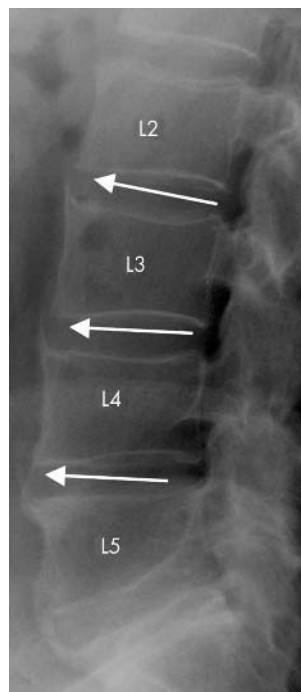


Figure 7 Anterior syndesmophytes between L2/3, L3/4, and L4/5 as detected by x ray examination. By using the Berlin x ray score, the VU L2/3 is scored with a scoring grade of 3 (small syndesmophytes) and VUs L3/4 and L4/5 are scored with a scoring grade of 5 (bridging syndesmophytes).

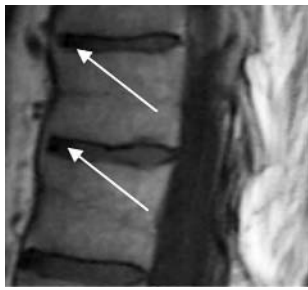


Figure 8 Vertebral anterior bridging between two VUs in the thoracic spine, indicating a score of 5 in the ASspiMRI-c (T_1 MR image).

one VU, was found in 58.3% of the patients, while affection of the CS and the LS was seen in only 27.8% and 25.7% of the patients, respectively.

For the analysis of relative involvement of the spinal segments on the basis of all scored VUs in each segment, we assessed the distribution of the individual scorings between 0 and 6 (table 5). Using this approach for the Berlin score, we found that 31.6% of all VUs of the CS, 31.8% of the TS, and 22.4% of all VUs of the LS had a grading ≥ 2 indicating at least one definite erosion.

These findings were rather similar using the ASspiMRI-c score: 25.6% of all VUs in the CS, 29.9% in the TS, and 22.0% of all VUs in the LS had a score ≥ 2 .

Finally, 18.3% of the VUs had a score of 2–3 and 10.3% had a score of 4–6 using the Berlin score and 20.2% and 5.7% using the ASspiMRI-c, respectively.

In more detail, using the Berlin score, major involvement of the TS (25.8%) was found for the lower gradings (2–3) and similar involvement between CS (17.5%) and LS (11.7%). In contrast, for the higher gradings there was less involvement of the TS (6%) than of the CS (14.1%) and the LS (10.7%).

Using the ASspiMRI-c, the TS showed higher involvement (23.3%) than the CS (19.5%) and the LS (17.7%) in scores 2–3. For the higher gradings, all three spinal segments were equally affected, having scores of 4–6 in 6.7% for the TS, 6.2% for the CS, and 4.2% for the LS.

Validation of the scoring systems based on their correlation with the BASMI and the BASFI

In addition to the comparison of the scoring systems with each other, we proved the validity of the three scoring systems (modified SASSS, Berlin score, ASspiMRI-c) for their correlation with clinical and functional findings, as assessed by the Bath Ankylosing Spondylitis Metronomy Index (BASMI) and the BASFI. As the BASMI does not assess all three spinal segments, the correlations were calculated by comparing the values of each spinal segment with the BASMI. The scoring systems were also compared individually: for the LS the lumbar flexion and the lateral lumbar flexion (as part of the BASMI) with the scores for the LS in each scoring system; for the CS the cervical rotation and the tragus to wall distance, but also the cervical rotation only, with the values of the CS in each scoring system. The lateral lumbar flexion correlated significantly with the three scoring systems: $r = 0.5$ compared with the Berlin score ($p = 0.03$), $r = 0.6$ compared with the SASSS ($p = 0.02$), and $r = 0.4$ compared with the ASspiMRI-c ($p = 0.04$). When only the cervical rotation of the BASMI was compared with the CS scores of the modified SASSS, a significant correlation of $r = 0.5$ was calculated ($p = 0.03$). This was similar for the lumbar part of the BASMI which correlated with the modified SASSS ($r = 0.5$, $p = 0.02$) and for the lateral flexion with the modified SASSS ($r = 0.5$, $p = 0.02$), but not for the other parameters.

Validation of the scoring systems based on aspects of the OMERACT filter

To further validate the two new scoring systems (Berlin x ray score and ASspiMRI-c), we tested their performance by using the aspects of the OMERACT filter: *truth*, *discrimination*, and *feasibility*. The modified SASSS has been found to be the best radiological scoring method for evaluation of spinal changes in AS compared with the original SASSS and the BASRI.¹¹ Therefore, in this study, the modified SASSS was set as the “gold standard” for comparison of the results of the new Berlin x ray score.

DISCUSSION

As far as we know, this is the first study that systematically compares the performance of conventional radiography and MRI in the detection of chronic spinal changes in patients with ankylosing spondylitis. Using the procedure of the OMERACT filter, we discuss several important findings of this study.

Not only acute but also chronic spinal changes in AS can be assessed by MRI. This is substantiated by the rather good correlations of the T_1 score ASspiMRI-c with the BASRI and the Berlin score.

The reliability of the ASspiMRI-c has improved in comparison with our recently published study, in which we reported on the first 20 patients who had undergone consecutive MRI examinations before and after anti-TNF therapy.²⁰ The data presented here suggest that bone changes in patients with AS can be assessed by T_1 weighted MR images and scored using the ASspiMRI-c score. However, these results also show that MRI is no better in the assessment of chronic lesions than conventional x ray examinations—at least for the CS and the LS. This result is relevant for clinical practice because it suggests that MRI should not be the first choice in the initial assessment of patients with AS when the clinical question relates to chronic changes and damage. This is different if the clinical question relates to acute changes.⁴

As shown by the x ray and MRI data, the TS, in comparison with the LS and the CS, is most commonly affected in AS. This is important, because this spinal segment is not assessed by the available x ray scoring systems^{7–9} owing to the reportedly bad reliability caused by technical difficulties raised by the overimposed lung tissue. This problem is confirmed by our data because the inter- and intrarater variances for the TS were worse than for the other segments. However, the variances reported here are still acceptable from a statistical point of view, especially for the Berlin score and the ASspiMRI-c score. This tendency is in accordance with the data from our recent study.²⁰ The intrarater variability was satisfactory, suggesting that the reliability of the scoring systems was rather good. This is especially relevant for the scoring of the TS, for which a high rate of disagreement can be expected because of the anatomy determined by the lung tissue. This well known problem is confirmed by our data showing higher but still fair inter- and intrarater variances in this segment. Because MRI is the best method for visualising the spinal anatomy in the TS, it clearly has a role in assessing chronic changes in this segment.

The Berlin x ray score proved valid and reliable. In comparison with other scoring systems, it is more detailed and it takes longer than the BASRI, but it is more likely to be sensitive to change as shown for the modified SASSS, which performed similar to the Berlin score in this analysis.

The frequent affection of the TS by AS was demonstrated by both the T_1 weighted MRI technique and conventional x ray examination. When the ASspiMRI-c and the Berlin x ray scoring systems were used, the calculation of the means for each single VU in each spinal segment suggests that the

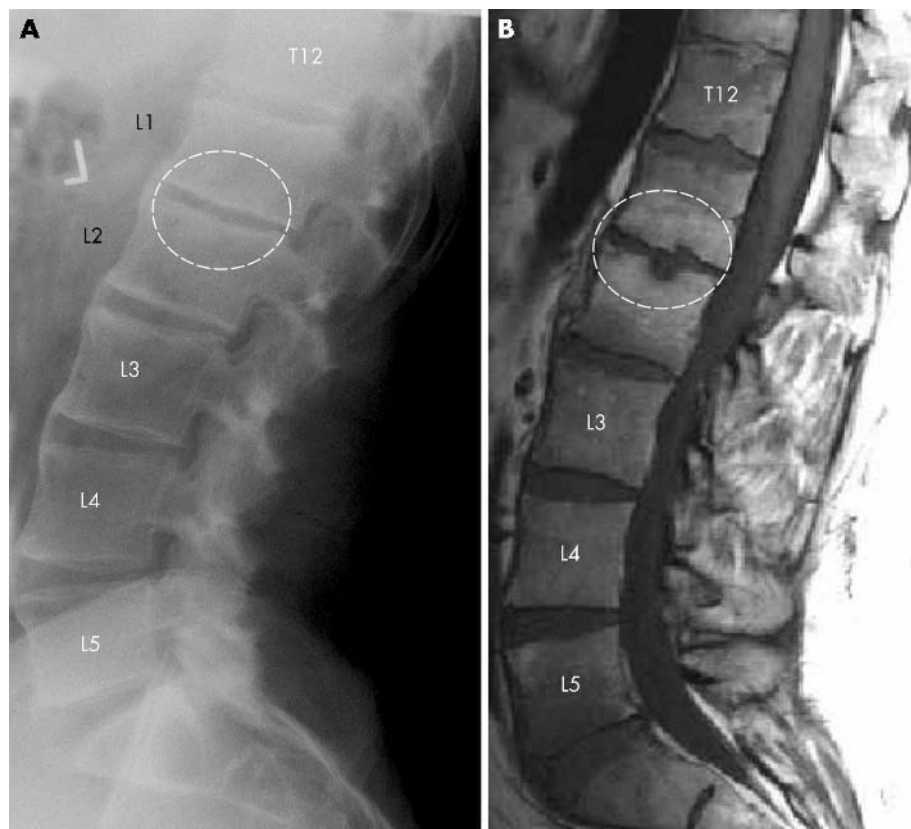


Figure 9 Ventral syndesmophytes and a medial erosion of the lumbar spine as a sign of chronic changes in the spine of a patient with AS. Berlin score (left image): grade 5 (bridging syndesmophytes); ASspiMRI-c score (right image): grade 4 (erosion). One difference between the two imaging techniques is that osteophytes are more easily seen with x rays than with MRI. T₁ weighted MRI can visualise the signs of erosion more clearly than x rays.

explanation for this observation is not only the higher number of vertebral bodies in the TS but also this localisation itself. This observation is of clinical relevance because it indicates the importance of imaging all spinal segments in AS.

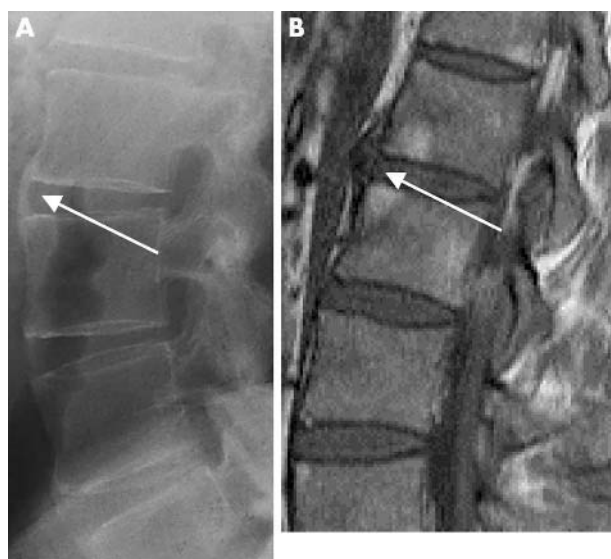


Figure 10 Thoracolumbar region of a patient with AS. The anterior syndesmophytes are clearly seen radiologically (grading 5 and 6 in the Berlin score) (A), whereas T₁ MR images (B) show more indirectly the hypertrophic reaction as a sign of chronic bony changes.

A weak point in all scoring systems is the minor and suspicious changes. Therefore, in order to be able to argue on the basis of reliable data, we took only changes ≥ 2 as a cut off point. Thus, values of 0–1 were excluded from the analysis. The relative frequency of high scoring points indicating severe changes was different for the three spinal segments and also for the scoring systems. Both the ASspiMRI-c and the Berlin score showed a higher frequency of chronic changes for the CS and the TS than for the LS, but the TS was found to be more commonly affected than the CS in the severe form only when using the ASspiMRI-c (table 5). This might be due to the better quality of the images in the MRI technique, especially in the TS, where the image quality of the conventional x ray images is not very satisfactory. Taken together, AS affects all three spinal segments but somewhat differently.

Finally, both new scoring systems presented in this study showed good results when they were compared with aspects of the OMERACT filter (truth, discrimination, feasibility), in comparison with other well accepted scoring systems like the modified SASSS and the BASRI.

Truth

The first aspect of the OMERACT filter is truth, which asks the question whether the proposed method does really measure what is intended. The Berlin score proved valid because it correlated statistically significant with the modified SASSS (table 3). The ASspiMRI-c correlated poorly with the modified SASSS but showed a good correlation with the Berlin score. This might be due to the different grading between these two scores as compared with the rather similar grading between the ASspiMRI-c and the Berlin score, where

both, erosions and hyperproliferation, such as syndesmophytes, are taken into account.

Furthermore, all scoring systems showed a larger number of chronic changes in the TS than in the CS and the LS. Indeed, the TS is not assessed by the modified SASSS but by the ASSpiMRI-c and the Berlin score.

In addition to these comparisons, the correlation between all three scoring systems with the individual aspects of the BASMI for each single segment was significant only for the lateral lumbar spinal flexion (lateral Schober), for the cervical part of the BASMI in comparison with the cervical part of the modified SASSS, and for the lumbar part of the BASMI and the lumbar part of the modified SASSS. No correlation between all the scoring systems in comparison with the BASFI was found.

Discrimination

The second aspect of the OMERACT filter is discrimination, which asks the question whether the proposed methods can differentiate between situations of interest by investigating reliability and sensitivity to change. The last aspect was not assessed in our study because we only evaluated images at a single time.

The reliability of the scoring methods was assessed in this study by comparing the inter- and intrarater variances. In all scoring systems used, variances were rather low when the CS and the LS were evaluated but always somewhat higher for the TS. This is owing to the known technical difficulty of reading thoracic *x* ray images, which is caused by the interference with the lung tissue.

Because the calculated variances of the Berlin score and the ASSpiMRI-c score are acceptable, these scoring methods represent a reliable tool for scoring spinal images of patients with AS, also in respect of the OMERACT filter.

Feasibility

The final aspect of the OMERACT filter is feasibility, which asks the question whether the proposed scoring system is easily performed and in a fair period of time, including the technical requirements for each imaging method and the costs needed for their performance. The time needed for scoring with the BASRI was less than the time needed for scoring with the modified SASSS and the Berlin score. This is because these scoring methods are much more detailed than the BASRI. Furthermore, the ASSpiMRI-c and the Berlin do also score the TS, which is clearly more difficult and time consuming to score than the other two spinal segments because of the overlaid lung tissue. The times needed for scoring with the modified SASSS or with the Berlin score were quite similar. The Berlin score takes only a little more time for the evaluation, as these scores grade the radiological spinal changes between 0 and 4 and 0 and 6, respectively. Scoring MR images is clearly the most time consuming procedure, in general, for the simple reason that MR images are always a set of images and we are dealing with anatomical sections. This means that one has always to look at many images and compare many different sections and mostly at least two different sequences.

The radiation exposure was measured for the *x* ray images with 0.07 mSv for the CS in the lateral view, 0.3 mSv for the TS in the lateral view, 0.5 mSv for the TS in the anteroposterior view, 0.9 mSv for the anteroposterior view and 0.7 mSv for the lateral view of the LS, respectively. Thus, the total radiation exposure for the use of the Berlin score in the lateral view was 1.07 mSv, for the use of the Berlin score in the anteroposterior view was 1.4 mSv, for the use of the BASRI it was 0.77 mSv, and for the use of the SASSS it was 1.07 mSv. In contrast, a clear advantage of the MRI is that it does not use radiation. Importantly, this means that patients

can undergo this examination as often as necessary. Of course, this is especially useful for clinical trials. For the time needed to produce the images: an *x* ray examination of the spine or one spinal segment is performed in a few minutes. In contrast, the time needed to perform an MRI scan is still of the order of 20-30 minutes, depending on the number of sequences needed.

In addition, MRIs are much more expensive than *x* ray examinations and they need more skill. Furthermore, MRIs are not yet widely available. However, MR images have other advantages: (a) they provide the best views and have the best anatomical precision and (b) they can be used to directly demonstrate inflammatory states by either using STIR (short T₁ inversion recovery) sequences or by application of contrast agents such as Gd-DTPA.

In conclusion, the TS is an important target in AS, and T₁ weighted MR images are useful to demonstrate chronic changes in this important part of the musculoskeletal system.

Authors' affiliations

J Braun, X Baraliakos, Rheumazentrum Ruhrgebiet, Herne, Germany
W Golder, Abt. f. Radiologie, DRK Kliniken Westend, Berlin, Germany
K-G Hermann, S Zuehlsdorf, Abt. f. Radiologie, Universitaetsmedizin Berlin, Campus Charité, Berlin, Germany
J Listing, Deutsches Rheumaforschungszentrum, Berlin, Germany
J Brandt, M Rudwaleit, J Sieper, Abt. f. Rheumatologie und Gastroenterologie, Universitaetsmedizin Berlin, Campus Benjamin Franklin, Berlin, Germany
M Bollow, Augusta Krankenhaus, Bochum, Germany
D van der Heijde, University of Maastricht, The Netherlands

REFERENCES

- Braun J, Bollow M, Remlinger G, Eggens U, Rudwaleit M, Distler A, et al. Prevalence of spondylarthropathies in HLA-B27 positive and negative blood donors. *Arthritis Rheum* 1998;**41**:58-67.
- Braun J, Sieper J. The sacroiliac joint in the spondyloarthropathies. *Curr Opin Rheumatol* 1996;**8**:275-87.
- Braun J, Bollow M, Sieper J. Radiologic diagnosis and pathology of the spondyloarthropathies. *Rheum Dis Clin North Am* 1998;**4**:697-735.
- Braun J, van der Heijde D. Imaging and scoring in ankylosing spondylitis. *Best Pract Res Clin Rheumatol* 2002;**16**:573-604.
- van der Heijde D, Boonen A, van der Linden S, Boers M. Reading radiographs in sequence, in pairs or random in rheumatoid arthritis: influence on sensitivity to change [abstract]. *Arthritis Rheum* 1997;**40**(suppl):287.
- 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.
- Averns HL, Oxtoby J, Taylor HG, Jones PW, Dziedzic K, Dawes PT. Radiological outcome in ankylosing spondylitis: use of the Stoke Ankylosing Spondylitis Spine Score (SASSS). *Br J Rheumatol* 1996;**35**:373-6.
- Creemers M, Franssen M, van't Hof MA, Gribnau F, van de Putte L, van Riel P, et al. Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system. *Ann Rheum Dis*. Published Online First: 29 March 2004, doi: 10.1136/ard.2004.020503.
- MacKay K, Mack C, Brophy S, Calin A. The Bath Ankylosing Spondylitis Radiology Index (BASRI): a new, validated approach to disease assessment. *Arthritis Rheum* 1998;**41**:2263-70.
- Spoorenberg A, de Vlam K, van der Heijde D, de Klerk E, Dougados M, Mielants H, et al. Radiological scoring methods in ankylosing spondylitis: reliability and sensitivity to change over one year. *J Rheumatol* 1999;**26**:997-1002.
- Wanders A, Landewé R, van der Linden S, Dougados M, Mielants H, van der Tempel H, et al. What is the most appropriate scoring method for ankylosing spondylitis in clinical trials? A comparison of the available methods based on the OMERACT Filter. In 67th Annual Scientific Meeting of the American College of Rheumatology. Orlando/FL, USA, 2003.
- Braun J, Bollow M, Eggens U, König H, Distler A, Sieper J. Use of dynamic magnetic resonance imaging with fast imaging in the detection of early and advanced sacroiliitis in spondylarthropathy patients. *Arthritis Rheum* 1994;**37**:1039-45.
- Braun J, Bollow M, Neure L, Seipelt E, Seyrekbasan F, Herbst H, et al. Use of immunohistologic and in situ hybridization techniques in the examination of sacroiliac joint biopsy specimens from patients with ankylosing spondylitis. *Arthritis Rheum* 1995;**38**:499-505.
- Bollow M, Fischer T, Reisschauer H, Backhaus M, Sieper J, Hamm B, et al. Quantitative analyses of sacroiliac biopsies in spondyloarthropathies: T cells and macrophages predominate in early and active sacroiliitis—cellularity

- correlates with the degree of enhancement detected by magnetic resonance imaging. *Ann Rheum Dis* 2000;**59**:135–40.
- 15 **McGonagle D**, Gibbon W, O'Connor P, Green M, Pease C, Emery P. Characteristic magnetic resonance imaging enthesal changes of knee synovitis in spondylarthropathy. *Arthritis Rheum* 1998;**41**:694–700.
 - 16 **Muche B**, Bollow M, Francois RJ, Sieper J, Hamm B, Braun J. Anatomic structures involved in early- and late-stage sacroiliitis in spondylarthritides: a detailed analysis by contrast-enhanced magnetic resonance imaging. *Arthritis Rheum* 2003;**48**:1374–84.
 - 17 **Bollow M**, Enzweiler C, Taupitz M, Golder W, Hamm B, Sieper J, *et al*. Use of contrast enhanced magnetic resonance imaging to detect spinal inflammation in patients with spondyloarthritides. *Clin Exp Rheumatol* 2002;**20**(suppl 28): S167–74.
 - 18 **Brandt J**, Haibel H, Cornely D, Golder W, Gonzalez J, Reddig J, *et al*. Successful treatment of active ankylosing spondylitis with the anti-tumor necrosis factor alpha monoclonal antibody infliximab. *Arthritis Rheum* 2000;**43**:1346–52.
 - 19 **Stone M**, Salonen D, Lax M, Payne U, Lapp V, Inman R. Clinical and imaging correlates of response to treatment with infliximab in patients with ankylosing spondylitis. *J Rheumatol* 2001;**28**:1605–14.
 - 20 **Braun J**, Baraliakos X, Golder W, Brandt J, Rudwaleit M, Listing J, *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.
 - 21 **Calin A**, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, *et al*. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. *J Rheumatol* 1994;**21**:2281–5.
 - 22 **Chmura Kraemer H**, Periyakoil VS, Noda A. Tutorial in biostatistics. Kappa coefficients in medical research. *Stat Med* 2002;**21**:2109–29.