Vertebral Erosions Associated with Spinal Inflammation in Patients with Ankylosing Spondylitis Identified by Magnetic Resonance Imaging: Changes After 2 Years of Tumor Necrosis Factor Inhibitor Therapy

Xenofon Baraliakos, Joachim Listing, Hildrun Haibel, Joachim Sieper, and Jurgen Braun

ABSTRACT. **Objective.** Spinal inflammation and erosions have been described in magnetic resonance imaging (MRI) examinations of patients with ankylosing spondylitis (AS). MRI scoring systems have implemented these observations.

**Methods.** MRI scans (T1 or short-tau inversion recovery) from tumor necrosis factor-α blocker (anti-TNF) trials with patients with active AS (n = 22) were analyzed at baseline and after 2 years based on vertebral units (VU). The analysis was based on the prevalence of spinal erosions in relation to inflammation (active erosions) or without it (inactive erosions) as an outcome measure on MRI and their course under anti-TNF therapy. The results of MRI scoring systems that include (ASspiMRI) or exclude (Berlin score) erosions were also compared.

**Results.** At baseline, there were more VU with inflammation (33.7%) than with erosions irrespective of activity (10.6%). After 2 years, active erosions decreased to 3.7% while inflammation was seen in a total of 12% of VU—a reduction of 58.9% and 64.5%, respectively (both p < 0.02). The overall extent of erosions decreased from 10.6% at baseline to 5.6% at 2 years. At the patient level, 73% and 32% of patients showed active erosions (p = 0.002), while 100% and 64% of patients showed inflammation (p = 0.029) at baseline and 2 years, respectively. Both scoring systems showed similar improvement, independent of inclusion or exclusion of erosions.

**Conclusion.** Inflammation with erosions was observed in the spine of most patients with AS but their contribution to changes observed upon anti-TNF therapy was small, indicating that erosions do not need to be included in quantitative scoring systems of inflammation. Spinal inflammation was still present after 2 years of anti-TNF therapy in two-thirds of patients. (First Release Aug 1 2013; J Rheumatol 2013;40:1891–6; doi:10.3899/jrheum.120533)

Key Indexing Terms:
- MAGNETIC RESONANCE IMAGING
- EROSIONS
- INFLAMMATION
- ANKYLOSING SPONDYLITIS

Ankylosing spondylitis (AS) is a chronic inflammatory rheumatic disease that mainly affects the axial skeleton and causes inflammation of the spine and sacroiliac joints (SIJ), leading to pain and stiffness. Sacroiliitis, spondylitis, and spondylodiscitis are the main inflammatory manifestations, while new bone formation, syndesmophytes, and ankylosis of the vertebral column are pathognomonic for the structural changes occurring in the course of the disease. The latter changes can be best identified by conventional radiographs. However, new imaging tools such as magnetic resonance imaging (MRI) have the major advantage of assessing spinal inflammation but can also depict chronic changes. In addition to the typical osteoproliferative changes, patients with AS may also show osteodestructive changes, such as erosions, in both the SIJ and the spine. Analyses of conventional radiographs indicated that such changes occur rather infrequently in AS, while in contrast, they seem to be characteristic and predictive for the development of AS when seen in the SIJ by MRI.

The Outcome Measures in Rheumatology Clinical Trials (OMERACT) filter has been developed to facilitate the evaluation and comparison of different outcome methods for use in rheumatology by application of the aspects of discrimination/sensitivity to change, truth, and feasibility. This procedure has been applied to the 3 scoring systems assessing inflammatory spinal lesions in AS by MRI: the AS spinal MRI (ASspiMRI) scoring system, the Berlin scoring system, and the SpondyloArthritis Research...
Consortium of Canada\textsuperscript{15} scoring system. Based on a multiple-reader MRI scoring exercise, the main result was that all systems worked well, no major differences in the sensitivity to change were found, and no prioritization of one particular system was proposed. However, it must be made clear that the scoring systems have different contents: the Berlin score has been developed on the basis of the ASspiMRI, and the part assessing spinal inflammation is identical in both systems. The additional and special feature of the ASspiMRI is the inclusion of inflammation with erosions: while inflammation without erosions is graded according to the extent of inflammation by grades 1–3, inflammation with erosions (erosions surrounded by inflammation) is quantified by scores of 4–6. This part of the ASspiMRI scoring system had initially been developed not to overlook such osteodestructive changes, which were initially thought to be a direct consequence of inflammation\textsuperscript{6}. Using both scoring systems, inactive erosions are scored only in the absence of inflammation (score 0), which shows that the scoring system is limited in this regard.

The primary objective of our study was to examine the significance of spinal erosions with (active erosions) or without inflammation (inactive erosions) on MRI and their course under tumor necrosis factor-\(\alpha\) inhibitor (anti-TNF) therapy. In addition, we also analyzed the effect of inclusion or exclusion of such inflammation with erosions on the performance of scoring systems, which are widely used as outcome measures in clinical trials.

**MATERIALS AND METHODS**

Patients with AS fulfilling the modified New York criteria\textsuperscript{16} who had participated in clinical trials with infliximab or etanercept\textsuperscript{17,18} were included in this analysis. The ethical committees of the Charité University Hospital in Berlin and the University of Muenster had approved the performance of MRI. All patients gave written informed consent for MRI examinations and agreed on subsequent use of the data from these examinations for retrospective analyses by the investigators.

Standard clinical and laboratory measures to assess disease activity [Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)]\textsuperscript{19}, C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), function (Bath Ankylosing Spondylitis Functional Index; BASFI\textsuperscript{20}), and mobility ([Bath Ankylosing Spondylitis Metrology Index; BASMI\textsuperscript{21}]) were available for all patients at all available timepoints.

MRI examinations included T1 and short-tau inversion recovery (STIR) sequences of the spine in all patients, as described\textsuperscript{13}. Both the STIR and the T1-weighted MR images had to be available at each timepoint for inclusion in our study.

All images were rescanned by one experienced reader who was blinded for the timepoint of performance of the MRI, using 3 different methods: (1) assessment of single vertebral units (VU) in both the anterior and the posterior part of the vertebrae and by quantifying between edges showing active or inactive erosions, but also edges with inflammation only in a binary method (presence yes/no for the different lesions); (2) scoring using the ASspiMRI; and (3) scoring using the Berlin scoring system. Scoring with the 2 scoring systems was based on a VU, defined as the region between 2 virtual lines through the middle of 2 neighboring vertebrae, including the intervertebral disc. In the ASspiMRI, the range of scores within 1 vertebral unit is 0–6. Spinal inflammation was quantified on the basis of the extent of the inflammatory signal (range 1–3), and by the size of potential erosions surrounded by inflammation (active erosions, range 4–6). In the Berlin score, the range of scores within 1 VU is 0–3 and spinal inflammation was quantified on the basis of the extent of the inflammatory signal, similar to the ASspiMRI. Active erosions were scored as inflammatory signal without the area or erosion, thus excluding the scorings of 4–6 taken from the ASspiMRI and concentrating only on the inflammatory part of the assessed VU. The range of the total score for all 23 VU evaluated from C2/3 to L5/S1 is 0–138 with the ASspiMRI and 0–69 with the Berlin score.

To determine the feasibility aspect of the OMERACT filter, the time to score with either system was also recorded.

The paired Wilcoxon rank-sum test was used to compare the readings of the 2 scoring systems as well as the mean number of lesions per patient at different timepoints. The McNemar test was used to compare paired proportions over time. The Spearman correlation coefficient was used to measure the associations among the imaging data between timepoints and single clinical and laboratory measures.

**RESULTS**

Status and change of different types of spinal lesions at baseline and after 2 years of followup. A total of 22 patients (\(n = 1012\) VU) had MRI at both baseline and the 2-year followup. The mean age of the patients was 38 ± 7.3 years, the mean symptom duration was 13.9 ± 8.4 years, and 17/22 (77.3\%) were male. The mean BASDAI at baseline was 6.4 ± 1.3, the mean BASFI was 5.6 ± 1.8, and the mean BASMI was 3.9 ± 2.0; mean CRP was 24.7 ± 25.8 mg/dl and mean ESR was 33.0 ± 24.4 mm/h. The detailed results based on the analysis of single VU for the numbers and rates of any type of lesions together with the magnitude of change between baseline and 2 years are shown in Table 1.

At least 1 active erosion was seen in 16 patients (73\%) at baseline and in 7 patients (32\%) after 2 years (\(p = 0.003\)). In comparison, all 22 patients had spinal inflammation at baseline, while after 2 years, 14/22 patients (64\%) still showed signs of inflammation (\(p = 0.005\); Figure 1).

Overall, the demographic baseline characteristics, including age, disease duration, and CRP, between patients with active or inactive erosions were similar. There was no predilection for active erosions in one particular segment of the spine, in neither the anterior nor the posterior part of the vertebrae (data not shown).

Erosions without inflammation on MRI in patients with AS. Overall, inactive erosions were seen in 0.8 VU per patient (1.8\% of all available VU) at baseline, and in 1 VU/patient (2.1\% of all available VU) after 2 yrs. At least 1 inactive erosion was seen in 8/21 patients (38\%) at baseline and in 11/21 patients (52\%) after 2 years.

The analysis of inactive erosions in the different parts of the spine showed no significant differences between spinal segments: inactive erosions at baseline and after 2 years were found in 3 (1.2\%) and 7 (2.8\%) of 252 VU in the cervical spine, 12 (1.2\%) and 9 (0.9\%) of 966 VU in the thoracic spine, and 2 (0.8\%) and 4 (1.6\%) of 252 VU in the lumbar spine, respectively.

**Correlation of MRI scores and clinical assessments.** Significant correlations were found only for baseline values
of BASMI and baseline values in the scoring systems, with $r = 0.379$ (p = 0.023) for BASMI and ASspiMRI and $r = 0.369$ (p = 0.027) for BASMI and Berlin score. No other assessment correlated with the MRI scores at baseline or at followup.

Further, significant correlations were found between the 2 different scoring systems for both baseline ($r = 0.947$, p < 0.001) and followup ($r = 0.626$, p < 0.001).

**Outcome of scoring systems after inclusion or exclusion of inflammation with erosions.** The ASspiMRI score improved from a total of 345 units at baseline to a total of 92 units after 2 years, which corresponds to an improvement of 73.3% and a mean difference of 11.5 units (95% CI 7.1–15.9) between timepoints. In comparison, the Berlin score improved from a total of 216 units at baseline to a total of 68 units after 2 years, which corresponds to an improvement of 68.5% and a mean difference of 6.7 units (95% CI 3.9–9.5) between timepoints (p < 0.001 for both scoring systems between baseline and 2-year followup, but no significant difference for the changes between scoring systems).

### Table 1. Course of different types of inflammatory lesions over 2 years from 22 patients [total n = 1012 vertebral units (VU)] for whom complete magnetic resonance imaging sets were available.

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>2 Yrs</th>
<th>% of Change (95% CI)</th>
<th>p (2 yrs vs baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erosions (any), n (%)</td>
<td>107</td>
<td>57</td>
<td>46.7 (37.0–56.6)</td>
<td>0.042</td>
</tr>
<tr>
<td>Inflammation (with or without erosions), n (%)</td>
<td>341</td>
<td>121</td>
<td>64.4 (59.2–69.6)</td>
<td>0.011</td>
</tr>
</tbody>
</table>

**Inflammatory lesions without erosion**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>251</td>
<td>84</td>
<td>66.5 (60.3–72.3)</td>
<td>0.001</td>
</tr>
<tr>
<td>% of all VU</td>
<td>24.8</td>
<td>8.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of all inflamed VU</td>
<td>73.6</td>
<td>69.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Active erosions**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>90</td>
<td>37</td>
<td>58.9 (48.0–69.2)</td>
<td>0.018</td>
</tr>
<tr>
<td>% of total VU</td>
<td>8.9</td>
<td>3.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of all inflamed VU</td>
<td>26.4</td>
<td>30.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Erosions (any): erosions with or without inflammation; active erosions: erosions surrounded by inflammation.

Figure 1. STIR-weighted and T1-weighted MRI of the lower spine at baseline and after 2 years of followup. Both possibilities of the course of inflammation with erosion can be detected: inflammation with erosion at baseline that becomes inactive after 2 years (white arrow), and inflammation with erosion at baseline that remains active after 2 years (black arrow).
Feasibility of scoring changes between timepoints with both scoring systems. The mean time for scoring both the STIR and the T1/Gd-DTPA images of 1 patient was 498.6 ± 128.3 s (range 136–909) with the ASspiMRI, and 317.7 ± 112.1 s (range 129–664) with the Berlin scoring system.

**DISCUSSION**

To our knowledge, this is the first MRI-based study that describes in detail the quantity and course of erosions in the spine of patients with active AS who had been treated with TNF blockers. Although not only inflammation but also erosions are frequently found in the majority of patients with active AS, we show that the number of prevalent erosions with or without inflammation was rather small. After 2 years of anti-TNF treatment, the number of erosions was even decreased to a large degree (47%). Whether this decrease was due to anti-TNF therapy seems possible, because osteodestructive changes in spondyloarthropathy (SpA), for example in psoriatic arthritis, have been shown to decrease with use of anti-TNF therapies. Nevertheless, based on our findings, it seems that spinal erosions are unlikely to significantly contribute to the main imaging outcome in studies with AS. This is different when hips are affected by AS.

All patients in our study had high disease activity scores at baseline and had not received antinflammatory medication other than nonsteroidal antinflammatory drugs. In that situation, about 33% of all VU showed signs of spinal inflammation, but only about 25% of these additionally had erosions. Overall, of course, osteoproliferative changes are more important in AS, because 67% of the patients may show osteoproliferative changes in 21% of the vertebral edges on the corresponding conventional radiographs of the cervical and lumbar spine at baseline, as reported.

Yet we still know little about the further course of erosions and their significance for the pathophysiology of the disease. It has been proposed that erosions may precede new bone formation, but the evidence for this, even regarding prediction of clinical or imaging outcomes, is scarce. On the other hand, in our study the total number of erosions decreased over time. It remains uncertain whether mechanisms such as the effect of anti-TNF therapy on bone and its metabolism or on the blockade of the wingless pathway may also play a role in the healing of erosions in the spine of patients with AS.

Regarding the methodology, there are still doubts about the capability of MRI to precisely detect inactive erosions. Based on our clinical experience, it is indeed easier to detect erosions in the spine and in the SIJ when they are surrounded by hyperintense MRI signals, such as inflammation (in STIR) or fatty degeneration (in T1 sequences). A possible solution to this methodological problem would be a systematic comparison of MR and computed tomography (CT) images, because CT is considered the gold standard to depict bony changes in the axial skeleton. Using CT in relatively young patients has been a matter of debate because of the radiation exposure.

Nevertheless, our data analysis shows that erosions were found in < 10% of all VU, and that their contribution to the total changes related to inflammation because of anti-TNF therapy, as frequently described, was rather limited.

Based on our previous studies, the expected reduction of spinal inflammation as detected by MRI was observed in the current study: after 2 years of anti-TNF therapy there were about two-thirds fewer inflamed vertebral edges. However, this implies that spinal inflammation was still present in 12% of all vertebral edges, about 30% of which still had inflammation with erosions. This remaining spinal inflammation was observed in 64% of the patients, 32% of whom still had inflammation with erosions. These data confirm previous MRI data of our group and others, showing that some inflammatory lesions are still present in many patients with AS even after 2 years of continuous anti-TNF treatment. However, even though the relationship between inflammation and new bone formation is incompletely understood and the suppression of spinal inflammation has remained a major aim of therapy in AS. Whether the persistence of spinal inflammation contributes to the new bone formation observed in patients treated with anti-TNF agents remains unclear.

The location of the erosions was not differentially influenced by anti-TNF therapy. Inflammation with erosions may occur at the edge of the vertebral body or in the middle.

While inflammation is considered more specific for AS, erosions cannot be considered per se a sign of AS because they are also detected in patients with degenerative disease. However, both types of lesions show a decrease of the surrounding spinal inflammatory activity upon anti-TNF therapy. The effect of these changes on the development of new bone formation needs to be studied with a larger study sample.

Overall, there were more patients with active than with inactive erosions in our study and the total amount of inflammation with erosions was also greater. Whether and how this is influenced by factors such as physical stress is not clear.

Finally, confirming older data, both scoring systems compared in our study performed similarly. The rate of improvement in the 2 groups was consistent with previous studies, with improvement of about 70% after 2 years. Thus, we could confirm that the Berlin score has a performance similar to that of the ASspiMRI at the group level. However, the Berlin score is more feasible because it could be performed faster (mean time for evaluation of the entire spine of 1 patient was 8 minutes using the ASspiMRI versus 5 minutes using the Berlin score). Of note, the Berlin modification does not count erosions but it does count the...
inflammatory signal irrespective of the presence of erosions. Thus, excluding the erosive part does not lead to loss of information on the magnitude of inflammation in the spine of patients with established AS. Overall, the data presented here confirm data from clinical studies showing that the Berlin and the ASspiMRI are valid scoring systems according to the rules of the OMERACT filter. These data do not challenge the usefulness of other scoring tools.

Our study had 2 main limitations. First, the images were read by one reader, which is different from other studies. However, we believe that the data are valid and that their interpretation is justified because our purpose of the study was to assess the importance of inflammation with erosions as the sole lesions, (2) because the reader was very experienced with the subject and with reading MRI of patients with AS, and (3) because the results of the MRI scorings between baseline and 2 years were very similar to what has been seen in several other studies including those having several readers. A second limitation is the assessment of different MRI lesions based on vertebral units and not on vertebral edges, as we have done in other studies. This may indeed increase the “background noise” and affect the overall quantity of pathologic findings, because the assessment of vertebral edges is more sensitive for the assessment of subtle changes. However, again because the reader was well aware of this problem, we think that this has not influenced the data to a large extent. Further, the MRI scoring system developed by our group some years ago, the ASspiMRI and its modification, the Berlin score, also quantify changes based on vertebral units and not vertebral edges. A similar approach has been used for the study on the link between inflammatory lesions and new bone formation in patients with AS recently. In addition, because we have differentiated the depiction of MRI changes by assessing single VU in both the anterior and the posterior part of the spine, we believe that the chosen approach is not a big limitation but rather a careful quantification of the pathologic MRI lesions studied in our analysis.

Erosions with inflammation occur in many patients with AS, but in < 10% of all vertebral edges, and the inflammation decreases irrespective of the presence of erosions. Thus, erosions have no major effect on MRI-based outcome measures of treatment with TNF blockers. However, our study confirms that there is persisting inflammation independent of erosions after 2 years of anti-TNF therapy. This might affect future therapeutic strategies, because more intensive treatment could be needed in AS patients with active disease and persistent spinal inflammation. Such a treatment change could alter the role of MRI in management strategies of axial SpA, because clinical remission and MRI remission do not necessarily match, as recently shown in the ESTHER trial. Our study also shows that the Berlin modification of the ASspiMRI score is sufficient for the assessment of spinal inflammation in patients with AS. Thus, future research in this area should include (1) more serial MRI scans in patients with and without anti-TNF therapy, and not only because inflammatory activity might fluctuate; and (2) clinical studies using MRI to possibly guide changes in therapy.

ACKNOWLEDGMENT
We thank the patients for their participation and willingness to undergo MR imaging. We thank Beate Buss for help in coordination of collection of MR images.

REFERENCES