Development and Validation of MRI Sacroiliac Joint Scoring Methods for the Semiaxial Scan Plane Corresponding to the Berlin and SPARCC MRI Scoring Methods, and of a New Global MRI Sacroiliac Joint Method

Pernille Hededal, Mikkel Østergaard, Inge Juul Sørensen, Anne Gitte Loft, Jens S. Hindrup, Gorm Thamsborg, Karsten Asmussen, Oliver Hendricks, Jesper Nørregaard, Jakob M. Møller, Anne Grethe Jurik, Lone Morsel, Lone Balding, and Susanne Juhl Pedersen

ABSTRACT. Objective. To develop semiaxial magnetic resonance imaging (MRI) scoring methods for assessment of sacroiliac joint (SIJ) bone marrow edema (BME) in patients with axial spondyloarthritis, and to compare the reliability with equivalent semicoronal scoring methods.

Methods. Two semiaxial SIJ MRI scoring methods were developed based on the principles of the semicoronal Berlin and Spondyloarthritis Research Consortium of Canada (SPARCC) methods. A global quadrant-based method was also developed. Baseline and 12-week MRI of the SIJ from 51 patients participating in a randomized double-blind placebo-controlled trial of adalimumab 40 mg every other week versus placebo were scored by the semiaxial and the corresponding semicoronal methods. Results were compared by linear regression analysis. The reproducibility and sensitivity were evaluated by intraclass correlation coefficients (ICC) and smallest detectable change [SDC, absolute values and percentage of the highest observed score (SDC-HOS)].

Results. Interreader and intrareader ICC were moderate to very high for semiaxial scoring methods (baseline 0.83–0.88 and 0.85–0.97; change 0.33–0.78), while high to very high for semicoronal scoring methods (baseline 0.90–0.92 and 0.93–0.97; change 0.77–0.89). Association between semiaxial and semicoronal scores were high for both the Berlin and SPARCC method (baseline: $R^2 = 0.93$ and 0.88; change: $R^2 = 0.82$ and 0.87, respectively), while lower for the global method (baseline: $R^2 = 0.79$; change: $R^2 = 0.54$). The SDC-HOS were 9.8–18.6% and 5.9–10.7% for the semiaxial and semicoronal methods, respectively.

Conclusion. Detection of SIJ BME in the semiaxial scan plane is feasible and reproducible. However, a slightly lower reliability of all 3 semiaxial methods supports the general practice of using the coronal scan-plane in therapeutic studies. (J Rheumatol First Release October 1 2017; doi:10.3899/jrheum.161583)

Key Indexing Terms: SACROILIAC JOINT MAGNETIC RESONANCE IMAGING SEMIAXIAL SCORING SEMICORONAL SCORING BONE MARROW EDEMA
Magnetic resonance imaging (MRI) of the sacroiliac joints (SIJ) has a key role in the diagnosis and followup of patients with axial spondyloarthritis (axSpA). This role relies on the ability of MRI to detect SIJ inflammation [i.e., bone marrow edema (BME) and/or osteitis] before structural lesions can be detected on conventional radiography and computed tomography, and on MRI being the most sensitive method to detect changes in inflammation.

Several MRI scoring methods for SIJ inflammation have been developed. In clinical trials of patients with axSpA, sacroiliitis is assessed with standardized and validated semiquantitative MRI scoring methods for inflammation, most often the Berlin MRI score or the Spondyloarthritis Research Consortium of Canada (SPARCC) score. For the assessment of inflammation with these methods, the acquisition of a water-sensitive sequence in the semicoronal plane is needed. Usually, a short-tau inversion recovery (STIR) sequence is used for this purpose.

However, there is no evidence concerning which scan plane should be used in clinical practice, and consequently it is most often decided locally based on the local radiologist’s experience and time, and quite frequently STIR images in the semiaxial rather than semicoronal plane are obtained. This may be explained by the impression of a higher sensitivity of the semiaxial scan plane to detection of changes in the ligamentous part of the joint and at the “sacral corners” of the cartilaginous part of the SIJ. Nevertheless, a direct comparison of the ability of the semicoronal versus semiaxial scan plane to detect and monitor lesions has never been performed, to our knowledge.

Consequently, the current MRI SIJ scoring methods do not allow assessment of semiaxial images, and this hinders the use of these methods for systematic followup of MRI of many patients treated in routine care. This is problematic because patients diagnosed and treated in routine care often present a broader spectrum of disease than patients recruited to clinical trials, and they may respond differently to treatment as compared to clinical trial patients.

Therefore, just as clinical registries have been very valuable in providing relevant additional clinical data on patients treated in routine practice, analyses of MRI scans performed in routine care may provide additional important objective information about treatment efficacy. However, to use such MRI in routine care, systematic and validated scoring methods that can be applied in both scan planes are needed.

The aim of our study was to develop and validate 3 semiaxial MRI scoring methods of SIJ inflammation, made by modifying the principles of the Berlin and SPARCC MRI SIJ methods and by developing a new global quadrant-based method. In addition, we compared the reproducibility and sensitivity to change of the semiaxial and semicoronal methods to investigate their performance.

MATERIALS AND METHODS

Patients. The study comprised 52 patients included in a previously described randomized, double-blind, placebo-controlled investigator-initiated study. Patients were randomized in blocks of 4 to treatment with either subcutaneous (SC) injection of adalimumab (ADA) 40 mg or placebo every other week for 12 weeks. To be included in the study, patients had to fulfill the European Spondyloarthropathy Study Group classification criteria for SpA, and have sacroiliitis on MRI and/or conventional radiography as judged by an SpA expert radiologist with 30 years of experience. All patients fulfilled the imaging arm of the Assessment of Spondyloarthritis international Society criteria for axSpA. Further, patients had to have a Bath Ankylosing Spondylitis Disease Activity Score > 40 mm (visual analog scale 0–100 mm) and clinical indication for treatment with tumor necrosis factor-α (TNF-α) inhibitor. Before the MRI were read, 1 patient was excluded because MRI of the SIJ at baseline had not been performed in both the semiaxial and semicoronal planes.

MRI details. MRI were performed at 1.5T according to a standardized protocol at 3 different departments of radiology. The SIJ scans were performed before treatment initiation (i.e., at baseline) and at Week 12. The MRI sequences included a STIR sequence of the SIJ performed in the semicoronal and semiaxial scan plane, respectively.

The STIR sequences were performed with echo time 26–60 ms, repetition time 2500–4200 ms, inversion time 150–170 ms, and field of view (FOV) 240 × 240 mm with matrix 256 × 189 or 256 × 256 (semicoronal) and 224 × 168 or 256 × 256 (semiaxial) or FOV 300 × 300 mm with matrix 240 × 180 or 368 × 274 (both planes). The slice thickness was 4–5 mm and gap 0.5–0.6 mm in both scan planes. During the 12 weeks of study, the individual patients were scanned according to the same MRI protocol, i.e., with the same scan measures within the individual patients.

The MRI were anonymized and evaluated with the readers blinded for clinical, biochemical, and other imaging data. All scans were read on the same MRI workstation. Two readers evaluated the scans according to a standardized reading methodology, which included standardized paper-based score diagrams.

Semicoronal Berlin and SPARCC scoring methods. In the semicoronal Berlin and SPARCC scoring methods, the SIJ are divided into 4 quadrants by a horizontal line going midway through the joint on each image slice, resulting in an upper and lower iliac and sacral quadrant, respectively. In the Berlin scoring method, the volume of BME is graded semiquantitatively for each SIJ quadrant into: 0 No BME; 1: < 33%; 2: 33–66%; and 3: > 66%. The total score range is 0–24.

The SPARCC MRI SIJ Inflammation Index is based on 6 consecutive semicoronal slices covering the cartilaginous part of the SIJ. BME is scored dichotomously (0: absent; 1: present) for each quadrant for each slice. An additional score of 1 is given per joint per slice if the signal is “intense,” corresponding to the signal intensity of the cerebrospinal fluid. Moreover, an additional score of 1 is given per joint per slice if there was a homogeneous and univocal increase in signal extending over a depth of at least 1 cm from the articular surface (“depth”). The total score range is 0–72.

The new semiaxial methods corresponding to the Berlin and SPARCC methods, and a new global method. In the semiaxial Berlin and SPARCC...
methods, the SIJ are also divided into 4 quadrants. On the semiaxial scan, the border between the upper and lower quadrants is defined by the particular image slice that displays the first sacral neural foramen (the image slice in Figure 1E), i.e., corresponding to the border between S1 and S2. The number of semiaxial slices that cover the SIJ from top to bottom is usually 12, and with the definition above there are most often 6 slices covering the upper and lower quadrants, respectively.

In the semiaxial Berlin method, the volume of BME is graded semiquantitatively for each SIJ quadrant with the same methodology as the semicoronal method, providing similar score range (0–24) to the conventional semicoronal Berlin method.

In the semiaxial SPARCC method, the 12 consecutive slices of the cartilaginous part are scored. The scoring approach for BME including depth and intensity is scored according to the same definitions as described for the semicoronal SPARCC. Total score range is 0–96. It is derived from a score of 0–48 for presence of BME + score of 0–24 for presence of “intensity” + score of 0–24 for presence of “depth.”

In the new semicoronal and semiaxial global scoring method, the SIJ are divided into 4 quadrants according to the same definitions as described above for the semicoronal and semiaxial Berlin and SPARCC methods. BME is scored dichotomously (0: absent; 1: present) per quadrant per joint. The total score range is 0–8 in both the semiaxial and the semicoronal scan planes.

Reading exercises. The study comprised 2 reading exercises. Exercise 1 was performed as a prereading exercise to calibrate the readers and to assess the reliability for status scores. Baseline MRI in the semicoronal and semiaxial scan plane were anonymized according to 2 different series of random anonymization numbers. First, the readers independently of each other scored the semicoronal MRI according to the Berlin, SPARCC, and global quadrant method. Thereafter, the readers evaluated all semiaxial MRI according to the corresponding 3 new semiaxial methods.

In exercise 2, the semicoronal and semiaxial scans were reanonymized according to different series of numbers. First, both readers independently of each other read all semicoronal scans and thereafter all semiaxial scans. Baseline and followup scans were scored simultaneously in known time order.

Statistical analysis. Baseline and change scores were characterized with descriptive statistics. Intraobserver and interobserver agreement was assessed with intraclass correlation coefficients (ICC). A 2-way mixed-effects model with the patient as a random factor and the observer as a fixed factor was used and the results were given as single-measure ICC, absolute agreement. An ICC ≤ 0.40 was considered fair, > 0.40 to ≤ 0.60 moderate, > 0.60 to ≤ 0.80 good, > 0.80 to ≤ 1.00 excellent. Scores were compared with independent t tests (between groups) and paired t tests (within groups). The association between the semicoronal and semiaxial methods was estimated for the mean scores of the 2 readers using ICC and linear regression analyses. Bland-Altman plots were performed to visualize any systematic difference between the readers in relation to status and changes scores (data not shown).

Effect size (ES), smallest detectable change (SDC), and smallest detectable difference (SDD) were used to evaluate the sensitivity to change of all methods. ES was calculated as the difference between the mean baseline scores and followup scores at Week 12 divided by the SD of the baseline scores. SDC was calculated as ± 1.96 × SD of the change scores divided by [√2 × √k] and SDD as ± 1.96 × SD of the baseline scores divided by √k, where k is number of readers. The SDC was also calculated as the percentage of the highest observed score (SDC-HOS), defined as the mean of the highest scores of reader 1 and 2, and the percentage of the highest possible score (SDC-HPS), defined as the highest possible score of the scoring method. ES was considered poor if < 0.20, small if...
also showed good to almost perfect correlation (0.72–0.93, methods.

Tables show that the global

p < 0.0001) for the Berlin and SPARCC methods, but with a

 semicoronal versus semiaxial methods for the 2 readers. ICC

92.2% (reader 1) and 88.0% (reader 2).

in the semicoronal versus the semiaxial plane was in exercise

1, 88.0% (reader 1) and 80.3% (reader 2), and in exercise 2,

higher in the treatment group than the placebo group. Our

methods were numerically, although not significantly,

methods. Finally, changes in MRI scores according to all

SDC-HOS were about twice as high for the semiaxial

methods both demonstrated good to excellent reliability,

SDC-HPS, were low for the Berlin and SPARCC methods

baseline to Week 12 by all MRI methods (results not shown).

Sensitivity to change. The SDC, including SDC-HOS and

Table 2 shows the results of

The absolute agreement on presence/absence of BME visualized

in the semicoronal versus the semiaxial plane was in exercise

1, 88.0% (reader 1) and 80.3% (reader 2), and in exercise 2,

92.2% (reader 1) and 88.0% (reader 2).

Table 3 provides the ICC of the comparison of the

semicoronal versus semiaxial methods for the 2 readers. ICC

of the baseline scores were good to excellent (ICC 0.90–0.98)
as compared to the semiaxial methods (ICC 0.77–0.92). For change scores, the

interreader ICC of the semicoronal MRI scores were also good to very good (0.77–0.89), whereas the ICC for the

semiaxial scores were fair to good (0.33–0.78, p < 0.0001).

Comparison of semicoronal and semiaxial methods. The

absolute agreement on presence/absence of BME visualized

in the semicoronal versus the semiaxial plane was in exercise

1, 88.0% (reader 1) and 80.3% (reader 2), and in exercise 2,

92.2% (reader 1) and 88.0% (reader 2).

Table 3 provides the ICC of the comparison of the

semicoronal versus semiaxial methods for the 2 readers. ICC

of the baseline scores were good to excellent (ICC 0.78–0.96, p < 0.0001) for all 3 methods. The ICC for the change score also showed good to almost perfect correlation (0.72–0.93, p < 0.0001) for the Berlin and SPARCC methods, but with a numerically lower correlation (ICC 0.54–0.77) for the global methods.

Table 1. Baseline characteristics of all patients and stratified according to treatment group. Data are mean (SD) unless otherwise indicated.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All Patients, n = 51</th>
<th>Adalimumab, n = 24</th>
<th>Placebo, n = 27</th>
<th>p*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male, n (%)</td>
<td>39 (76.5)</td>
<td>18 (75.0)</td>
<td>21 (77.8)</td>
<td>0.8</td>
</tr>
<tr>
<td>Age, yrs</td>
<td>39.5 (11.1)</td>
<td>40.8 (12.6)</td>
<td>38.3 (9.5)</td>
<td>0.4</td>
</tr>
<tr>
<td>Disease duration, yrs</td>
<td>10.2 (8.3)</td>
<td>11.5 (7.2)</td>
<td>10.2 (9.1)</td>
<td>0.5</td>
</tr>
<tr>
<td>HLA-B27-positive, n (%)</td>
<td>42 (82.4)</td>
<td>23 (95.8)</td>
<td>19 (79.1)</td>
<td>0.09</td>
</tr>
<tr>
<td>BASDAI, 0–100 mm</td>
<td>60.9 (16.4)</td>
<td>60.3 (14.3)</td>
<td>60.6 (18.8)</td>
<td>0.9</td>
</tr>
<tr>
<td>BASFI, 0–100 mm</td>
<td>45.4 (19.0)</td>
<td>47.8 (17.6)</td>
<td>43.3 (20.2)</td>
<td>0.4</td>
</tr>
<tr>
<td>ASDAS</td>
<td>4.4 (1.1)</td>
<td>4.5 (1.1)</td>
<td>4.4 (1.1)</td>
<td>0.8</td>
</tr>
<tr>
<td>CRP, mg/l</td>
<td>15.6 (21.5)</td>
<td>15.5 (14.6)</td>
<td>15.7 (26.5)</td>
<td>1.0</td>
</tr>
</tbody>
</table>

* Independent t test and chi-squared test for comparison of baseline values. BASDAI: Bath Ankylosing Spondylitis Disease Activity Score; BASFI: Bath Ankylosing Spondylitis Functional Index; ASDAS: Ankylosing Spondylitis Disease Activity Score; CRP: serum concentration of C-reactive protein.

Figure 2 shows the semicoronal versus the semiaxial baseline and change scores for exercise 2 with the linear equations best fitting the observed data. Linear regression analyses revealed very good correlation between the semicoronal and semiaxial methods for the Berlin and SPARCC methods (baseline: R² = 0.93 and 0.88 and change: R² = 0.82 and 0.87, respectively) and moderate correlation for the global methods (R² baseline: 0.79 and change: 0.54).

Comparison of treatment groups. No statistically significant differences between the treatment groups were observed (Table 4). Both the treatment group and the placebo group had statistically significant decreases in BME scores from baseline to Week 12 by all MRI methods (results not shown). Sensitivity to change. The SDC, including SDC-HOS and SDC-HPS, were low for the Berlin and SPARCC methods and higher for the global method (Table 2). Moreover, the SDC-HOS were about twice as high for the semiaxial methods as compared to the corresponding semicoronal methods (Table 2). ES and the SDC stratified according to treatment group are presented in Table 4. ES was poor for all the scoring methods in patients treated with ADA and placebo. The standardized response mean was moderate in both treatment groups for all MRI methods.

DISCUSSION

In our study, we developed and validated new SIJ MRI scoring methods for BME applied on the semiaxial scan plane and compared these with the semicoronal methods. The semiaxial methods based on the SPARCC and Berlin methods both demonstrated good to excellent reliability, and good to excellent agreement between semicoronal and semiaxial scores. The SDC were low, indicating high sensitivity to change, although the SDC-HOS were about twice as high for the semiaxial methods as for the semicoronal methods. Finally, changes in MRI scores according to all methods were numerically, although not significantly, higher in the treatment group than the placebo group. Our results support that semiaxial MRI can provide reliable
Our finding of an SDC 3.2 for the semicoronal SPARCC MRI scoring methods for assessment of SIJ inflammation. The ES for all MRI methods (semicoronal and semiaxial) were small for patients treated with ADA and poor for patients treated with placebo. This may partly be explained by not all patients having BME at baseline. In the current study, the ES for the semicoronal Berlin method for patients treated with ADA were higher than previously reported in a cohort of patients with ankylosing spondylitis (AS) treated with TNF-α inhibitor (ES 0.22)\(^2\). The ES for the semicoronal methods indicates that the semicoronal scan plane supports the general practice of using this scan plane in therapeutic studies.

### Table 2. Baseline and change scores for exercise 1 and 2 for the two readers including reliability and sensitivity to change.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Exercise 1</th>
<th>Exercise 2</th>
<th>Exercise 1 vs 2</th>
<th>Exercise 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td></td>
<td>Baseline</td>
<td>Week 0–12</td>
</tr>
<tr>
<td>Score</td>
<td>Mean (SD)</td>
<td>Reliability</td>
<td>ICC (interreader)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Reader 1</td>
<td>1</td>
<td>2</td>
<td>1 vs 2</td>
<td>1</td>
</tr>
<tr>
<td>Berlin</td>
<td>3.2 (5.3)</td>
<td>3.4 (5.8)</td>
<td>0.97</td>
<td>2.8 (5.2)</td>
</tr>
<tr>
<td>SPARCC</td>
<td>10.1 (11.9)</td>
<td>7.4 (12.4)</td>
<td>0.98</td>
<td>6.4 (12.0)</td>
</tr>
<tr>
<td>Global</td>
<td>1.8 (2.5)</td>
<td>1.9 (2.5)</td>
<td>0.98</td>
<td>1.6 (2.3)</td>
</tr>
<tr>
<td>Semi axial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Berlin</td>
<td>2.2 (3.9)</td>
<td>3.9 (5.4)</td>
<td>0.79</td>
<td>2.8 (4.5)</td>
</tr>
<tr>
<td>SPARCC</td>
<td>6.3 (12.1)</td>
<td>8.0 (11.8)</td>
<td>0.91</td>
<td>6.7 (11.4)</td>
</tr>
<tr>
<td>Global</td>
<td>1.4 (2.2)</td>
<td>2.2 (2.5)</td>
<td>0.77</td>
<td>1.8 (2.2)</td>
</tr>
</tbody>
</table>

### Table 3. Comparison of the semicoronal and semiaxial scores for the 2 individual readers and for mean of readers.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Berlin</th>
<th>SPARCC</th>
<th>Global</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reader 1</td>
<td>Baseline Reader 2 Mean of Readers*</td>
<td>Reader 1 Change Reader 2 Mean of Readers*</td>
<td></td>
</tr>
<tr>
<td>0.95</td>
<td>0.91</td>
<td>0.96</td>
<td>0.85</td>
</tr>
<tr>
<td>0.95</td>
<td>0.89</td>
<td>0.94</td>
<td>0.83</td>
</tr>
<tr>
<td>0.89</td>
<td>0.78</td>
<td>0.89</td>
<td>0.77</td>
</tr>
</tbody>
</table>

*Analyses based on mean score of both readers. p < 0.0001 for interreader and intrareader ICC. ICC: intraclass correlation coefficient; SPARCC: Spondyloarthritis Research Consortium of Canada.
SPARCC method is in concordance with results of 2 randomized double-blind placebo-controlled trials of patients with nonradiographic axSpA treated with etanercept (ES 0.39) versus placebo (0.08) for 12 weeks and patients with AS treated with ADA (0.36) versus placebo (0.12) for 24 weeks.

Regardless of the observed high agreement between the semiaxial and semicoronal MRI SIJ scores, we experienced several potential important differences when assessing MRI of the SIJ in semiaxial versus the semicoronal scan plane. On the semiaxial MRI, we had the impression that the most...
The anterior part of the upper anterior surface of the cartilaginous part of the SIJ was poorly visualized because of partial volume effects of bone and psoas muscle owing to the scan plane not being perpendicular to the steep slope of the joint surface in this area. This may be important because our experience is that many patients have BME and capsulitis in relation to the upper anterior superior surface of the SIJ, and it may contribute to the lower semiaxial scores. In contrast, the semiaxial scans easily displayed capsulitis and enthesitis in relation to the anterior inferior and posterior surfaces of the SIJ, which are poorly visualized in the semicoronal scan plane because of slice direction and partial volume effects. Because the SPARCC and Berlin scoring methods focus on BME and not capsulitis and enthesitis, we consider this a minor issue in relation to systematic assessment of the SIJ with the MRI scoring methods. Nevertheless, the very complex location and anatomy of the SIJ call for further studies of the utility of MRI scans performed in the semicoronal versus semiaxial scan plane for diagnosing and prognosticating axSpA and other conditions affecting the SIJ. Moreover, future studies should also investigate whether combined evaluation of the semiaxial and semicoronal sequences are superior to evaluation of one of the scan planes.

Highly standardized MRI protocols of the SIJ are important to avoid reader variation due to differences in scan measures and suboptimal angulation irrespective of scan plane, when comparing changes over time. However, this occurred in only a few patients, and because it concerned both semiaxial and semicoronal slices, we do not regard it as having any major effect on the results. A full validation of a new method requires other steps, e.g., comparison with histology/pathology. Nevertheless, in our study we have shown that both the semiaxial Berlin and SPARCC methods are reproducible and have face, construct, and discriminative validity. Other limitations of our study were the relatively small sample sizes.

This first comparison of semiaxial and semicoronal scoring methods documents that scoring of MRI sacroiliac joint inflammation according to the principles of the SPARCC and Berlin methods can also be done reliably by using semiaxial slices. This indicates that these modified scoring methods can be used to get important additional information from clinical practice cohorts in which semiaxial but not semicoronal STIR images were acquired. Our study also indicates that there is no advantage in using semiaxial in therapeutic studies as compared to semicoronal slices because the reliability and sensitivity to change are slightly inferior, thus supporting that acquisition of semicoronal slices continues to be the standard method used in clinical trials.

ACKNOWLEDGMENT

We thank personnel at the Department of Radiology, Sygehus Lillebaelt, Veje, and the MRI Center Thava Aabenraa for their assistance with the MRI scans and radiography, and the staff at Slagelse Hospital for including and following patients.

REFERENCES


