Development and Preliminary Validation of the Spondyloarthritis Research Consortium of Canada Magnetic Resonance Imaging Sacroiliac Joint Structural Score

Walter P. Maksymowych, Stephanie Wichuk, Praveena Chiowchanwisawakit, Robert G. Lambert, and Susanne J. Pedersen

**Abstract.** Objective. There is an unmet need for reliable assessment of structural progression in the sacroiliac joints (SIJ) of patients with spondyloarthritis (SpA), but radiography is unreliable and lacks responsiveness. We aimed to develop and validate a new scoring method for structural lesions based on magnetic resonance imaging (MRI), the Spondyloarthritis Research Consortium of Canada (SPARCC) SIJ Structural Score (SSS).

Methods. The SSS method for assessment of structural lesions is based on T1-weighted spin echo MRI, validated lesion definitions, slice selection according to well-defined anatomical principles, and dichotomous scoring (lesion present/absent) of 5 consecutive slices through the cartilaginous portion of the joint. Scoring ranges are fat metaplasia (0–40), erosion (0–40), backfill (0–20), and ankylosis (0–20). We progressively conducted 3 validation exercises with 2–4 readers on baseline, and either 2-year (exercises 1 and 2) or 1-year (exercise 3) scans from 147 patients with SpA assessed blinded to timepoint. Interobserver reliability was assessed by intraclass correlation coefficient (ICC) and smallest detectable change (SDC).

Results. Interobserver reliability for status score was good to excellent for ankylosis (ICC 0.79–0.98), consistently good for fat metaplasia (ICC 0.71–0.78), moderate to good for erosion (ICC 0.58–0.62), and fair to good for backfill (ICC 0.35–0.66). Reliability for change scores was moderate to good for all structural lesions despite the relatively small changes in scores, and was highest for fat metaplasia when both ICC and SDC values were compared.

Conclusion. The new SPARCC MRI SSS method can detect structural changes in the SIJ with acceptable reliability over a 1–2-year timeframe, and should be further validated in patients with SpA. (First Release Oct 15 2014; J Rheumatol 2015;42:79–86; doi:10.3899/jrheum.140519)

**Key Indexing Terms:** MRI, VALIDATION, SACROILIAC JOINT, SPONDYLOARTHRITIS
Development of the SPARCC SSS method. Lesion definitions. We conducted standardized definitions of structural lesions of the SIJ on MRI, which were developed by the Canada-Denmark MRI Working Group19 and then extended in a subsequent report to include backfill8. For fat metaplasia to be scored in the SSS method, the lesion has to demonstrate homogeneous signal across the lesion that must extend more than 1 cm in depth from the joint surface because our previous work has shown that this enhances reliability (data not shown). Backfill is defined on T1WSE sequences as the complete loss of the iliac or sacral cortical bone at its anticipated location and increased signal that is clearly demarcated from adjacent normal marrow by irregular dark signal reflecting sclerosis at the border of the eroded bone (Figure 1).

We developed a training module that comprised a detailed description of the SSS scoring method with examples, a Digital Imaging and Communication in Medicine (DICOM)-based reference image set of 45 cases with baseline and 2-year scans, and consensus reader scores for these DICOM images. This comprehensive module is aimed at facilitating calibration of non-expert readers and is available at www.carearthritsis.com. Bone sclerosis and abnormalities of the synovial cavity are not addressed in the SSS method because of poor reproducibility in previous reading exercises (data not shown).

Scoring methodology. Evaluation of structural lesions in the SIJ proceeds sequentially in the following steps: (1) the transitional slice is identified by scrolling from anterior to posterior through the SIJ and viewing DICOM images depicting semicoronal slices through the joint. The transitional slice is defined as the first slice in the cartilaginous portion that has a visible portion of the ligamentous joint when viewed from anterior to posterior (Figure 2); (2) all timepoints are anatomically matched according to the transitional semicoronal SIJ slice. The link function on the DICOM viewing software allows simultaneous scrolling of anatomically matched images from the transitional slice anteriorly, thereby facilitating detection of change in lesions between different timepoints; (3) five consecutive semicoronal slices are assessed starting from the transitional slice and scrolling anteriorly. This number of slices was chosen because our preliminary work showed that the SIJ cavity starting with adjacent bone marrow are still clearly visible at the most anterior slice in virtually all patients (data not shown). Scoring slices posterior to the transitional slice is not appropriate because those slices are predominantly or entirely composed of the ligamentous portion of the joint, which often has irregular bone cortices resembling erosions. Consequently, additional anterior or posterior slices do not provide useful additional information, depict areas that are frequently the hardest to interpret even in normal subjects, and including them in the exercise reduced reader reliability; and (4) the presence/absence of lesions is scored in SIJ quadrants (fat, erosion) or halves (backfill, ankylosis) using a direct online data entry system based on a schematic of the SIJ.

Scoring ranges are fat metaplasia (0–40), erosion (0–40), backfill (0–20), and ankylosis (0–20). The scoring for erosion and fat metaplasia is based on the assessment of SIJ quadrants because those abnormalities can occur on either the iliac and/or sacral sides. Scoring of those lesions is therefore based on all 8 SIJ quadrants. Ankylosis extends from the iliac to the sacral side while backfill fills the joint cavity. Scoring of these lesions is therefore based on SIJ halves. An example of the approach to scoring is provided in Figure 3.

Reading exercises. We conducted 3 formal reading exercises aimed at the assessment of feasibility and reliability. Readings were conducted blinded to patient demographics, time sequence, and treatment according to the following steps: (1) feasibility was first assessed in a pilot exercise of 20 cases randomly selected from the cohort with baseline and 2-year scans scored by 4 readers blinded to timepoint. The average time required to read...
a pair of scans (baseline, 2 yrs) per case was estimated. Because reliability for baseline as well as 2-year change scores was at least good [intraclass correlation coefficient (ICC) ≥ 0.6] for most features (status/change ICC for fat metaplasia = 0.72/0.68, for erosion = 0.60/0.59, for backfill = 0.86/0.55, for ankylosis = 0.98/0.79), further validation was undertaken after debriefing of discrepant cases and clarification of lesion definitions for scoring purposes; (2) in reading exercise 1, 4 readers scored 45 cases with baseline and 2-year scans randomly selected from the cohort to assess whether interobserver reliability for the 4 readers was consistently evident, particularly for change scores, when assessed blinded to timepoint; (3) in

Figure 1. T1WSE MRI scan illustrating an example of backfill. There is complete loss of the right iliac cortical bone at its anticipated location and increased signal (white arrow) that is clearly demarcated from adjacent normal marrow by irregular dark signal, reflecting sclerosis at the border of the eroded bone. T1WSE: T1-weighted spin echo; MRI: magnetic resonance imaging.

Figure 2. T1WSE MRI scan illustrating the transitional slice. This is defined as the first slice in the cartilaginous portion of the SIJ that has a visible portion of the ligamentous joint (arrow) when viewing DICOM images of semicoronal slices through the SIJ from anterior to posterior. T1WSE: T1-weighted spin echo; MRI: magnetic resonance imaging; DICOM: Digital Imaging and Communication in Medicine; SIJ: sacroiliac joints.
Figure 3. Example of SSS method applied to a T1WSE MRI scan from a patient with SpA assessed at baseline and 2 years. Scoring for ER, BF, FAT, and ANK is indicated on the schematic of the SIJ to the right of each MRI scan. Each lesion is indicated as being present/absent on a dichotomous basis (present = 1, absent = 0). A. SSS scores for ER and BF at baseline and 2 years. B. SSS scores for fat and ANK at baseline and 2 years. Fat metaplasia must extend more than 1 cm in depth from the joint surface to be scored. At 2 years, left upper iliac quadrant is not scored even though fat metaplasia is present because it does not extend more than 1 cm from the joint surface. SSS: Sacroiliac Joint Structural Score; T1WSE: T1-weighted spin-echo; MRI: magnetic resonance imaging; SpA: spondyloarthritis; ER: erosion; BF: backfill; FAT: fat metaplasia; ANK: ankylosis; SIJ: sacroiliac joints.
RESULTS

Baseline characteristics and descriptive SP ARCC SSS data. The majority of patients in all 3 exercises were men with longstanding disease receiving TNF-α inhibitor therapy (Table 1). A minority received only prescribed NSAID therapy and some patients received no prescribed therapy, but took over-the-counter antiinflammatory agents. At baseline, the majority of patients in each of the 3 exercises had at least 1 SIJ quadrant with fat metaplasia, erosion, and backfill in the SIJ, and about half in each exercise had at least 1 SIJ quadrant with ankylosis. Mean SSS erosion score decreased over followup while scores for other structural lesions increased. At the patient level, an increase in SSS fat metaplasia score was observed in a mean of 19.7% and a decrease in 15.6% for the 2 readers who read all 147 available baseline and 2-year scans. Similarly, for SSS erosion score, an increase was observed in a mean of 14.6% patients and a decrease in 37.8%. An increase in SSS backfill was observed in a mean of 23.1% of patients and a decrease in 16.7%. The SSS method detected change in structural lesion score for each lesion as soon as 1 year in the majority of patients with any change and the mean (SD) change in each of the 4 structural lesion scores over 2 years. Interobserver reliability for baseline, 1-year, and 2-year change scores was assessed using ICC (3,1). A 2-way mixed effects model with patient as a random factor and observer as a fixed factor was used, and the results are given as single measures. The ICC are presented as absolute agreement for the individual reader pairs and for all readers together (overall ICC). An ICC value of < 0.4 was designated fair; ≥ 0.4 but < 0.6 moderate; ≥ 0.6 but < 0.8 good; ≥ 0.8 but < 0.9 very good; and ≥ 0.9 excellent. We also calculated the smallest detectable change (SDC), which provides an absolute measure of agreement using the Bland-Altman 80% levels of agreement.

Feasibility and interobserver reliability. The time required to assess a pair of scans per case for all structural lesions ranged from 5–15 min. The selection of 5 slices from the transitional slice anteriorly allowed the assessment of structural lesions in the bone marrow of both sides of the joint cavity in all patients. Interobserver reliability for status score assessed at baseline was good to excellent for ankylosis, consistently good for fat metaplasia, moderate to good for erosion, and fair to good for backfill (Table 2). Very good to excellent reliability was achieved by some reader pairs.

Interobserver reliability for change scores was moderate to good for all structural lesions and for all exercises with the exception of erosion in exercise 1 despite the relatively small changes in scores that were recorded (Table 3). Moderate to good reliability was evident for scoring 1-year change in all structural lesions and was highest for fat metaplasia when both ICC and SDC values were compared. SDC values were consistently less than 5% of the maximum score for fat metaplasia in all 3 exercises. Backfill was the most difficult lesion to detect reliably with SDC values of between 5–10% of the maximum score for each of the 3 exercises.

DISCUSSION

We have developed and conducted preliminary validation of a scoring method for structural lesions in the SIJ that is based on the same scoring principles used in the SPARCC MRI SIJ inflammation scoring method. The approach to the selection of MRI slices is anatomically defined, the majority of the cartilaginous portion of the joint is assessed on consecutive slices in the semicoronal plane, and scoring is dichotomous (present/absent), which simplifies assessment and improves reliability. Assessment of status and change scores on a pair of scans can be completed in a feasible time frame. Interobserver reliability for status scores is good to excellent for all lesions and can be achieved early in the calibration process. Good reliability can also be achieved for change scores at the 1-year followup.

Scoring with this method begins with the identification of the transitional slice and then proceeds by evaluating 5 consecutive slices ventrally through the cartilaginous portion of the joint. Scoring slices dorsal to the transitional slice is not appropriate because those slices are predominantly or entirely composed of the ligamentous portion of the joint. In addition, scoring slices more ventrally often results in failure to depict the joint clearly. In both cases, these additional ventral or dorsal slices do not provide useful additional information in the context of scoring structural damage, depict areas that are frequently the hardest to interpret even in normal subjects, and including them in the exercise reduced reader reliability (data not shown). A variable that could influence the application of this method is that the comparison of scores between timepoints can be more challenging if the coronal tilt of slices through the SIJ varies between timepoints. This is easily prevented by stipulating appropriate methodological detail in protocols for imaging in clinical trials and research. We recommend the following protocol for defining the tilted coronal plane. After a triplanar series of scout images, a dedicated small field-of-view sagittal scout series of the sacrum is performed. With reference to the midline sagittal image of the sacrum, the tilted coronal sequences are angled parallel to the posterior cortex of the second sacral vertebral body, which also represents the anterior border of the sacral spinal canal at the S2 level.
A defining principle of the SPARCC methodology for scoring inflammatory and structural lesions in SpA is the application of a dichotomous scoring method based on the simple presence or absence of a lesion per slice in an anatomically defined location such as an SIJ quadrant. The pathological abnormalities visible on MRI often include mixed lesions with complex appearances that may hinder scoring approaches based on estimates of percent volume of an anatomical region occupied by the lesion, especially in the SIJ. Several lesions may be evident concomitantly even in the same SIJ quadrant, and reliable assessment of the volume of the SIJ quadrant occupied by a specific lesion in a joint as morphologically complex as the SIJ may be exceptionally difficult. For example, a lesion visible only in anterior slices will occupy a greater percentage of an SIJ quadrant or the iliac/sacral portion of the joint as compared to the same-sized lesion present only in dorsal slices. For lesions extending across several slices from ventral to dorsal, scoring based on the percent volume occupied by the lesion requires that the reader attempt to mentally compile a 3-D estimate based on all slices assessed.

The SDC was calculated based on 80% limits of agreement and resulted in values that are higher than the mean change scores observed in all 3 exercises. This indicates that there is a risk that patients may be misclassified as having progression when in fact there is measurement error. In our dataset, however, it is also important to note that change in structural lesions may occur in both directions, i.e., increase or decrease, and the mean score reflects a summation at the group level. This limits

Table 1. Baseline patient characteristics and SPARCC MRI SSS, and change over 2 years (exercises 1 and 2) or 1 year (exercise 3) from patients with SpA.

<table>
<thead>
<tr>
<th>Patient Characteristics</th>
<th>Exercise 1, n = 45</th>
<th>Exercise 2, n = 102</th>
<th>Exercise 3, n = 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. readers</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Age, mean (SD)</td>
<td>41.4 (12.6)</td>
<td>39.9 (11.3)</td>
<td>36.9 (9.4)</td>
</tr>
<tr>
<td>Males, n (%)</td>
<td>34 (76)</td>
<td>77 (75)</td>
<td>34 (85)</td>
</tr>
<tr>
<td>Symptom duration, yrs, mean (SD)</td>
<td>15.9 (10.5)</td>
<td>16.9 (9.9)</td>
<td>14.8 (8.4)</td>
</tr>
<tr>
<td>BASDAI, mean (SD)</td>
<td>5.3 (2.5)</td>
<td>4.7 (2.4)</td>
<td>5.5 (2.2)</td>
</tr>
<tr>
<td>CRP, mg/l, mean (SD)</td>
<td>13.5 (15.0)</td>
<td>13.0 (17.0)</td>
<td>11.5 (13.7)</td>
</tr>
<tr>
<td>Taking only NSAID, n (%)</td>
<td>17 (38)</td>
<td>39 (38)</td>
<td>17 (43)</td>
</tr>
<tr>
<td>Taking TNF-α inhibitor, n (%)</td>
<td>23 (51)</td>
<td>56 (55)</td>
<td>21 (53)</td>
</tr>
<tr>
<td>SSS fat score, 0–40*, mean (SD) (range)</td>
<td>3.38 (4.07) (0–17)</td>
<td>3.84 (6.54) (0–31.5)</td>
<td>4.61 (6.30) (0–26)</td>
</tr>
<tr>
<td>Change</td>
<td>0.21 (1.43) (–4.25 – +5.75)</td>
<td>0.33 (2.03) (–6 – +11)</td>
<td>0.73 (1.80) (–4 – +6)</td>
</tr>
<tr>
<td>SSS erosion, 0–40*, mean (SD) (range)</td>
<td>3.09 (3.37) (0–15)</td>
<td>3.03 (4.01) (0–21)</td>
<td>4.63 (2.77) (0–18)</td>
</tr>
<tr>
<td>Change</td>
<td>–1.32 (2.04) (–8.25 – +4.25)</td>
<td>–0.94 (2.56) (–12 – +6)</td>
<td>–0.63 (1.67) (–9.33 – +6)</td>
</tr>
<tr>
<td>SSS backfill, 0–20*, mean (SD) (range)</td>
<td>3.68 (4.34) (0–17.8)</td>
<td>2.76 (3.68) (0–19)</td>
<td>4.93 (3.09) (0–18)</td>
</tr>
<tr>
<td>Change</td>
<td>0.5 (2.55) (–7.25 – +10.75)</td>
<td>0.06 (2.45) (–14.5 – +8)</td>
<td>–0.03 (1.61) (–8.33 – +7.33)</td>
</tr>
<tr>
<td>SSS ankylosis, 0–20*, mean (SD) (range)</td>
<td>5.27 (7.96) (0–20)</td>
<td>6.16 (8.02) (0–20)</td>
<td>2.36 (3.86) (0–14.7)</td>
</tr>
<tr>
<td>Change</td>
<td>0.15 (1.13) (–2.25 – +6.5)</td>
<td>0.36 (1.73) (–5 – +9)</td>
<td>0.58 (1.39) (–1.33 – +7.67)</td>
</tr>
</tbody>
</table>

* Maximum scoring range for SSS score. Baseline SSS score > 0 denotes at least 1 SIJ quadrant in the 5 coronal slices assessed in the SSS method with the structural lesion. Change SSS score > 0 denotes at least 1 SIJ quadrant of the 5 SIJ slices with a new or resolved structural lesion. SPARCC: the Spondyloarthritis Research Consortium of Canada; MRI: magnetic resonance imaging; SSS: Sacroiliac joint Structural Score; SpA: spondyloarthritis; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; CRP: C-reactive protein; NSAID: nonsteroidal antiinflammatory drug; TNF-α: tumor necrosis factor-α; SIJ: sacroiliac joint.
Several scoring methods have been developed for quantification of fat metaplasia, and the reliability for assessment of status scores was reported as very good although none as yet have been validated to show that change scores can be reliably detected. A report assessed fat lesions dichotomously (present/absent) according to the extent of subcortical bone affected (0 = no erosion, 1 = < 25%, 2 = 25–50%, 3 = > 50%). Reliability for status score was good, but data for change scores was not reported. In a second method, erosion was graded in both the cartilaginous and ligamentous compartments, and severity graded according to the extent of subcortical bone affected (0 = no erosion, 1 = < 25%, 2 = 25–50%, 3 = > 50%). Reliability for status score was good, but data for change score was not reported. A limitation of scoring based on the number of erosions per SIJ is that it is not unusual to observe erosion affecting the entire vertical height of the iliac or sacral bone on a coronal scan so that one cannot consider these as discrete lesions. In addition, this scheme appears to preclude the possibility that there may be a substantial reduction in the size of an erosion without affecting the number of erosions.

The scoring methodology we have developed suggests that reliable detection of change in structural lesions in the SIJ may be possible even after 1 year. Ankylosis is defined by the presence of bright marrow signal traversing the joint space between the iliac and sacral bones and is therefore easier to discern reliably on T1-weighted MRI than erosion, which requires loss of dark signal, signifying breach of cortical bone. Nevertheless, reliability for detecting change in erosion and ankylosis was comparable, suggesting that application of validated and standardized lesion definitions and calibration based on DICOM images illustrating the spectrum of abnormalities and changes over time can be an effective mode of knowledge transfer. Backfill presents a grading scheme based on the number of erosions per SIJ (1 = 1–2, 2 = 3–5, 3 = > 5 erosions per SIJ).
more heterogeneous MRI appearance with increased signal on the T1-weighted scan that is clearly demarcated from adjacent normal marrow by irregular dark signal reflecting sclerosis at the border of the eroded bone. This is more difficult to define in a standardized manner and is the likely reason why more extensive calibration is required to achieve more reliable detection of this lesion. We have shown that resolution of inflammation and reduction in erosion is significantly associated with the development of backfill, although further validation of this lesion using computed tomography would be helpful.

We have developed and validated a scoring methodology for the assessment of structural lesions on MRI in the SIJ of patients with SpA that is based on the same principles as the widely used SPARCC MRI SIJ inflammation score. These include selection of MRI slices according to well-defined anatomical principles, the scoring of lesions on a dichotomous basis as being present/absent on consecutive, oblique coronal slices through the cartilaginous portion of the joint, and calibration of readers using standardized definitions and DICOM-based reference cases. Further validation should now be undertaken aimed at assessment of responsiveness and the pathophysiological and prognostic significance of these lesions observed on MRI.

REFERENCES


