

# Feasibility and Reliability of the Spondyloarthritis Research Consortium of Canada Sacroiliac Joint Structural Score in Children

Pamela F. Weiss, Walter P. Maksymowych, Robert G. Lambert, Jacob L. Jaremko, David M. Biko, Joel Paschke, Timothy G. Brandon, Rui Xiao, and Nancy A. Chauvin

**ABSTRACT. Objective.** There is a critical need for measures to evaluate structural progression in the pediatric sacroiliac joint (SIJ). We aimed to evaluate the construct validity and reliability of the Spondyloarthritis Research Consortium of Canada SIJ Structural Score (SSS) in children with suspected or confirmed juvenile spondyloarthritis.

**Methods.** The SSS assesses structural lesions of the SIJ on magnetic resonance imaging (MRI) through the cartilaginous part of the joint. We conducted 3 sequential reading exercises with 6 readers (1 adult and 3 pediatric radiologists, 1 adult and 1 pediatric rheumatologist). Each exercise was preceded by a calibration module. Interobserver reliability was assessed using intraclass correlation coefficients (ICC). Prespecified acceptable reliability thresholds were  $ICC > 0.5$  for erosion, backfill, and sclerosis, and  $ICC > 0.7$  for ankylosis and fat metaplasia.

**Results.** The SSS had face validity and was feasible to score in pediatric cases for all 3 reading exercises. Of the cases used in the 3 exercises, 58% were male and the median age was 14 years (range 6.8–18.7 yrs). After calibration, median ICC across all readers for each SSS component were the following: erosion 0.67 (interquartile range 0.54–0.80), backfill 0.33 (0.19–0.52), fat metaplasia 0.74 (0.62–0.85), sclerosis 0.63 (0.48–0.77), and ankylosis 0.44 (0.28–0.62). Prespecified reliability thresholds were achieved in the third exercise for erosion, sclerosis, and fat metaplasia but not for backfill or ankylosis.

**Conclusion.** The SSS was feasible to score and had acceptable reliability for pediatric SIJ MRI evaluation. The ICC improved with additional calibration and reading exercises, even for readers with limited experience. (First Release June 15 2018; J Rheumatol 2018;45:1411–17; doi:10.3899/jrheum.171329)

## Keyword Indexing Terms:

SACROILIAC JOINT

SPONDYLOARTHRITIS

CALIBRATION

MAGNETIC RESONANCE IMAGING

DISEASE PROGRESSION

From the Departments of Pediatrics and Radiology, Division of Rheumatology, and Center for Pediatric Clinical Effectiveness (CPCE) at the Children's Hospital of Philadelphia; Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA; Department of Medicine at the University of Alberta; Department of Radiology and Diagnostic Imaging, University of Alberta; Canadian Research and Education (CaRE) Arthritis Organization, Edmonton, Alberta, Canada.

Dr. Weiss' work was supported by the Rheumatology Research Foundation.

P.F. Weiss, MD, MSCE, Department of Pediatrics, Division of Rheumatology at the Children's Hospital of Philadelphia and Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania; W.P. Maksymowych, MD, FRCP, Department of Medicine at the University of Alberta; R.G. Lambert, MB BCh, FRCP, FRCP, Department of Radiology and Diagnostic Imaging at the University of Alberta; J.L. Jaremko, MD, PhD, Department of Radiology and Diagnostic Imaging at the University of Alberta; D.M. Biko, MD, Department of Radiology at the Children's Hospital of Philadelphia; J. Paschke, BSc, CaRE Arthritis; T.G. Brandon, MPH, Department of Pediatrics, Division of Rheumatology at the Children's Hospital of Philadelphia; R. Xiao, PhD, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania; Department of Radiology and Diagnostic Imaging, University of Alberta; N.A. Chauvin, MD, Department of Radiology at the Children's Hospital of Philadelphia.

Address correspondence to Dr. P.F. Weiss, The Children's Hospital of Philadelphia, Roberts Center for Pediatric Research, 2716 South St., Room 11121, Philadelphia, Pennsylvania 19146, USA.  
E-mail: weisspa@email.chop.edu

Accepted for publication February 28, 2018.

Children with spondyloarthritis (SpA) are distinct from children with other types of juvenile arthritis in their propensity to have HLA-B27 positivity, enthesitis (inflammation around the tendon and ligament attachments), associated bowel inflammation, and axial arthritis (sacroiliac or spinal). One-third to one-half of children with juvenile SpA develop sacroiliitis within several years of diagnosis<sup>1,2,3,4</sup>. Efficacy and effectiveness studies are critically needed for this understudied pediatric population, but we lack the biomarkers and tools necessary to conduct them.

It is unclear what factors are associated with the progression of axial disease in children. Factors that have been reported in adults include smoking<sup>5</sup>, elevated inflammatory markers<sup>6</sup>, and structural damage at baseline<sup>6</sup>. In

adults, studies have suggested that fat metaplasia and backfill are 2 imaging biomarkers associated with later development of sacroiliac joint (SIJ) ankylosis<sup>7</sup>. There is also preliminary evidence in a mouse model indicating stress and strain may promote inflammation and disease progression<sup>8</sup>. Validated measures to assess progression of sacroiliitis include the New York (NY) and modified NY criteria<sup>9</sup> and the Spondyloarthritis Research Consortium of Canada (SPARCC) SIJ Structural Score (SSS)<sup>10</sup>. The NY and modified NY criteria rely on structural changes on conventional radiographs. The SPARCC SSS assesses a spectrum of structural lesions of the SIJ including fat metaplasia, erosion, backfill, and ankylosis on magnetic resonance imaging (MRI). These components are scored 0–20 (backfill and ankylosis) or 0–40 (fat metaplasia and erosion), with higher numbers indicative of more progression. The SPARCC SSS has been used as a clinical trial outcome<sup>11,12</sup>. In a 12-week placebo-controlled trial of etanercept for adults with nonradiographic axial SpA, significant changes in the SPARCC structural component score for erosion were detectable as early as 12 weeks<sup>13</sup>.

Although the SPARCC SSS has been validated in adults, it is critical to also specifically assess its performance in the pediatric population. The pediatric SIJ undergoes normal development-related changes over time, including those that might be mistaken for pathology, which could affect the reliability and validity of the tool in this population. In addition, structural damage lesions seen primarily with longstanding disease might be more difficult to detect reliably in children who, on average, have much shorter disease duration at the time of imaging. Lastly, lesions such as sclerosis, which were not ultimately included in the SPARCC adult score, are worth evaluating in the pediatric population, because degenerative changes are typically uniformly absent.

We evaluated the reliability and construct validity of the SPARCC SSS in children with suspected or confirmed juvenile SpA from 2 large tertiary care centers.

## MATERIALS AND METHODS

The protocol for our study was reviewed and approved by the Children's Hospital of Philadelphia's Committee for the Protection of Human Subjects (16-012641, 16-013477, and 17-013883). Waivers of consent and Health Insurance Portability and Accountability Act of 1996 authorization were granted for these studies because the procedures did not represent more than minimal risk to the subjects and did not adversely affect the rights and welfare of the subjects.

Cases were children (ages 5–18 yrs at the time of imaging) with suspected or confirmed SpA (European Spondyloarthritis Group SpA criteria, International League of Associations for Rheumatology juvenile arthritis criteria for enthesitis-related arthritis or psoriatic arthritis) who had dedicated pelvic MRI performed at the Children's Hospital of Philadelphia (Philadelphia, Pennsylvania, USA) or Stollery Children's Hospital (Edmonton, Alberta, Canada) between January 2011 and November 2016. All images were centrally collected, digitized, anonymized, randomized (time order), and scored as detailed below. Imaging studies that did not include semicoronal T1-weighted (T1W) and short-tau inversion recovery (STIR) views were excluded.

The SSS<sup>10</sup> assesses a spectrum of structural lesions of the SIJ on MRI

including erosion, backfill, fat metaplasia, and ankylosis on 5 consecutive slices through the cartilaginous part of the joint (Table 1). These components are scored 0–20 (backfill and ankylosis) or 0–40 (erosion and fat metaplasia; Figure 1). For the evaluation of pediatric cases we also included sclerosis, scored 0–40. Sclerosis was tested as a component of the SPARCC SSS adult scoring but was not included secondary to lack of specificity, responsiveness, and difficulty in differentiating from degenerative sclerosis. These issues were not felt to apply to the pediatric SIJ. In total we conducted 3 calibration modules and 3 reading exercises over the course of about 18 months.

All training and scoring were performed within a Web-based environment. Prior to exercise 1, all readers (1 adult and 3 pediatric radiologists, 1 adult and 1 pediatric rheumatologist) reviewed a pediatric training module that included a detailed description of each SSS component plus sclerosis (0–40), scoring methodology, and numerous examples based on PowerPoint slides. Two of the readers, the adult rheumatologist (WM) and adult radiologist (RL), were the SPARCC SSS developers. After reviewing the module, the 6 readers used an online viewing and scoring system to score 30 Digital Imaging and Communication in Medicine (DICOM)-based anonymized cases comprising semicoronal T1W sequences (reading exercise 1). The same structure was followed leading up to the second exercise: all readers reviewed studies within a training module comprising both T1W and STIR scans, then independently scored an additional 29 studies (reading exercise 2).

After reading exercise 2, all readers, except WPM and RGL (the 2 SPARCC SSS developers), participated in an interactive calibration module comprising 40 adult cases, each with studies of T1W scans from baseline and 2 years after initiation of tumor necrosis factor inhibitor therapy ([www.carearthritis.com/mriportal/sss/index/](http://www.carearthritis.com/mriportal/sss/index/)). Each study was scored in pairs blinded to timepoint for each SSS component. For the first 20 cases, readers received instantaneous feedback on concordance/discordance of scoring with expert readers (WPM and RGL) after scoring each individual semicoronal slice. For the second 20 cases, feedback was provided after scoring the entire case. Intraclass correlation coefficient (ICC) with expert reader scores was provided after the first 20 cases had been scored, then again after the next 10 cases, and finally after all 40 cases had been scored. *A priori* scores anticipated for this calibration were ICC of > 0.7 for ankylosis and fat metaplasia status scores, and > 0.5 for erosion and backfill status scores. After the interactive training module, we conducted a third reading exercise based on DICOM images comprising T1W and STIR scans from 30 pediatric studies, which were a mix of 27 studies from the previous 2 exercises and 3 new studies.

Interobserver reliability was assessed by calculating the ICC using 2-way random effects models<sup>14</sup>. Given our data, we chose to report ICC to provide comparable measures of interrater reliability for different subgroups of raters. Use of ICC in cases of 2 raters is justifiable because both the inter- and intrarater variations are estimable. The ICC are presented as agreement for all readers together, SPARCC developers (n = 2), pediatric radiologists (n = 3), and rheumatologists (n = 2). One investigator was both a SPARCC developer and a rheumatologist. We applied the prespecified ICC targets from the interactive calibration module to the interpretation of the results from the 3 reading exercises (> 0.5 for erosion and backfill, and > 0.7 for fat metaplasia and ankylosis status score; in addition, we anticipated an ICC of sclerosis > 0.5).

We evaluated the absolute change in the ICC between exercises 1 and 3 among all readers and within groups of readers categorized by professional background (SPARCC developers, pediatric radiologists, rheumatologists). We also assessed the interrater agreement between exercise 1 and 3 (stratified by the groups of readers) by testing whether the difference in scores on a case-by-case level decreased significantly in exercise 3 compared to exercise 1. Wilcoxon rank-sum test was used for 2 reader categories (rheumatologists and SPARCC developers) and mixed effects regression was used for the 3 reader categories (pediatric radiologists).

Construct validity between the mean SPARCC SSS developers' scores for each domain from reading exercise 3 and disease duration (peripheral or axial manifestations) was assessed with Spearman correlation. Discrimination was tested by comparing the mean SPARCC SSS developers' scores

Table 1. Scoring system used to evaluate 5 consecutive slices in the SPARCC SSS.

Feature	Definition	Score per Slice
Fat metaplasia	Increased signal on T1W sequences; homogeneous increased signal that extends $\geq 1$ cm from the joint surface	0–8
Erosion	Full thickness loss of cortical bone at its anticipated location and loss of normal bright appearance of adjacent bone marrow	0–8
Backfill	Complete loss of cortical bone at its anticipated location and increased T1 signal clearly demarcated from adjacent normal marrow by dark band with irregular contour on T1W sequence	0–4
Ankylosis	T1W bone marrow signal extending between the sacral and iliac bone marrow with full thickness loss of dark appearance of cortical bone	0–4
Sclerosis	Dark appearance of cortical bone on T1W sequence extending $> 5$ mm beyond the cortical margin	0–8

SPARCC: Spondyloarthritis Research Consortium of Canada; SSS: Sacroiliac joint Structural Score; T1W: T1-weighted.

for each component with the following: (1) patient-reported pain (axial or peripheral)  $\geq 4$  versus  $< 4$ ; (2) patient-reported global disease activity  $\geq 3$  versus  $< 3$ ; (3) physician-reported disease activity  $\geq 3$  versus  $< 3$ ; (4) elevated C-reactive protein ( $\geq 1$  mg/dl) or erythrocyte sedimentation rate ( $\geq 21$  mm/h); and (5) HLA-B27 status. Cutoff values for the discrimination analysis were determined based upon the distribution of scores for each variable. Clinical values were included in the assessment if they were  $\pm 60$  days from the date of the MRI. All analyses were run using Stata Statistical Software version 14.2 (StataCorp).

## RESULTS

**Subjects.** Table 2 shows the demographic and clinical characteristics of the cases included in the reading exercises. In total there were 62 unique cases used for the 3 reading exercises. Thirty-six (58%) were male and median age at the time of imaging was 14.3 years (IQR 11.9–16.4, range 6.8–18.7 yrs). At the time of imaging, most children and adolescents had limited peripheral disease activity (median active joint and tender entheses counts of 0) but moderate self-reported pain (median 3.8) and self- and physician-reported disease activity (2.8 and 3, respectively).

**Feasibility and interobserver reliability.** The SSS had face validity and was feasible to score. The time required to assess a case for all structural lesions for a single case ranged from 5 to 20 min and the online scoring platform was considered easy to use by the readers.

The ICC for the SSS components from all 3 reading exercises are shown in Table 3. In the first exercise, the ICC for the group of all readers ranged between 0.31 and 0.42. The prespecified ICC targets of  $> 0.5$  for erosion, backfill, and sclerosis status scores and  $> 0.7$  for fat metaplasia and ankylosis status scores were not met. The only group to exceed all prespecified ICC was from the SPARCC developers. Rheumatologists were able to achieve agreement in line with *a priori* thresholds for fat metaplasia and sclerosis status scores. In exercise 1, the number of studies with scores  $> 0$  for all readers were erosion 131 (73%), sclerosis 100

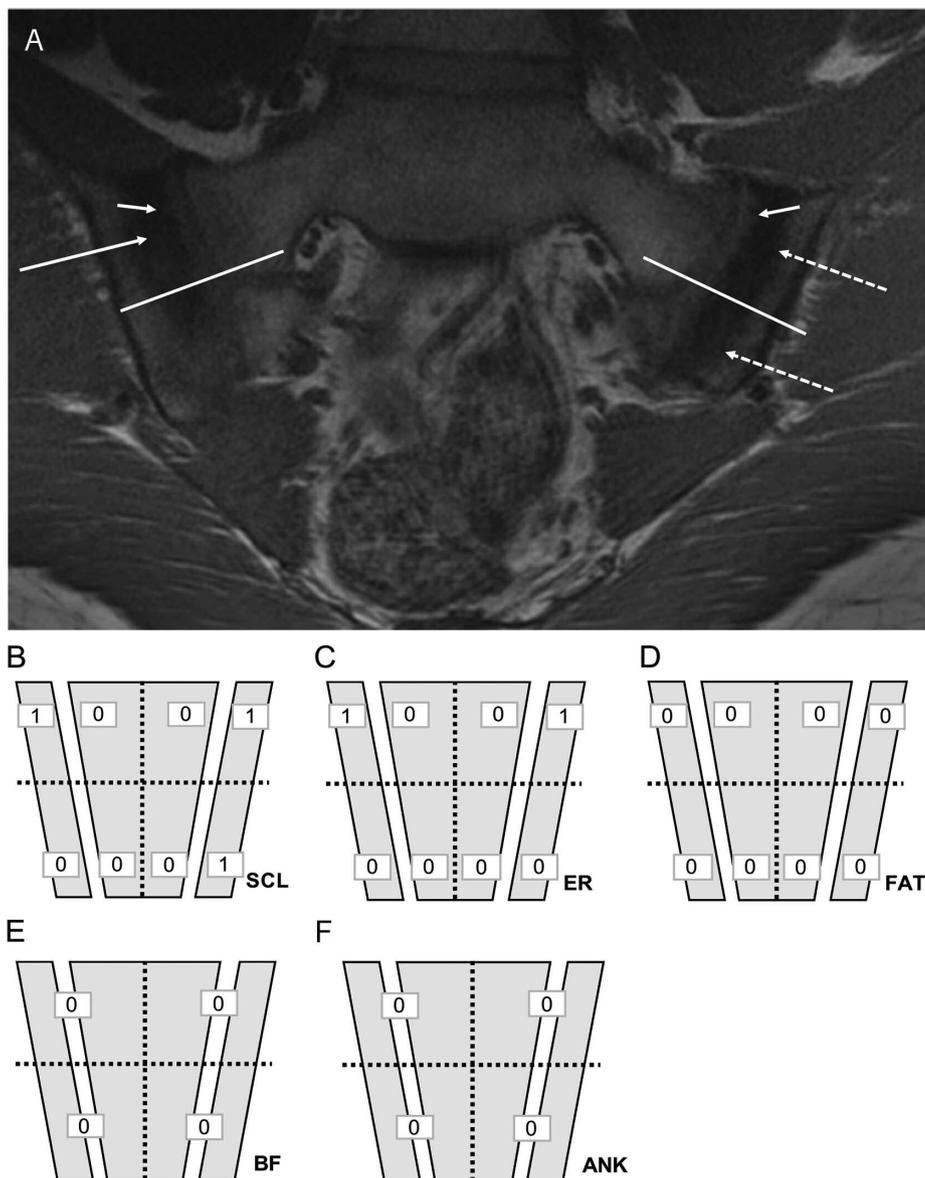
(56%), backfill 56 (31%), fat metaplasia 31 (17%), and ankylosis 21 (12%).

In the second reading exercise, the erosion ICC among all readers surpassed the prespecified ICC for erosion (ICC 0.54, IQR 0.34–0.72); ICC for backfill and sclerosis remained lower than anticipated. Ninety-one (52%), 18 (10%), and 49 (28%) studies from all readers had a score  $> 0$  for erosion, backfill, and sclerosis in exercise 2, respectively. Reliability for fat metaplasia and ankylosis were not calculated owing to low frequency of lesions [8 (5%) and 3 (2%), respectively].

In between the second and third reading exercises, the 4 readers who were not SPARCC SSS developers completed an interactive calibration module comprising 40 adult cases. After the initial 20 cases that provided real-time feedback regarding concordance or discordance of scores with expert readers, all 4 readers achieved an ICC status score  $> 0.7$  for backfill, fat metaplasia, and ankylosis. All ICC for erosion status score were  $\geq 0.86$ , fat metaplasia status score  $\geq 0.97$ , and ankylosis status score  $\geq 0.86$ . All 4 readers achieved an ICC  $\geq 0.7$  for backfill.

In the third reading exercise, the ICC among all readers surpassed the prespecified targets for erosion (0.67, IQR 0.54–0.80), fat metaplasia (0.74, IQR 0.62–0.85), and sclerosis (0.63, IQR 0.48–0.77) status scores. The ICC for backfill was  $< 0.40$  for all readers as a group, but slightly higher at 0.43 for the SPARCC developers and 0.58 for the rheumatologists. In exercise 3, the number of studies that had component status scores  $> 0$  for all readers were the following: erosion 124 (69%), sclerosis 69 (38%), backfill 47 (26%), fat metaplasia 26 (24%), and ankylosis 16 (9%).

The ICC for all components of the SSS across the reading exercises are shown in Figure 2. The ICC among all readers improved with each reading exercise with the exception of backfill, which increased between exercises 1 and 2 but decreased between exercises 2 and 3. In testing whether the



**Figure 1.** Sample slice and scoring methods for the SPARCC SSS. A. Coronal oblique T1-weighted image of the SIJ shows each joint divided into 4 quadrants. The long, white non-arrow lines divide the joints into upper and lower quadrants. There is sclerosis along the upper quadrant of the right iliac bone (long arrow) as well as within both the upper and lower quadrants of the left iliac bone (dashed arrows). In addition, there are erosions within the upper quadrants of both iliac bones (short arrows). B. Sclerosis (SCL). C. Erosions (ER). D. Fat metaplasia (FAT). E. Backfill (BF). F. Ankylosis (ANK). Each box represents a location to be scored. Based on panel A, sclerosis is present within the right upper iliac quadrant and the left upper and lower iliac quadrants, for a total score of 3. Erosions are present within the right and left upper iliac quadrants, for a total score of 2. There is no fat metaplasia, backfill, or ankylosis on this image slice. Total cumulative score for this image slice is 5. SPARCC: Spondyloarthritis Research Consortium of Canada; SIJ: sacroiliac joints; SSS: SIJ Structural Score.

difference in scores between readers within professional background categories changed from reading exercise 1 to exercise 3, we found that deltas in status scores were trending in the direction of improvement suggested by the improvement in ICC, but there was not enough evidence to conclude that the populations of delta values in exercise 1 were different from exercise 3 for any of the SSS compo-

nents except for erosion measured by rheumatologists ( $p = 0.01$ ).

Correlations of symptom (axial or peripheral) duration with each component of the SSS developers' mean scores for exercise 3 were low (fat metaplasia  $r = -0.15$ ,  $p = 0.43$ ; erosion  $r = 0.09$ ,  $p = 0.65$ ; backfill  $r = 0.23$ ,  $p = 0.23$ ; sclerosis  $r = 0.28$ ,  $p = 0.14$ ; ankylosis  $r = 0.20$ ,  $p = 0.28$ ). Fat

Table 2. Case characteristics (62 unique cases in 3 reading exercises). Values are n (%) or median (IQR) unless otherwise specified.

Characteristics	N	Values
Age	62	14.3 (11.9–16.4)
Sex, male	62	36 (58.1)
HLA-B27+	55	31 (56.4)
Race	56	
White		46 (82.1)
African American		6 (10.7)
Other		4 (7.1)
Active joint count	52	0 (0–1)
Tender enthesis count	52	0 (0–5)
Patient-reported pain*	46	4 (1–6)
Patient-reported disease activity	45	3 (1–5)
Physician global VAS	46	3 (2–4)
cJADAS10 <sup>#</sup>	43	6.5 (2–9.7)
CRP, elevated, ≥ 1 mg/dl	37	19 (51.4)
ESR, elevated, ≥ 21 mm/h	41	17 (41.5)
Disease duration, days	33	154 (23–717)

\* General pain recorded on VAS. <sup>#</sup> A validated composite disease measure specifically developed for juvenile idiopathic arthritis and ranges from 0 (inactive disease) to 30, composed of the active joint count (maximum 10 joints), and physician global and parent global disease activity evaluation. VAS: visual analog scale; cJADAS10: Clinical Juvenile Arthritis Disease Activity Score; CRP: C-reactive protein; ESR: erythrocyte sedimentation rate; IQR: interquartile range.

metaplasia, erosion, and ankylosis did not discriminate between children with patient-reported pain (peripheral or axial) ≥ 4 versus < 4 (all  $p > 0.05$ ), patient-reported global disease activity ≥ 3 versus < 3 (all  $p > 0.05$ ), physician-reported disease activity ≥ 3 versus < 3 (all  $p > 0.05$ ),

elevated C-reactive protein or sedimentation rate (all  $p > 0.05$ ), or HLA-B27 status ( $p > 0.05$ ). Sclerosis scores were significantly different between those with a patient-reported pain score of ≥ 3 versus < 3 (median scores of 0 and 3, respectively;  $p = 0.03$ ); sclerosis scores did not significantly discriminate between the other clinical measures. Backfill scores did not significantly discriminate by HLA-B27 status (median scores of 0 for B27-negative and 0.5 for B27-positive,  $p = 0.02$ ); backfill scores did not significantly discriminate between the other clinical measures (data not shown).

## DISCUSSION

To our knowledge, this is the first study to assess the performance of the SPARCC SSS in a pediatric population. The SPARCC SSS method has been validated in adults with SpA<sup>10</sup> and has been used as an outcome in clinical trials<sup>11,13</sup>. We adapted the original instrument for use in the pediatric population by adding sclerosis to the assessment. The exclusion of sclerosis from adult scoring is primarily based on lack of specificity, lack of responsiveness, and difficulty in scoring sclerosis in adults. However, none of these issues may apply in the rapidly evolving skeleton of children, and pediatric SIJ are also not subject to the confounding factor of degenerative sclerosis. The SSS was feasible and easy to score for pediatric SIJ MRI evaluation. After calibration we attained the prespecified acceptable reliability for erosion, sclerosis, and fat metaplasia but not for backfill or ankylosis. Importantly, the ICC for each component of the SSS improved with additional calibration exercises based on DICOM, even for readers with limited experience. The components of the SPARCC SSS did not have construct

Table 3. ICC for all calibration exercises. Values are ICC (IQR).

Variables	All Readers, n = 6	Pediatric Radiologists, n = 3	SPARCC Developers, n = 2	Rheumatologists, n = 2
Exercise 1				
Erosion	0.37 (0.22–0.55)	0.39 (0.15–0.61)	0.72 (0.48–0.86)	0.16 (–0.17 to 0.47)
Backfill	0.39 (0.25–0.58)	0.36 (0.14–0.58)	0.82 (0.56–0.92)	0.45 (0.12–0.69)
Fat metaplasia	0.40 (0.25–0.58)	0.46 (0.25–0.67)	0.89 (0.77–0.95)	0.82 (0.65–0.91)
Sclerosis	0.42 (0.25–0.60)	0.42 (0.17–0.65)	0.61 (0.08–0.83)	0.72 (0.49–0.86)
Ankylosis	0.31 (0.17–0.49)	0.46 (0.24–0.66)	0.72 (0.49–0.85)	0.19 (–0.16 to 0.51)
Exercise 2				
Erosion	0.54 (0.34–0.72)	0.51 (0.16–0.74)	0.96 (0.89–0.98)	0.61 (0.32–0.79)
Backfill	0.47 (0.31–0.64)	0.12 (–0.08 to 0.38)	0.90 (0.80–0.95)	0.98 (0.95–0.99)
Fat metaplasia	–	–	–	–
Sclerosis	0.47 (0.30–0.65)	0.35 (0.13–0.58)	0.88 (0.75–0.94)	0.81 (0.63–0.90)
Ankylosis	–	–	–	–
Exercise 3				
Erosion	0.67 (0.54–0.80)	0.67 (0.49–0.81)	0.76 (0.21–0.91)	0.66 (0.40–0.82)
Backfill	0.33 (0.19–0.52)	0.28 (0.07–0.51)	0.43 (0.09–0.68)	0.58 (0.27–0.77)
Fat metaplasia	0.74 (0.62–0.85)	0.76 (0.59–0.87)	0.74 (0.53–0.87)	0.75 (0.53–0.88)
Sclerosis	0.63 (0.48–0.77)	0.57 (0.36–0.75)	0.69 (0.34–0.85)	0.59 (0.30–0.78)
Ankylosis	0.44 (0.28–0.62)	0.20 (–0.02–0.45)	0.84 (0.70–0.92)	0.86 (0.73–0.93)

Score ranges for each component of the SSS: erosion (0–40), backfill (0–20), fat metaplasia (0–40), sclerosis (0–40), ankylosis (0–20). Exercise 2 ICC scores not calculable for fat metaplasia and ankylosis owing to an insufficient number of abnormalities present within the cohort of selected subjects for that exercise. ICC: intraclass correlation coefficients; IQR: interquartile range; SPARCC: Spondyloarthritis Research Consortium of Canada; SSS: Sacroiliac joint Structural Score.

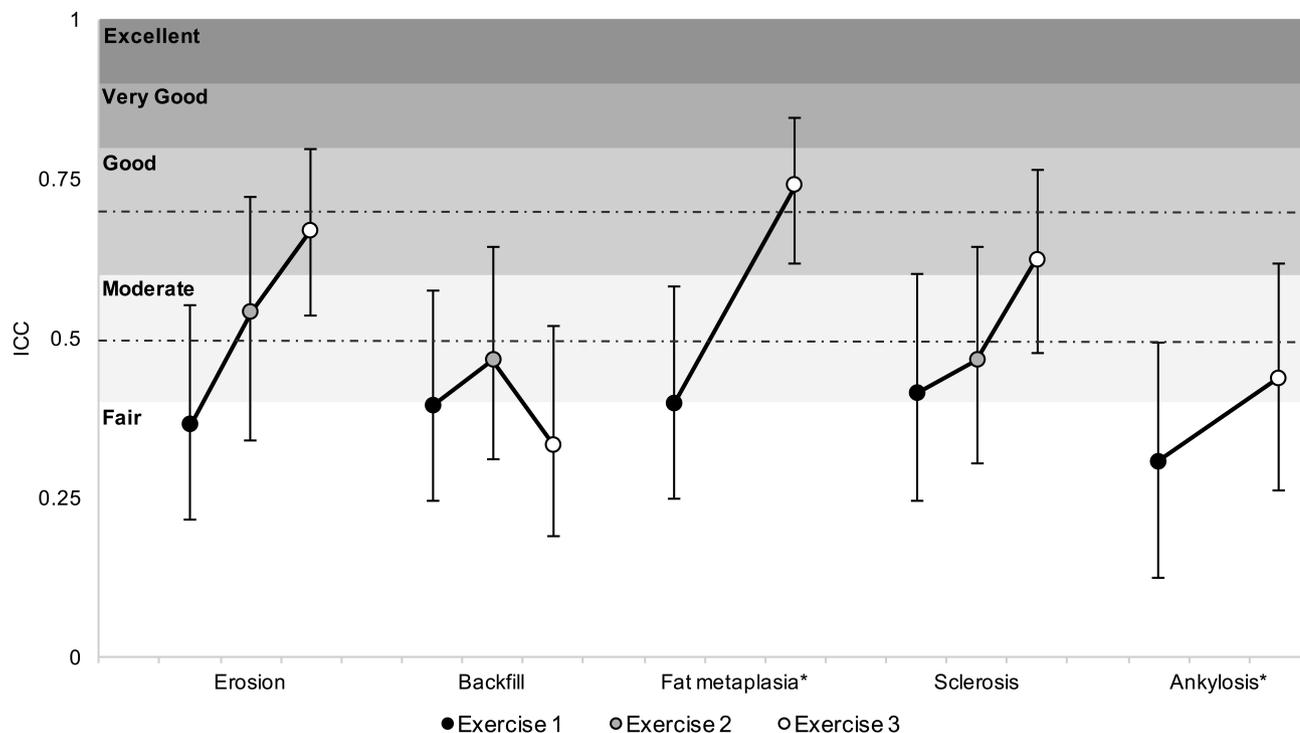


Figure 2. The ICC for each scored abnormality for all raters across 3 calibration exercises. \* Exercise 2 ICC scores not calculable for ankylosis and fat metaplasia because of an insufficient number of abnormalities present within the cohort of selected subjects for that exercise. Vertical bars depict the 95% CI for each ICC. Shaded regions reflect literature-defined interpretations of ICC and dashed lines represent the preexercise target acceptable agreement thresholds for erosion, backfill, and sclerosis (0.5) and ankylosis and fat metaplasia (0.7) based on the inherent difficulty of interpreting the given abnormalities. ICC: intraclass correlation coefficients.

validity with disease duration and only backfill discriminated between HLA-B27 status.

A few key points from our study deserve additional discussion. First, the detail that the ICC improved among each category of reader background as well as within all 6 readers with each additional calibration and reading exercise, even for readers with limited experience, is critical. In pediatrics, there is a relative paucity of dedicated musculoskeletal radiologists. Further, not all pediatric musculoskeletal radiologists have extensive experience interpreting imaging of the SIJ and few have any experience with scoring structural lesions of these joints. Our results indicate that readers of various levels of experience can be trained to read and score these studies reliably through the calibration and reading exercises publicly available on the CaRE Arthritis platform. This is heartening because it suggests increased feasibility of conducting effectiveness and efficacy studies of axial disease across centers and countries.

Second, the ICC achieved by each of the 4 readers who participated in the interactive calibration exercise was substantially better than the ICC among all 6 readers for the SSS components in the third reading exercise. Part of this difference in reliability may be explained by the immediate feedback provided regarding concordance/discordance of scoring with the expert readers, allowing for more accelerated

learning. The interactive calibration exercise, however, was based upon studies performed in adults with SpA, not children. It may be that identification of some of the lesions may be more challenging in pediatric studies. For example, in adult SPARCC scoring, the definition of erosion includes full-thickness cortical loss and loss of underlying bright T1 marrow signal. In younger children, the bony cortex may not be fully ossified and may not appear as a dark line, while underlying marrow may still be “red” marrow with relatively low T1 signal, leading to confusion between physiologic appearances and erosion. Another explanation for the difference in the magnitude of the ICC between the interactive calibration and reading exercise 3 may be a result of the frequency or conspicuity of the lesions. Ankylosis, backfill, and fat metaplasia are more common in adult cases with longstanding disease than pediatric cases with relatively short disease duration. The relative rarity of these lesions in children can increase the numeric effect of discrepant readings in just a few cases, resulting in lower ICC values. Having a higher proportion of lesions close to threshold for scoring similarly may tend to reduce ICC.

Third, the SSS components did not have construct validity when compared to disease duration. This is not surprising because most children in our study, typical of most children with SpA seen in routine practice, had a relatively short

disease duration. Discrimination of patient- and physician-reported clinical outcomes was low. It has been previously reported that the clinical findings have low sensitivity and positive predictive value for inflammation or chronic changes consistent with sacroiliitis in children<sup>15</sup>, so the limited discrimination is also not surprising. What the SPARCC sacroiliac SSS offers is a more systematic and objective evaluation that lends itself to repeatability and consistency in measurements that is missing in current clinical and imaging evaluations. In addition, sclerosis can be observed with the same degree of consistency as the other domains of structural damage, and because its utility is not yet known, it is recommended that sclerosis be included in the MRI assessment of structural damage in future studies.

There are several limitations to our study. First, the number of available pediatric studies was limited, largely because of missing sequences important for use in SPARCC scoring. This forced our team to reevaluate cases used in the first and second reading exercises again during the third exercise. We do not believe using the same studies more than once affected the results because the cases were not discussed among the readers, and each reading exercise was separated by about 6 months. The limited number of studies also means we could not assess for minimal detectable difference or change scores. This should be addressed in future work. Second, the studies used in these reading and calibration exercises were primarily from 2 North American hospitals. However, both these institutions are large tertiary care and referral centers, so the studies included herein are likely to be widely representative of cases that also appear in other geographic areas. Third, some patient-reported outcomes and physician disease activity assessments were missing, which is inevitable in a retrospective study. While this limited the number of cases available to use for assessment of construct validity and discriminative ability, this does not affect the feasibility or reliability assessment. These relatively minor limitations are to be expected in the first systematic assessment of the feasibility and reliability of an objective scoring system for pediatric axial disease.

We have demonstrated the feasibility and reliability of the SPARCC SSS methodology in children with established or suspected SpA. This scoring system is based upon dichotomous scoring of lesions on consecutive coronal oblique slices through the cartilaginous part of the joint. In addition to demonstrating feasibility we have shown that inexperienced readers can be calibrated using standardized definitions, DICOM reference cases, and interactive calibration modules. Further work is needed to assess responsiveness and the prognostic significance of the SPARCC SSS MRI lesions in children.

## REFERENCES

1. Bollow M, Braun J, Biedermann T, Mutze S, Paris S, Schauer-Petrowskaja C, et al. Use of contrast-enhanced MR

- imaging to detect sacroiliitis in children. *Skeletal Radiol* 1998;27:606-16.
2. Stoll ML, Bhore R, Dempsey-Robertson M, Punaro M. Spondyloarthritis in a pediatric population: risk factors for sacroiliitis. *J Rheumatol* 2010;37:2402-8.
3. Flato B, Hoffmann-Vold AM, Reiff A, Forre O, Lien G, Vinje O. Long-term outcome and prognostic factors in enthesitis-related arthritis: a case-control study. *Arthritis Rheum* 2006;54:3573-82.
4. Pagnini I, Savelli S, Matucci-Cerinic M, Fonda C, Cimaz R, Simonini G. Early predictors of juvenile sacroiliitis in enthesitis-related arthritis. *J Rheumatol* 2010;37:2395-401.
5. Poddubnyy D, Haibel H, Listing J, Marker-Hermann E, Zeidler H, Braun J, et al. Cigarette smoking has a dose-dependent impact on progression of structural damage in the spine in patients with axial spondyloarthritis: results from the GERMAN SPondyloarthritis Inception Cohort (GESPIC). *Ann Rheum Dis* 2013;72:1430-2.
6. Poddubnyy D, Rudwaleit M, Haibel H, Listing J, Marker-Hermann E, Zeidler H, et al. Rates and predictors of radiographic sacroiliitis progression over 2 years in patients with axial spondyloarthritis. *Ann Rheum Dis* 2011;70:1369-74.
7. Maksymowych WP, Wichuk S, Chiowchanwisawakit P, Lambert RG, Pedersen SJ. Fat metaplasia and backfill are key intermediaries in the development of sacroiliac joint ankylosis in patients with ankylosing spondylitis. *Arthritis Rheum* 2014;66:2958-67.
8. Jacques P, Lambrecht S, Verheugen E, Pauwels E, Kollias G, Armaka M, et al. Proof of concept: enthesitis and new bone formation in spondyloarthritis are driven by mechanical strain and stromal cells. *Ann Rheum Dis* 2014;73:437-45.
9. Goie TH, Steven MM, van der Linden SM, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis: a comparison of the Rome, New York and modified New York criteria in patients with a positive clinical history screening test for ankylosing spondylitis. *Br J Rheumatol* 1985;24:242-9.
10. Maksymowych WP, Wichuk S, Chiowchanwisawakit P, Lambert RG, Pedersen SJ. Development and preliminary validation of the spondyloarthritis research consortium of Canada magnetic resonance imaging sacroiliac joint structural score. *J Rheumatol* 2015;42:79-86.
11. Pedersen SJ, Poddubnyy D, Sorensen IJ, Loft AG, Hindrup JS, Thamsborg G, et al. Course of magnetic resonance imaging-detected inflammation and structural lesions in the sacroiliac joints of patients in the randomized, double-blind, placebo-controlled Danish multicenter study of adalimumab in spondyloarthritis, as assessed by the Berlin and Spondyloarthritis Research Consortium of Canada methods. *Arthritis Rheum* 2016;68:418-29.
12. Maksymowych WP, Wichuk S, Dougados M, Jones HE, Pedersen R, Szumski A, et al. Modification of structural lesions on MRI of the sacroiliac joints by etanercept in the EMBARK trial: a 12-week randomised placebo-controlled trial in patients with non-radiographic axial spondyloarthritis. *Ann Rheum Dis* 2018;77:78-84.
13. Lambert RG, Salonen D, Rahman P, Inman RD, Wong RL, Einstein SG, et al. Adalimumab significantly reduces both spinal and sacroiliac joint inflammation in patients with ankylosing spondylitis: a multicenter, randomized, double-blind, placebo-controlled study. *Arthritis Rheum* 2007;56:4005-14.
14. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 1979;86:420-8.
15. Weiss PF, Xiao R, Biko DM, Chauvin NA. Assessment of sacroiliitis at diagnosis of juvenile spondyloarthritis by radiography, magnetic resonance imaging, and clinical examination. *Arthritis Care Res* 2016;68:187-94.