

Metric Properties of the SPARCC Score of the Sacroiliac Joints — Data from Baseline, 3-month, and 12-month Followup in the SPACE Cohort

Rosaline van den Berg, Manouk de Hooge, Pauline A.C. Bakker, Floris van Gaalen, Victoria Navarro-Compán, Karen Minde Fagerli, Robert Landewé, Maikel van Oosterhout, Roberta Ramonda, Monique Reijnierse, and Désirée van der Heijde

ABSTRACT. Objective. To evaluate metric properties of the SpondyloArthritis Research Consortium of Canada (SPARCC) score of the sacroiliac (SI) joints.

Methods. Patients with back pain (≥ 3 months, ≤ 2 years, onset < 45 years) were included in the SPACE cohort (SpondyloArthritis Caught Early). Patients with (possible) axial spondyloarthritis had followup visits after 3 and 12 months and were treated according to clinical practice. Magnetic resonance imaging (MRI) of the SI joints (MRI-SI) was scored in 2 independent campaigns (campaign 1: at baseline and 3 months; campaign 2: at baseline, 3 months, and 12 months) by 2 different blinded reader pairs, applying the Assessment of Spondyloarthritis International Society (ASAS) definition (MRI-SI+ vs MRI-SI–; discordant cases were adjudicated by a third reader) and SPARCC score (mean of 2 agreeing readers). Calculations were made for agreement between SPARCC score cutoff values and a consensus judgment of MRI-SI+ (ASAS definition) as external standard, change in SPARCC score, and smallest detectable changes (SDC) over 3 and 12 months.

Results. SPARCC score ≥ 2 showed best agreement with MRI-SI+ in both campaigns. Regarding observed changes in relation to SDC, SPARCC score changed in 70/151 patients; 26/70 patients changed $> \text{SDC}$ (3.4), of whom 20 patients received stable treatment over 3 months in campaign 1. Over 3 months, 20/68 patients showed changes in SPARCC score; 11/20 $> \text{SDC}$ (2.1), of whom 8 patients received stable treatment. Over 1 year, 23/74 patients changed their SPARCC score; 14/23 changed $> \text{SDC}$ (2.4), of whom 7 received stable treatment in campaign 2.

Conclusion. SPARCC score ≥ 2 can be used as surrogate for a consensus judgment of MRI-SI+ (ASAS definition) in clinical trials. The SDC ranged from 2.1–3.4 dependent on reader pair and were close to the proposed minimum important change of 2.5. (First Release May 1 2015; J Rheumatol 2015;42:1186–93; doi:10.3899/jrheum.140806)

Key Indexing Terms:

SPONDYLOARTHRITIS
INFLAMMATION

MAGNETIC RESONANCE IMAGING
SACROILIAC JOINTS

From the Department of Rheumatology, and the Department of Radiology, Leiden University Medical Center, Leiden, the Netherlands; Department of Rheumatology, University Hospital La Paz, Madrid, Spain; Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway; Amsterdam Rheumatology Center, Amsterdam; Department of Rheumatology, Groene Hart Ziekenhuis (GHZ) hospital, Gouda, the Netherlands; Rheumatology Unit, Department of Medicine, DIMED University of Padua, Padua, Italy.

Supported by the Dutch Rheumatism Association (Reumafonds).

R. van den Berg, PhD; M. de Hooge, MSc; P.A. Bakker, MD; F. van Gaalen, MD, PhD, Department of Rheumatology, Leiden University Medical Center; V. Navarro-Compán, MD, Department of Rheumatology, Leiden University Medical Center, and Department of Rheumatology, University Hospital La Paz; K.M. Fagerli, MD, PhD, Department of Rheumatology, Diakonhjemmet Hospital; R. Landewé, MD, PhD, Amsterdam Rheumatology Center; M. van Oosterhout, MD, PhD, Department of Rheumatology, GHZ hospital; R. Ramonda, MD, Rheumatology Unit, Department of Medicine, DIMED University of Padua; M. Reijnierse, MD, PhD, Department of Radiology, Leiden University Medical Center; D. van der Heijde, MD, PhD, Department of Rheumatology, Leiden University Medical Center.

Address correspondence to R. van den Berg, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, The Netherlands.

E-mail: r.van_den_berg@lumc.nl

Accepted for publication March 12, 2015.

A positive magnetic resonance imaging (MRI) of the sacroiliac joints according to the Assessment of Spondyloarthritis International Society (ASAS) definition (positive MRI)¹ is part of the ASAS axial spondyloarthritis (axSpA) criteria² and is increasingly used to test eligibility of patients with axSpA for clinical trials^{3,4,5}. Within clinical trials, MRI-SI is often repeated over short periods (e.g., 12 weeks) to test the efficacy of treatment (especially biological) regarding changes in inflammation. For this efficacy read, the SpondyloArthritis Research Consortium of Canada (SPARCC) score is frequently used because it measures inflammation on a continuous scale with good sensitivity to change^{6,7}. It is unknown what SPARCC score cutoff value is the equivalent of a positive MRI, a value that is needed to link the read for eligibility and the efficacy reading. For example, this information would be useful to define groups with MRI scored according to SPARCC scores as having or not having a “positive MRI,” to study differences in treatment response over time⁴.

Treatment with biologicals may dramatically influence inflammatory signs on MRI^{8,9,10,11}, but inflammation may also spontaneously change over time in patients without treatment and in patients taking stable nonbiological treatment^{12,13,14}. However, it is not clear how many SPARCC score units these spontaneous changes represent^{12,13,14}. Moreover, these spontaneous changes are likely to be different with variable lengths of followup. A minimally important change (MIC) of 2.5 SPARCC units is proposed based on the patient global assessment as external anchor¹⁵. It is known that interreader reliability of SPARCC scores for both a fixed timepoint and changes over time ranges from moderate to high (ICC 0.69–0.97^{16,17,18} and ICC 0.51–0.89^{7,8,18}, respectively)⁶. It would be of additional value to know about interreader reliability in terms of smallest detectable change (SDC), to be able to judge whether the SDC is sufficiently small to detect the proposed MIC.

The aim of our study is 3-fold: first, to define which SPARCC score best approximates a “positive MRI” judgment; second, to establish an SDC for a 3-month period and for a 1-year period; third, to describe which variation in SPARCC score over a 3-month and 1-year period can be expected in patients without (change in) treatment.

MATERIALS AND METHODS

Study population. Data from the SPondyloArthritis Caught Early (SPACE) cohort are used for this analysis. An extensive description of the SPACE cohort is given elsewhere¹⁹. In short, the SPACE cohort is an ongoing cohort started in January 2009, including patients aged 16 years and older with back pain (≥ 3 months, ≤ 2 years, onset age < 45 years) visiting the rheumatology outpatient clinics of 5 participating centers. Patients were not included if they had other painful conditions (not related to SpA) that could interfere with the evaluation of the disease. After signing informed consent forms, all patients underwent a diagnostic examination at baseline, including MRI and plain radiographs of the SI joints, HLA-B27 testing, and examining for other SpA features. Patients fulfilling the ASAS axSpA criteria or patients with possible axSpA were included for followup visits after 3 and 12 months. Possible axSpA was defined as the presence of at least 1 specific SpA feature with a high positive likelihood ratio (LR+ above 6) or at least 2 less specific SpA features (LR+ below 6), but not fulfilling the ASAS axSpA criteria²⁰.

MRI of the SI joints. MR imaging was performed on a 1.5T scanner, acquiring T1-weighted turbo spin echo (T1TSE; TR 550/TE 10) and short-tau inversion recovery (STIR; TR 2500/TE 60) sequences. Slices were of 4 mm thickness in the coronal oblique view of the SI joints.

All readers in our study ($n = 4$) were extensively trained in reading MRI according to the ASAS definition and the SPARCC score during a calibration session, supervised by a senior radiologist (MR) and a senior rheumatologist (DvdH), discussing definitions of lesions, examples, and pitfalls. Next, all readers independently read 30 blinded MRI to calculate agreement on the ASAS definition ($\kappa = 0.75$ to $\kappa = 0.87$ for the different pairs of readers), and to calculate agreement on SPARCC scores (ICC 0.81 to ICC 0.95 for the different reader pairs on status scores and ICC 0.78 to ICC 0.97 on change scores). The mean baseline SPARCC score was 10.3 (SD 11.7); the mean followup score was 7.4 (10.0); the mean change score was -2.9 (9.9). A consensus meeting followed, in which the same supervising rheumatologist and radiologist participated. In addition, all readers participated in a reading exercise in which the original developers of the SPARCC score participated as well. The mean SPARCC score of the 32 evaluated cases was 5.3 (7.1). Agreement on the status SPARCC scores was acceptable (ICC 0.77) for all readers including the original developers. A consensus meeting with the

original developers was organized, and agreement was considered sufficiently high to start scoring the SPACE cohort.

Two separate reading campaigns were performed by different pairs of readers (RvdB and MdH in exercise 1; PB and MdH in exercise 2), with partly overlapping patients and images. Patients in the first reading exercise were included between January 2009 and November 2012 in 5 different centers and patients in the second reading exercise were included between January 2009 and October 2013 in 1 center. In exercise 1, baseline and 3-month MRI-SI were evaluated; in exercise 2, baseline, 3-month, and 1-year MRI-SI were evaluated. In both campaigns, MRI-SI were independently read by the 2 trained readers for the fulfillment of the ASAS definition¹ and according to the SPARCC score⁶, blinded for the time sequence of the MRI-SI as well as for clinical and laboratory data.

Every inflammatory lesion typical for SpA was marked according to the SPARCC score. In the next step, the readers took a look at the MRI again, and marked whether the MRI was positive or negative according to the ASAS definition based on the global evaluation of the entire MRI-SI. An MRI-SI can be marked positive according to the ASAS definition if ≥ 1 bone marrow edema (BME) lesion highly suggestive of SpA is present on ≥ 2 consecutive slices, or if several BME lesions highly suggestive of SpA are visible on a single slice. The presence of synovitis, enthesitis, or capsulitis only, without BME, is not sufficient for a positive MRI-SI¹. In case the 2 readers disagreed on the presence of a positive MRI, a third trained reader served as adjudicator (VNC in campaign 1; RvdB in campaign 2).

According to the SPARCC score, the presence of increased signal corresponding to BME lesions highly suggestive of SpA is marked on 6 consecutive slices of an MRI-SI, starting on the slice on which at least 1 cm of vertical height of the cartilage compartment is visible, from anterior to posterior, assessing the cartilaginous compartment of the SI joints and the antero-inferior portion of the SI joint. At the posterior aspect of the SI joints there is a natural division into upper and lower quadrants by intervening fat and fibrous tissue. When less than 1 cm of a quadrant is visible, it is no longer scored¹⁸. Each SI joint is divided into 4 quadrants (upper iliac, lower iliac, upper sacrum, and lower sacrum)^{6,18}. The maximum score for 2 SI joints on each slice is 8. In addition to these 8 points per slice, a score for “intensity” (adding 1 point) may be assigned to each SI joint if an “intense signal” is seen in any quadrant on each slice. The signal from presacral blood vessels defined a lesion that is scored as intense. Further, a score for depth (adding 1 point) may be assigned to each SI joint if a homogeneous and unequivocal increase in signal is extending over a depth of at least 1 cm from the articular surface on each slice, resulting in a maximum score of 12 points per slice. The total maximum SPARCC score is 72^{6,18}. The mean SPARCC scores of the 2 readers were used; in case there was a third reader involved because of disagreement among the 2 initial readers regarding a positive MRI according to the ASAS definition, the mean of the SPARCC scores of the 2 readers in agreement of a “positive MRI” for that particular case was used.

For both the SPARCC score and the ASAS definition assessment on the STIR sequence, the readers took into account the findings on the T1TSE sequences as well.

Treatment. Patients in the SPACE cohort are not treated according to a fixed protocol, but according to usual clinical practice by their rheumatologist. Treatment with nonsteroidal antiinflammatory drugs (NSAID) was recorded according to the ASAS recommendations, resulting in a 0–100 score whereby 0 means no NSAID intake at all and 100 means a daily intake at a full dose over the whole period of interest²¹. Treatment with disease-modifying antirheumatic drug (DMARD) and anti-tumor necrosis factor (TNF) therapy was recorded as present or absent.

To investigate variation in SPARCC scores over time, patients were categorized according to their treatment over the period of interest: no treatment, stable NSAID and/or DMARD intake, and change in NSAID and/or DMARD intake. Patients receiving anti-TNF therapy during the period of interest were excluded from the analysis on variation in SPARCC scores.

Statistical analysis. Baseline characteristics of patients in both groups were investigated using descriptive statistics. Agreement (Cohen κ) between MRI positivity based on several SPARCC score cutoff values (≥ 1 , ≥ 2 , ≥ 3 , and ≥ 4) and the consensus judgment of a “positive MRI” as external standard was calculated using cross-tabulation. Agreement on positive cases (positive agreement) and on negative cases (negative agreement) was also calculated²².

Changes in SPARCC score over the period of interest [baseline to 3 months (both campaigns); baseline to 1 year (campaign 2)] were visualized in cumulative probability plots in which patients were grouped based on treatment. Next, SDC were calculated based on a 95% level of agreement (95% LoA) between the 2 readers on the change scores for both baseline to 3-month and baseline to 1-year intervals, using the following formula:

$$SDC = (1.96 * SD_{\Delta \text{change-scores}}) / (\sqrt{2} * \sqrt{k})$$

whereby k represents the number of readers (2 in our study)²³. The SDC are also displayed in Bland-Altman plots that plot the mean SPARCC score changes of the 2 readers (X axis) and the interreader differences in SPARCC score changes (Y axis). In addition, the mean of the interreader differences in SPARCC score changes (which is a reflection of the systematic error between the 2 readers) and the 95% LoA are presented in these plots. SPSS software version 20.0 was used for statistical analysis.

RESULTS

Patients with available baseline MRI-SI were included in the analysis of the agreement between the SPARCC score cutoff value and positive MRI [n = 294 (campaign 1) and n = 249 (campaign 2)]. There is a partial overlap (49.1%) between patients included in campaign 2 and those included in campaign 1. In both campaigns the population was young, with short symptom duration. Around one-third of the patients were male and around one-third fulfilled the ASAS axSpA criteria (Table 1).

A 3-month followup MRI-SI was available for 154 patients in campaign 1. However, 3/154 patients received anti-TNF therapy during this period and were therefore excluded from the followup part of the analysis of the SPARCC score changes over time and SDC. In campaign 2, a 3-month followup MRI-SI was available in 70 patients, and in 76 patients a 1-year followup MRI-SI was available. Two patients received anti-TNF therapy, leaving MRI-SI of 68 (3-month period) and 74 patients (1-year period) for followup analyses.

SPARCC score cutoff. In both campaigns, there was a high level of agreement between MRI positivity based on all tested SPARCC score cutoff values and the consensus judgment of a “positive MRI” as external standard (Table 2). A cutoff value of ≥ 2 showed the highest κ values (0.94 in campaign 1 and 0.98 in campaign 2) and provided the best balance in terms of misclassifications in comparison to the external standard; 5 false-positive and 1 false-negative classification in campaign 1; 0 false-positive and 1 false-negative classification in campaign 2.

Smallest detectable change of SPARCC score. Of the patients with available followup MRI, the mean SPARCC score at baseline was 4.0 (8.3) and 2.3 (5.7; campaign 1 and 2, respectively). At 3 months, the mean SPARCC score was 3.4 (6.7) and 1.6 (3.8) (campaigns 1 and 2, respectively),

Table 1. Baseline characteristics of patients in reading campaign 1 and patients in reading campaign 2. A proportion (49.1%) of the patients was included in both campaigns. Data are n (%) unless indicated otherwise.

Characteristics	Reading Campaign 1, n = 294	Reading Campaign 2, n = 249
Age at inclusion, yrs, mean \pm SD	31.2 \pm 10.4	31.1 \pm 11.5
Male	102 (34.7)	81 (32.5)
Duration of back pain, mos, mean \pm SD	13.1 \pm 7.1	13.3 \pm 7.4
HLA-B27–positive	113 (38.4)	79 (31.7)
Positive family history SpA	113 (38.4)	89 (35.7)
IBP	195 (66.3)	142 (57.0)
Psoriasis	28 (9.5)	26 (10.4)
Dactylitis	16 (5.4)	8 (3.2)
Enthesitis	49 (16.7)	24 (9.6)
Uveitis	24 (8.2)	18 (7.2)
IBD	20 (6.8)	19 (7.6)
Good response to NSAID	112 (38.1)	69 (27.7)
Elevated CRP/ESR	58 (19.7)	42 (16.9)
Asymmetric lower limb arthritis	48 (16.3)	26 (10.4)
Radiographic sacroiliitis*	23 (7.8)	24 (9.6)
Sacroiliitis on MRI**	67 (22.8)	31 (12.4)
SPARCC score, mean \pm SD	2.9 \pm 7.7	1.3 \pm 4.4
SPARCC score, mean \pm SD of subgroup of patients with CBP	0.2 \pm 0.5 (n = 58)	0.2 \pm 0.5 (n = 54)
CRP, mean \pm SD	6.9 \pm 13.0	7.3 \pm 11.6
ASDAS, mean \pm SD	2.6 \pm 1.1	2.7 \pm 0.8
BASDAI, mean \pm SD	4.6 \pm 2.5	4.6 \pm 2.1
BASFI, mean \pm SD	3.0 \pm 2.3	3.2 \pm 2.4
ASAS axSpA criteria–positive	119 (40.5)	83 (33.3)

*Radiographic sacroiliitis according to the modified New York criteria²⁵.

**Sacroiliitis on MRI according to the ASAS definition (consensus judgment)¹. MRI: magnetic resonance imaging; IBP: inflammatory back pain; IBD: inflammatory bowel disease; CRP: C-reactive protein; ESR: erythrocyte sedimentation rate; CBP: chronic back pain; SpA: spondyloarthritis; NSAID: nonsteroidal antiinflammatory drugs; SPARCC: SpondyloArthritis Research Consortium of Canada; ASDAS: Ankylosing Spondylitis Disease Activity Score; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASFI: Bath Ankylosing Spondylitis Functional Index; ASAS: Assessment of Spondyloarthritis international Society; axSpA: axial SpA.

and at 1 year the mean SPARCC score was 1.4 (SD 4.0; campaign 2).

Bland-Altman plots show the mean of the 2 readers in SPARCC score changes over the 3-month (campaign 1; Figure 1) and over the 3-month and 1-year period (campaign 2; Supplementary Figure 1, available online at jrheum.org) against the difference between the 2 readers in SPARCC score changes over those periods. The plots show that a large number of observations is clustered around the mean difference of 0, and that differences between readers occur with similar amplitude across the entire range of the SPARCC score (a homoscedastic pattern). To visualize the high number of overlapping observations, series of ranges were defined. All observations were grouped into their corresponding range, increasing exponentially on the positive side of 0 and

Table 2. Various SPARCC cutoff values (based on the mean SPARCC score of the 2 readers) tested against the ASAS definition of a positive MRI, in reading campaign 1 and reading campaign 2.

	Positive MRI (ASAS)	Negative MRI (ASAS)
Reading campaign 1, n = 294		
SPARCC ≥ 1	67	21
SPARCC < 1	0	206
κ: 0.82	PA: 95.2%	NA: 86.5%
SPARCC ≥ 2	66	5
SPARCC < 2	1	222
κ: 0.94	PA: 98.7%	NA: 95.7%
SPARCC ≥ 3	57	1
SPARCC < 3	10	226
κ: 0.89	PA: 97.6%	NA: 91.2%
SPARCC ≥ 4	47	1
SPARCC < 4	20	226
κ: 0.77	PA: 95.6%	NA: 81.7%
Reading campaign 2, n = 249		
SPARCC ≥ 1	31	5
SPARCC < 1	0	213
κ: 0.91	PA: 98.8%	NA: 92.5%
SPARCC ≥ 2	31	1
SPARCC < 2	0	217
κ: 0.98	PA: 99.8%	NA: 98.4%
SPARCC ≥ 3	25	0
SPARCC < 3	6	218
κ: 0.88	PA: 98.6%	NA: 89.3%
SPARCC ≥ 4	21	0
SPARCC < 4	10	218
κ: 0.79	PA: 97.8%	NA: 80.8%

SPARCC cutoff values are based on the mean SPARCC scores of the 2 initial readers; in case there was a third reader involved, the mean of the SPARCC scores of the 2 readers in agreement of a “positive MRI” (ASAS definition) for that particular case were used. SPARCC: SpondyloArthritis Research Consortium of Canada; MRI: Magnetic resonance imaging; ASAS: Assessment of Spondyloarthritis International Society; PA: positive agreement; NA: negative agreement.

decreasing exponentially on the negative side, displayed on the X axis. The SDC based on the 95% LoA in campaign 1 over the 3-month period is 3.4 SPARCC units, depicted in Figure 1 as the dark grey area reflecting the SDC of both increased and decreased SPARCC scores over time. The SDC (95% LoA) in campaign 2 over the 3-month period is 2.1 SPARCC units (Supplementary Figure 1, top panel, available online at jrheum.org) and over the 1-year period, 2.4 SPARCC units (Supplementary Figure 1, bottom panel).

In comparison, the SDC based on the 80% LoA are 2.2 (80% LoA -4.3 to 4.5), 1.4 (80% LoA -2.5 to 2.9), and 1.6 (80% LoA -3.2 to 3.0) SPARCC units, respectively.

Change in SPARCC scores over 3 months and 1 year. Eighty-one out of 151 patients in campaign 1 (53.6%) showed no change in SPARCC score over the 3-month period of which 75/81 (92.6%) had a SPARCC score of 0 at both timepoints. In the 70 out of 151 patients (46.4%) showing a change in SPARCC score, 27 increased and 43 decreased

[mean change -1.1 (6.3); median change -0.5 (range -16.5 to 16.0); Figure 2 and Table 3]. In 26 out of 70 patients (37.1%) with SPARCC score changes, the change was more than the SDC (3.4); in 16 patients the SPARCC score decreased (2 patients without treatment, 11 with stable NSAID intake, 2 with stable NSAID and DMARD intake, 1 started NSAID intake) and in 10 patients it increased (2 without treatment, 7 with stable NSAID intake, 1 started NSAID intake). In the remaining 44 patients (62.9%) the SPARCC score changes were within the area still compatible with measurement error.

In campaign 2, two followup intervals for the same patients were available. Over the 3-month period, SPARCC score did not change in 48 out of 68 patients (70.6%); 46/48 patients (95.8%) had a SPARCC score of 0 at both timepoints. In the remaining 20 patients (29.4%) the SPARCC score changed; 14 patients showed a decrease and 6 patients an increase [mean change -3.1 (4.6); median change -1.5 (range -12.5 to 5); Supplementary Figure 2a, available online at jrheum.org, and Table 3]. Eleven out of 20 patients (55.0%) showed a SPARCC score change > SDC (2.1); 10 patients showed a decrease (1 without treatment, 6 with stable NSAID intake, 2 with stable NSAID and DMARD intake, 1 started NSAID intake) and 1 patient increased (started NSAID intake). The remaining 9 patients (45.0%) had SPARCC score changes still compatible with measurement error.

The results over the 1-year period in campaign 2 are similar to the results over the 3-month period in campaign 2, although more variation between patients is seen; 51/74 patients (68.9%) did not show a change in SPARCC score; of them, 50 patients (98.0%) had a SPARCC score of 0 at both timepoints. The remaining 23 patients (31.1%) showed a change in SPARCC score; 16 patients showed a decrease and 7 an increase [mean change -2.9 (7.5); median change -1.0 (range -18.0 to 12.0); Supplementary Figure 2b, available online at jrheum.org, and Table 3]. Fourteen out of the 23 patients (60.9%) showed a SPARCC score change of more than the SDC (2.4); 10 patients showed a decrease (2 without treatment, 4 with stable NSAID intake, 2 with stable DMARD intake, 1 stopped NSAID intake, 1 started but stopped again NSAID intake), and 4 patients showed an increase (1 with stable NSAID intake, 1 stopped NSAID intake, 1 started NSAID intake, 1 stopped DMARD intake but continued NSAID intake). In the remaining 9 patients (39.1%), SPARCC score changes were not beyond measurement error.

The majority of the patients showing changes in SPARCC score of more than the SDC in both campaigns [20/26 (76.9%; campaign 1), 8/11 (72.7%; 3-month period campaign 2) and 7/14 (50.0%; 1-year period campaign 2)] had stable NSAID and/or DMARD intake.

DISCUSSION

Our study, performed in the SPACE cohort, has shown in 2

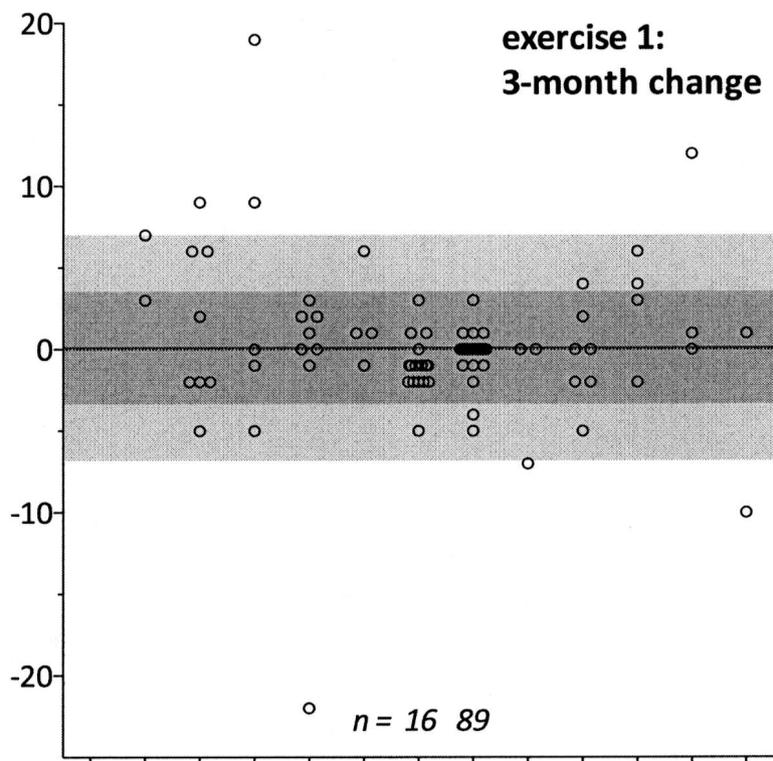


Figure 1. Bland-Altman plots showing the mean SPARCC score change of the 2 readers (X axis) versus the Δ SPARCC score changes of the 2 readers (Y axis). The large number of overlapping observations clustered around the mean difference of 0 are displayed in a series of ranges increasing exponentially on the positive side of 0 and decreasing exponentially on the negative side (X axis). The “n” above the X axis shows the number of observations per group. Solid grey line represents the overall mean of the Δ SPARCC score changes (equivalent to systematic error between the 2 readers). The light grey area represents the 95% levels of agreement (LoA), and the dark grey area represents the smallest detectable change (SDC) in both directions (increase in SPARCC score and decrease in SPARCC score over time). Mean of the Δ SPARCC scores is 0.1 (95% LoA -6.8 to 7.0); SDC 3.4. Observations are clustered in the range -0.5 to 0.5 (n = 89) and the range -1 to -0.5 (n = 16).

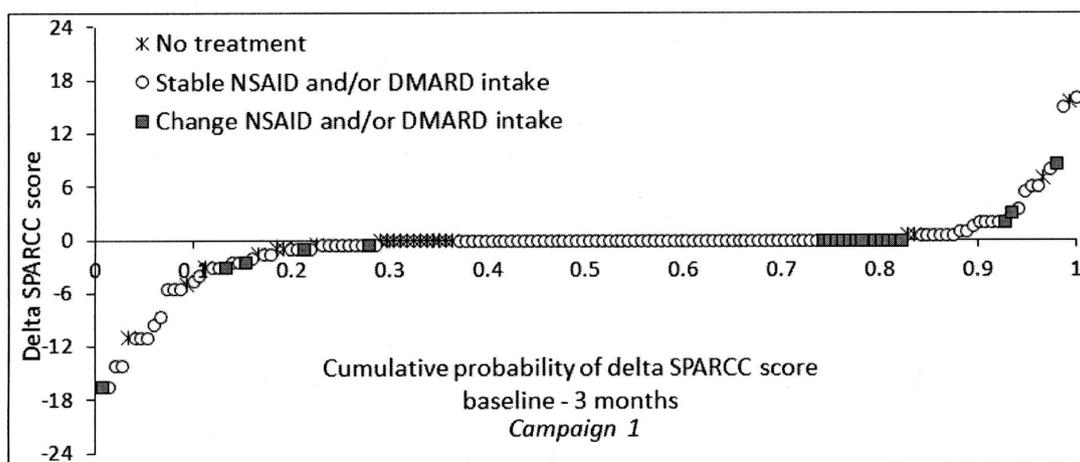


Figure 2. Cumulative probability plots of all Δ SPARCC scores over a 3-month period in campaign 1, with symbols indicating treatment over the investigated period. NSAID: nonsteroidal antiinflammatory drugs; DMARD: disease-modifying antirheumatic drug; SPARCC: SpondyloArthritis Research Consortium of Canada.

Table 3. All changes in SPARCC score in patients grouped according to treatment.

	No Treatment	Stable NSAID/DMARD	Start NSAID/DMARD	Stop NSAID/DMARD
Campaign 1 – baseline to 3 mos				
No SPARCC score change	n = 13	n = 56	n = 7	n = 5
Increase in SPARCC score	n = 4, mean: 5.9 (SD 7.1), range: 0.5 to 15.5	n = 20, mean: 3.7 (SD 4.6), range: 0.5 to 16.0	n = 3, mean: 4.5 (SD 3.5), range: 2 to 8.5	—
Decrease in SPARCC score	n = 9, mean: -2.8 (SD 3.4), range: -11.0 to -0.5	n = 29, mean: -4.9 (SD 4.9), range: -16.5 to -0.5	n = 5, mean: -4.7 (SD 6.7), range: -16.5 to -0.5	—
Campaign 2 – baseline to 3 mos				
No SPARCC score change	n = 4	n = 31	n = 8	n = 5
Increase in SPARCC score	—	n = 5, mean: 0.6 (SD 0.2), range: 0.5 to 1.0	n = 1, mean: 5 (—)	—
Decrease in SPARCC score	n = 2, mean: -5.0 (SD 6.4), range: -12.5 to -0.5	n = 10, mean: -4.6 (SD 3.3), range: -10.5 to -0.5	n = 2, mean: -6.5 (SD 8.5), range: -12.5 to -0.5	—
Campaign 2 – baseline to 1 yr				
No SPARCC score change	n = 10	n = 28	n = 7	n = 6
Increase in SPARCC score	—	n = 3, mean: 3.0 (SD 1.3), range: 1.5 to 4.0	n = 1, mean: 12 (—)	n = 3, mean: 3.3 (SD 4.5), range: 0.5 to 8.5
Decrease in SPARCC score	n = 3, mean: -6.8 (SD 8.5), range: -16.5 to -0.5	n = 8, mean: -5.8 (SD 5.1), range: -14.5 to -1.0	n = 1, mean: -0.5 (—)	n = 4, mean: -7.6 (SD 8.4), range: -18.0 to -0.5

SPARCC: SpondyloArthritis Research Consortium of Canada; NSAID: nonsteroidal antiinflammatory drugs; DMARD: disease-modifying antirheumatic drug.

campaigns that a cutoff value of 2 SPARCC units is best compatible with a consensus judgment of a positive versus negative MRI according to the ASAS definition. These results were expected because the ASAS definition of a positive MRI-SI includes (apart from a qualitative part, i.e., BME lesions highly suggestive of spondyloarthritis) a quantitative part that requires at least 1 BME lesion visible on at least 2 consecutive slices or several lesions on a single slice¹. However, in theory, a SPARCC score can be high because of the presence of several small lesions (highly suggestive of SpA) scattered over several slices (e.g., 1 lesion on slice 1, another lesion on slice 4 and another lesion on slice 6), but still not fulfilling the ASAS definition. A SPARCC score can also be high if 1 lesion is assigned as “intense” or “deep” while it is visible on only 1 slice. Moreover, the SPARCC score prescribes that lesions are scored in the 6 slices representing the largest proportion of the cartilaginous component of the SI joints, while the ASAS definition takes all slices into account^{1,6,18}.

Occasionally, part of a lesion may be visible on only 1 of the 6 selected slices, while the remaining part of the lesion is visible outside those 6 selected slices, or a slice outside those selected 6 shows several lesions. However, these considerations are mainly theoretical and do not appear very frequently (1 case in our study). Therefore, a SPARCC cutoff level of 2 units may serve as a surrogate for the ASAS definition of a positive MRI and could be used in clinical

trials with central efficacy reading to derive a dichotomy (positive vs negative).

The SDC in campaign 2 (2.1 SPARCC units over 3 mos and 2.4 over 1 yr) are close to the proposed MIC of 2.5 SPARCC units, which was calculated using pooled changes over 12 and 52 weeks¹⁵, but the SDC of campaign 1 (3.4) is slightly higher. This suggests that the previously proposed MIC is close to measurement error in our study based on 2 different reader pairs and different periods of followup.

A large proportion of the SPARCC score changes seen in the patients in both reading campaigns could be considered noise because these changes were smaller than the SDC [62.9% and 45% (3-months, campaign 1 and 2) and 39.1% (1-year in campaign 2)]. To investigate the influence of nonbiological treatment on inflammation on MRI-SI, only patients with SPARCC score changes greater than the SDC were taken into account. Somewhat surprisingly, the majority of patients with a change in SPARCC score were on stable NSAID and/or DMARD treatment. Some patients taking stable doses of NSAID increased in SPARCC score while others who also had stable NSAID intake showed a decrease in SPARCC score. These results are in line with the results found in trials where patients using NSAID — either in an open-label trial or in a placebo group — showed also both increased and decreased inflammation scores on MRI-SI over 6 and 16 weeks, respectively^{12,24}. Moreover, patients with stable background treatment in the placebo group of the

ABILITY-1 trial had slightly decreased SPARCC scores at group level, as we found in our study⁴.

Although too few patients in the SPACE cohort used DMARD to draw conclusions on the effect of DMARD, comparable effects can be expected. The comparator group in the ESTHER trial using sulfasalazine showed a mean decrease of 1.7 and 1.9 SPARCC units over 24 and 48 weeks, respectively¹⁴. In the comparator group of another trial where patients used methotrexate, a mean of 1.4 (95% CI -0.8, 3.5) inflammatory lesions resolved over 30 weeks²⁵. Although an overall decrease in inflammation score was seen in these trials, some patients increased in inflammation score on MRI-SI when looking at the individual level^{14,25}. These results indicate that in patients with stable treatment, changes may occur in BME on MRI-SI that are beyond measurement error, which may point to true fluctuation in inflammatory activity over time.

The direct comparisons of our results with the results of drug efficacy trials is difficult because the SPACE cohort is an observational cohort including unselected patients with back pain of short duration resulting in a heterogeneous patient population, with low numbers of a positive MRI and low baseline mean SPARCC scores, while drug efficacy trials select patients with high levels of disease activity. In patients selected because of a high level of disease activity, a decrease in scores is more likely (regression to the mean) in comparison to an unselected group of patients. Thus, the patients in the SPACE cohort will likely not be representative of patients in trials. Nevertheless, we have also observed an overall decrease in the SPACE cohort, just as in the trials. This might occur because patients preferably seek help in case of maximum complaints, which is by default the timepoint of inclusion in the SPACE cohort. It is possible that the results would have been different if our study had been performed in a longstanding or severely diseased group of patients. Further, the SPACE cohort is not designed to investigate the effects of treatment for inflammation on MRI. For example, and in contrast to drug efficacy trials, there is not a good relation between the start date of therapy and the date of the MRI.

Another possible limitation is that the readers have given their judgment based on the ASAS definition immediately after the evaluation according to the SPARCC score. Because the quantitative part of the ASAS definition resembles a SPARCC score of 2, the choice of the value of 2 as the best SPARCC score to serve as cutoff level for negative and positive MRI may not be entirely independent. Yet, it should be stressed that the readers were not trained in scoring the MRI positive according to the ASAS definition if the SPARCC score was 2 or higher, but based their judgment of a positive MRI only on the complete MRI view. Nevertheless, it would have been better if different scores were acquired independently or even by different readers, as is frequently the case in clinical trials. Besides, we repeat that our study

was not primarily designed to develop lesion-based cutoffs of a positive MRI for the purpose of disease classification.

A SPARCC score of 2 as cutoff value best reflects the caesura between a positive and negative MRI according to the ASAS definition. This cutoff can be used (in clinical trials) to create a dichotomous MRI variable of potential prognostic interest. The SDC we have obtained in our 2 experiments are close enough to the proposed MIC of 2.5 SPARCC units, which adds credibility to a cutoff level of 2.5 units in that it represents a true difference rather than only measurement error. Surprisingly, while patients experience stable treatment, true (> SDC) changes in SPARCC score over time (both increases and decreases) were frequently observed. This observation strongly suggests that MRI activity fluctuates over time.

ACKNOWLEDGMENT

The authors thank Maarten Boers, MD, PhD, from the VU University Medical Center, Amsterdam, the Netherlands, for his help and advice in creating the Bland-Altman plots.

ONLINE SUPPLEMENT

Supplementary data for this article are available online at jrheum.org.

REFERENCES

1. Rudwaleit M, Jurik AG, Hermann KG, Landewe R, van der Heijde D, Baraliakos X, et al. Defining active sacroiliitis on magnetic resonance imaging (MRI) for classification of axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI group. *Ann Rheum Dis* 2009;68:1520-7.
2. Rudwaleit M, van der Heijde D, Landewe R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;68:777-83.
3. Landewe R, Braun J, Deodhar A, Dougados M, Maksymowych WP, Mease PJ, et al. Efficacy of certolizumab pegol on signs and symptoms of axial spondyloarthritis including ankylosing spondylitis: 24-week results of a double-blind randomised placebo-controlled Phase 3 study. *Ann Rheum Dis* 2014;73:39-47.
4. Sieper J, van der Heijde D, Dougados M, Mease PJ, Maksymowych WP, Brown MA, et al. Efficacy and safety of adalimumab in patients with non-radiographic axial spondyloarthritis: results of a randomised placebo-controlled trial (ABILITY-1). *Ann Rheum Dis* 2013;72:815-22.
5. Sieper J, Lenaerts J, Wollenhaupt J, Rudwaleit M, Mazurov VI, Myasoutova L, et al. Efficacy and safety of infliximab plus naproxen versus naproxen alone in patients with early, active axial spondyloarthritis: results from the double-blind, placebo-controlled INFAST study, Part 1. *Ann Rheum Dis* 2014;73:101-7.
6. Maksymowych WP, Inman RD, Salonen D, Dhillon SS, Williams M, Stone M, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. *Arthritis Rheum* 2005;53:703-9.
7. Landewe RB, Hermann KG, van der Heijde DM, Baraliakos X, Jurik AG, Lambert RG, et al. Scoring sacroiliac joints by magnetic resonance imaging. A multiple-reader reliability experiment. *J Rheumatol* 2005;32:2050-5.
8. 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.
9. Braun J, Baraliakos X, Hermann KG, van der Heijde D, Inman RD, Deodhar AA, et al. Golimumab reduces spinal inflammation in ankylosing spondylitis: MRI results of the randomised, placebo-controlled GO-RAISE study. *Ann Rheum Dis* 2012; 71:878-84.
 10. Baraliakos X, Brandt J, Listing J, Haibel H, Sorensen H, Rudwaleit M, et al. Outcome of patients with active ankylosing spondylitis after two years of therapy with etanercept: clinical and magnetic resonance imaging data. *Arthritis Rheum* 2005;53:856-63.
 11. Braun J, Landewe R, Hermann KG, Han J, Yan S, Williamson P, et al. Major reduction in spinal inflammation in patients with ankylosing spondylitis after treatment with infliximab: results of a multicenter, randomized, double-blind, placebo-controlled magnetic resonance imaging study. *Arthritis Rheum* 2006;54:1646-52.
 12. Jarrett SJ, Sivera F, Cawkwell LS, Marzo-Ortega H, McGonagle D, Hensor E, et al. MRI and clinical findings in patients with ankylosing spondylitis eligible for anti-tumour necrosis factor therapy after a short course of etoricoxib. *Ann Rheum Dis* 2009;68:1466-9.
 13. van Onna M, Jurik AG, van der Heijde D, van Tubergen A, Heuft-Dorenbosch L, Landewe R. HLA-B27 and gender independently determine the likelihood of a positive MRI of the sacroiliac joints in patients with early inflammatory back pain: a 2-year MRI follow-up study. *Ann Rheum Dis* 2011;70:1981-5.
 14. Song IH, Hermann K, Haibel H, Althoff CE, Listing J, Burmester G, et al. Effects of etanercept versus sulfasalazine in early axial spondyloarthritis on active inflammatory lesions as detected by whole-body MRI (ESTHER): a 48-week randomised controlled trial. *Ann Rheum Dis* 2011;70:590-6.
 15. Maksymowych WP, Lambert RG, Brown LS, Pangan AL. Defining the minimally important change for the SpondyloArthritis Research Consortium of Canada spine and sacroiliac joint magnetic resonance imaging indices for ankylosing spondylitis. *J Rheumatol* 2012;39:1666-74.
 16. van der Heijde D, Sieper J, Maksymowych WP, Brown MA, Lambert RG, Rathmann SS, et al. Spinal inflammation in the absence of sacroiliac joint inflammation on magnetic resonance imaging in patients with active nonradiographic axial spondyloarthritis. *Arthritis Rheumatol* 2014;66:667-73.
 17. Weber U, Maksymowych WP, Jurik AG, Pfirrmann CW, Ruffbach K, Kissling RO, et al. Validation of whole-body against conventional magnetic resonance imaging for scoring acute inflammatory lesions in the sacroiliac joints of patients with spondylarthritis. *Arthritis Rheum* 2009;61:893-9.
 18. Maksymowych W, Dhillon SS, Chiowchanwisawakit P, Pedersen SJ, Martinez B, Ostergaard M, et al. Development and validation of Web-based training modules for systematic evaluation of active inflammatory lesions in the spine and sacroiliac joints in spondyloarthritis. *J Rheumatol Suppl* 2009;84:48-57.
 19. van den Berg R, de Hooze M, van Gaalen F, Reijnen M, Huizinga T, van der Heijde D. Percentage of patients with spondyloarthritis in patients referred because of chronic back pain and performance of classification criteria: experience from the Spondyloarthritis Caught Early (SPACE) cohort. *Rheumatology* 2013;52:1492-9.
 20. Rudwaleit M, van der Heijde D, Khan MA, Braun J, Sieper J. How to diagnose axial spondyloarthritis early. *Ann Rheum Dis* 2004;63:535-43.
 21. Dougados M, Simon P, Braun J, Burgos-Vargas R, Maksymowych WP, Sieper J, et al. ASAS recommendations for collecting, analysing and reporting NSAID intake in clinical trials/epidemiological studies in axial spondyloarthritis. *Ann Rheum Dis* 2011;70:249-51.
 22. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990;43:551-8.
 23. Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. *Ann Rheum Dis* 2005;64:179-82.
 24. Barkham N, Keen HI, Coates LC, O'Connor P, Hensor E, Fraser AD, et al. Clinical and imaging efficacy of infliximab in HLA-B27-Positive patients with magnetic resonance imaging-determined early sacroiliitis. *Arthritis Rheum* 2009; 60:946-54.
 25. Marzo-Ortega H, McGonagle D, Jarrett S, Haugeberg G, Hensor E, O'Connor P, et al. Infliximab in combination with methotrexate in active ankylosing spondylitis: a clinical and imaging study. *Ann Rheum Dis* 2005;64:1568-75.