The gestalt of spondyloarthritis: From early recognition to long-term imaging outcomes

Sepriano, A.R.

Chapter 11

Which imaging outcomes for axSpA are most sensitive to change? A 5-Year analysis of The DESIR Cohort

Alexandre Sepriano, Sofia Ramiro, Désirée van der Heijde, Maxime Dougados, Pascal Claudepierre, Antoine Feydy, Monique Reijnierse, Damien Loeuille, Robert Landewé

Accepted for publication, Arthritis Care & Research

ABSTRACT

Objective: To compare the sensitivity to change of different imaging scoring methods in patients with early axial spondyloarthritis (axSpA).

Methods: Patients from the DESIR cohort fulfilling the ASAS axSpA criteria were included. Radiographs and MRI of the sacroiliac joints (SIJ) and spine were obtained at baseline, 1, 2 and 5 years. Each image was scored by 2 or 3 readers in 3 separate 'reading-waves'. The rate of change of outcomes measuring spinal and SIJ inflammation (e.g. SPARCC score) and structural damage on MRI (e.g. ≥3 fatty lesions) and radiographs (e.g. mNY grading) was assessed using multilevel generalized estimating equations (GEE) models (taking all readers and waves into account). To allow comparisons across outcomes, rates were standardized (difference between the individual's value and the population mean divided by the standard deviation).

Results: In total, 345 patients were included. Inflammation on MRI-SIJ (standardized rate range: -0.278; -0.441) was more sensitive to change compared to spinal inflammation (range: -0.030; -0.055). Structural damage in the SIJ showed a higher standardized rate of change on MRI-SIJ (range: 0.015-0.274) compared to X-SIJ (range: 0.043-0.126). MRI-SIJ damage defined by ≥3 fatty lesions showed the highest sensitivity to change (0.274). Spinal structural damage slowly progressed over time with no meaningful difference between radiographic (range: 0.037-0.043) and MRI structural outcomes (range: 0.008-0.027).

Conclusion: Structural damage assessed in pelvic radiographs has low sensitivity to change, while fatty lesions detected on MRI-SIJ are a promising alternative. In contrast, MRI-spine is not better than X-spine in detecting structural changes in early axSpA patients.

INTRODUCTION

Several imaging outcomes have been developed to assess inflammation and structural damage over time in patients with axial spondyloarthris (axSpA). A recent systematic literature review (SLR) informing the EULAR recommendations for the use of imaging in the diagnosis and management of SpA in clinical practice identified several studies testing the utility of magnetic resonance imaging (MRI) and radiographs of the sacroiliac joints (SIJ) and spine on monitoring disease activity and structural damage over time.[1] However, these studies mostly assessed only one score each, and focused on comparing imaging to clinical measures of disease activity, disability and mobility, which means they mostly addressed their validity.

In addition to validity, in order to prioritize imaging outcomes measuring similar aspects of the disease (i.e. inflammation or structural damage), the other aspects of the Outcome Measures in Rheumatology (OMERACT) filter, namely discrimination (sensitivity to change and reliability) and feasibility should also be taken into account.[2] However, direct comparisons of the discriminative ability and feasibility of imaging outcomes in axSpA have been seldom performed, and almost only in later phases of the disease (radiographic axSpA; r-axSpA).[3-5] An exception to this, is the comparison of the different spinal radiographic scoring methods performed in the DESIR cohort and previously reported by us.[6]

A better understanding on which imaging findings (reflecting inflammation or structural damage), imaging modality (MRI or radiographs) and anatomical location (SIJ or spine) are most informative to monitor axial changes in the entire spectrum of axSpA (also including non-radiographic axSpA; nr-axSpA) over time is still a major unmet need. We aimed to compare the sensitivity to change of different MRI and radiographic scoring methods in patients with early axSpA.

METHODS

Patients and study design

Five-year data from patients with early axSpA from the DEvenir des Spondylarthopathies Indifférenciées Récentes (DESIR) cohort have been used (clinicaltrials.gov ID: NCT01648907).[7] Patients had to fulfill the Assessment of SpondyloArthritis international Society (ASAS) axSpA criteria and to have ≥1 radiograph and/or MRI reading available during the 5-year follow-up to be included in the current study. The database used for the current analysis was locked on 20th of June 2016. The study was approved by the appropriate local medical ethical committees. All patients signed the informed consent upon participation.

Imaging scoring procedures

Radiographs and MRIs of the SIJ (X-SIJ; MRI-SIJ) and spine (X-Spine; MRI-Spine) were obtained at baseline, 1, 2 and 5 years. Each image was independently scored, in 3 ‘reading-waves’ by trained central readers, blinded to chronology, clinical data and to the results of other imaging modalities. In wave 1 baseline images were scored by two readers and one adjudicator (in case of disagreement). In wave 2, images from baseline, 1 and 2 years were also scored by 2 readers and one adjudicator. In wave 3, images from baseline, 2 and 5 years were scored by 3 central

readers. Readers and adjudicators varied across modalities and waves (Online Supplementary Table S1).[8] By protocol, radiographs have been performed in all 25 participating centers at each time point, but MRIs were only performed in all centers at baseline, while MRIs at 1, 2 and 5 years were only obtained in 9 centers from Paris.

Inflammation outcomes

Inflammation on MRI-SIJ was assessed using the ASAS definition (positive/negative) and the Spondyloarthritis Research Consortium of Canada (SPARCC) score (range: 0-72).[9-11]

Bone marrow edema (BME) on MRI-Spine was defined according to the ASAS definition (≥3 vertebral corner lesions; yes/no).[12] In addition, a cut-off of 5 vertebral corner BME lesions (typical of axSpA and present in ≥2 consecutive slices) was also assessed, according to the Canada-Denmark method, as it has been shown to be highly specific of axSpA.[13] The total spine SPARCC (range: 0-414) and Berlin (range: 0-69) scores were used as continuous inflammatory outcomes.[3, 14]

Structural outcomes

Structural damage on X-SIJ was assessed according to the modified New York (mNY) system as continuous (range: 0-8) and as a binary (positive / negative) score.[15] Two additional binary definitions were assessed: worsening of ≥1 grade in ≥1 SIJ (yes/no); and worsening of ≥1 grade in ≥1 SIJ, with a 5-year grade ≥2 in the worsened joint (yes/no).[16]

An adaptation of the MRI-SIJ Structural score by Weber *et al*, previously described by us,[17] was used to define individual structural lesions on MRI-SIJ.[18] In summary, fatty lesions, erosions and ankylosis/partial ankylosis are scored as originally described. Sclerosis was added. Fatty lesions, erosions and sclerosis were marked as present if seen on ≥2 consecutive slices (maximum 5 lesions in 6 slices per each of the 8 quadrants in both SIJs). Ankylosis or partial ankylosis was considered present if seen on a single slice. Partial ankylosis and ankylosis cannot occur simultaneously in a quadrant, and ankylosis always involves two quadrants; therefore, the corresponding scoring range is 0–24.In the absence of a formal definition of presence of structural damage on MRI-SIJ, we considered 3 definitions previously shown most discriminatory in early axSpA: ≥5 fatty lesions and/or erosions; ≥3 erosions; and ≥ 3 fatty lesions.[13] Continuous structural lesions on MRI-SIJ were defined as number of fatty lesions and/or erosions (range: 0-80), number of erosions (range: 0-40), number of fatty lesions (range: 0-40) and total number of lesions with (range: 0-144) and without (range: 0-104) sclerosis.

Structural lesions on X-Spine were assessed as the presence of ≥1 syndesmophyte (yes/no), and by the modified Stoke Ankylosing Spondylitis Spinal Score (mSASSS; range: 0-72).[19]

Structural lesions on MRI-Spine were scored according to the Canada–Denmark (CANDEN) method.[20, 21] In the absence of a formal definition, we define structural damage as ≥5 fatty lesions, also previously shown to be highly specific for axSpA.[13] The total number of structural lesions (fatty lesions, erosions, bone spurs, ankylosis; range: 0-322) was assessed, as well as the total number of fatty lesions, erosions and bone spurs (range: 0-92; for all).

A detailed description of all scores is provided in Online Supplementary Tables S2-S10. The interreader reliability of the radiographic and MRI outcomes used in this study has been reported in detail elsewhere and is summarized in Online Supplementary Text S1.[6, 17]

Statistical analysis

The baseline value for each outcome was defined by a 'combination algorithm' of the scores from the 3 readers from wave 3 (agreement between ≥2 out of 3 for binary, and mean of 3 readers for continuous outcomes).

The rate of change of each outcome was analyzed by generalized estimating equations (GEE), with 'time' in years as the explanatory variable of interest. Each outcome was analyzed per patient, per time-point and per individual reader and the yearly rate of change estimated using the so-called 'integrated-analysis', including all patients with ≥1 score from ≥1 reader from ≥1 'reading-wave'. Different to traditional measures of sensitivity to change (e.g. Cohen's effect size), this method, which we have previously explained in detail,[8] appropriately handles the multilevel data structure of our data. All patients had to have ≥1 score from all outcomes, thus ensuring that the same patients are used across all analyses. All variables were standardized. A standardized variable (metric-free) was defined at the patient level as: difference between the individual's value and the population mean divided by the population standard deviation (SD). Each standardized variable has a mean of 0 and a variance of 1 and reads as the number of SD above (positive) or below (negative) the mean.

In addition, the relative standardized rate of change (i.e. the standardized yearly rate of change of an outcome divided by the corresponding rate of a reference imaging outcome) was calculated. For this calculation, a value > 1 means larger sensitivity and a value <1 lower sensitivity compared to the reference (the further away from 1 the larger the difference). Three types of references were defined: i. 'Inflammation common reference': comparing all inflammation outcomes to sacroiliitis on MRI-SIJ (ASAS definition); ii. 'Structural common reference': comparing all structural outcomes to sacroiliitis on X-SIJ (mNY); and iii. 'Modality reference': comparing outcomes to a reference within each modality and anatomical site.

Goodness-of-fit statistics (quasi-likelihood under the independence model criterion; QIC), were used to get an impression on how much of the outcome variability is explained by each model. Different transformations of time were tested to assess which yielded the lowest QIC (better fit). A non-linear model was chosen if best fitting the data, and if the non-linear factor (e.g. quadratic term) added to the model was significant ($p<0.05$). Stata V15.1 was used for the analyses.

RESULTS

Baseline characteristics

In total, 345 patients were included [mean (SD) symptom duration: 1.6 (0.9) years; 53% were males and 89% HLA-B27 positive; Table 1]. Baseline inflammation on MRI was more frequently present at the SIJ (active sacroiliitis: 39%) than at the spine level (BME ≥5 lesions: 6%) (Table 2).

Structural damage at baseline was limited in the SIJ (21% mNY positive) and even more in the spine (≥1 syndesmophyte: 6%) (Table 3).

Sensitivity to change of the different imaging outcomes

Inflammation on MRI-SIJ showed a higher sensitivity to change than on MRI-spine, the latter remaining essentially unchanged over time. This was true for the dichotomous ASAS MRI-SIJ score (standardized yearly rate of change -0.278) and especially for the continuous SPARCC score (standardized yearly rate of change -0.441), while the standardized yearly rates of change for MRI-spine ranged only between -0.030 and -0.055 (Table 2). The differences between SIJ and spine inflammation outcomes become especially evident with the relative standardized rate of change. Compared to the ASAS definition of a positive MRI-SIJ ('inflammation common reference'; i.e. value of 1) all inflammation outcomes in the spine were much less sensitive to change (range of relative standardized rates: 0.094; 0.531; i.e. all values far below 1).

Table 1. Patient- and disease-characteristics at baseline and during follow-up

	Baseline (N=345)	1 year (N=345)	2 years (N=342)	5 years (N=320)
Age at baseline (years), mean (SD)	31.0 (7.0)			
Male gender, n (%)	183 (53)			
Symptoms duration (years), mean (SD)	1.6 (0.9)			
Current smokers*, n (%)	135 (39)	127 (39)	118 (37)	92 (34)
HLA-B27, n (%)	307 (89)			
Radiographic sacroiliitis (mNY)**	73 (21)	NA	68 (23)	68 (27)
BASDAI*, mean (SD) (0-10)	4.1 (2.0)	3.2 (2.2)	3.1 (2.2)	2.9 (2.0)
ASDAS-CRP**, mean (SD)	2.6 (1.0)	2.1 (0.9)	2.0 (0.9)	2.0 (0.9)
Elevated CRP** (≥6 mg/L), n (%)	109 (33)	64 (20)	69 (22)	57 (22)
BASFI*, mean (SD) (0-10)	2.7 (2.2)	2.1 (2.1)	2.1 (2.2)	2.0 (2.0)
TNFi treatment**, n (%)	0 (0)	76 (24)	94 (29)	111 (42)
NSAID treatment*, n (%)	329 (95)	250 (77)	216 (68)	180 (66)

*Missing data <15% in each visit; ** Missing data <20% in each visit. BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; ASDAS, Ankylosing Spondylitis Disease Activity Score; CRP, C-reactive protein; BASFI, Bath Ankylosing Spondylitis Functional Index; TNFi, tumor necrosis factor inhibitors; NSAID, non-steroidal anti-inflammatory drugs; mNY, modified New York criteria (scored in wave 3); NA, not applicable (imaging in wave 3 is only scored at baseline, 2 and 5 years)

Structural damage in the SIJ increased over time but with a larger yearly rate on MRI-SIJ (standardized rate range: 0.015-0.274) compared to X-SIJ (standardized rate range: 0.043-0.126) (Table 3). Three or more fatty lesions on MRI-SIJ was the SIJ structural outcome with highest sensitivity to change (standardized rate: 0.274; relative rate of 6.227 comparing to mNY). On the contrary, ≥3 erosions on MRI-SIJ was the least sensitive (standardized rate: 0.015) of all SIJ structural outcomes (including both MRI-SIJ and X-SIJ). Importantly, ≥3 fatty lesions alone was

slightly more sensitive to change than combining fatty lesions with erosions, i.e. ≥ 5 fatty lesion and/or erosions (relative rate of 1.151 for the former compared to the latter).

Amongst the X-SIJ structural outcomes, worsening of ≥1 grade in ≥1 SIJ and worsening of ≥1 grade in ≥1 SIJ, with a 5-year grade ≥2 in the worsened joint were far more sensitive to change compared to the mNY binary definition as the 'modality reference' (relative rate: 2.864 and 2.705, respectively). Of note, the mNY continuous grading and the mNY binary score had comparable sensitivity to change (relative rate of the continuous vs the reference binary score = 0.977).

Overall, the standardized yearly rate of change of the spinal radiographic outcomes (range: 0.037-0.043) was higher as compared to MRI-Spine structural outcomes (range: 0.012-0.027) (Table 3), although all are relatively low. Amongst MRI-Spine outcomes, the total number of bone spurs was the outcome that most captured change (standardized rate: 0.027; and relative rate of 2.077 compared to ≥ 5 fatty lesions – the 'modality reference'). Yet, the best MRI-Spine outcome is still less sensitive to change as compared to X-spine outcomes, with a standardized rate of 0.037 for ≥ 1 syndesmophyte and of 0.043 for the continuous mSASSS.

Table 2. Baseline score and standardized yearly rate of change of inflammatory imaging outcomes over 5 years of follow-up in early axSpA patients fulfilling the ASAS axSpA classification criteria

Imaging outcomes	Baseline score* (N=334-344)	Standardized rate of change/year¥	Relative sRoC (Common Reference: ASAS MRI-SIJ)	Relative sRoC per modality & anatomical site
Inflammatory lesions (MRI-SIJ)[9-11]				
Sacroiliitis (ASAS criteria)	134 (39.2%)	-0.278£	1	1
SPARCC SIJ score (0-72)	4.7 (7.9)	-0.441£	1.586	1.586
Inflammatory lesions (MRI-Spine)[3,12-14]				
BME: ≥ 3 lesions	32 (9.4%)	-0.032	0.319	1
BME: ≥ 5 lesions	19 (5.6%)	-0.030	0.094	0.938
23 DVU SPARCC Spine score (0-414)	2.6 (7.7)	-0.050	0.531	1.563
Berlin Spine score (0-69)	0.9 (2.7)	-0.055	0.104	1.719

* Agreement of ≥2 out of 3 readers for binary variables and mean (SD) of 3 readers for continuous variables from wave 3; ¥ Estimated from a model where all independent variables (time, reader and wave) and the outcome are standardized; £ Quadratic transformation led to a better model goodness of fit (QIC: quasi-likelihood under the independence model criterion); ASAS, Assessment of SpondyloArthritis international Society; BME, bone marrow edema; MRI, magnetic resonance imaging; SIJ, sacroiliac joints; SPARCC, spondyloarthritis research consortium of Canada; DVU, discovertebral unit; sRoC, standardized rate of change.

Table 3. Baseline score and standardized yearly rate of change structural imaging outcomes over 5 years of follow-up in early axSpA patients fulfilling the ASAS axSpA classification criteria

Imaging outcomes	Baseline score* (N=313-344)	Standardized rate of change/year¥	Relative sRoC (Common Reference: mNY)	Relative sRoC per modality and anatomical site
Structural lesions (X-SIJ)[15, 16]				
mNY dichotomous	73 (21.2%)	0.044	1	1
mNY 1-grade change**	NA	0.126	2.864	2.864
mNY 1-grade change and value ≥ 2††	NA	0.119	2.705	2.705
mNY continuous grade (0-8)	1.7 (1.8)	0.043	0.977	0.977
Structural lesions (MRI-SIJ)[17]				
≥ 5 fatty lesion and/or erosions	66 (19.5%)	0.238£	5.409	1
≥ 3 erosions	60 (17.7%)	0.015	0.341	0.063
≥ 3 fatty lesions	56 (16.5%)	0.274£	6.227	1.151
Number of fatty lesions/erosions (0-80)	2.9 (4.9)	0.111	2.523	0.466
Number of erosions (0-40)	1.3 (2.2)	0.030	0.682	0.126
Number of fatty lesions (0-40)	1.5 (3.5)	0.140	3.182	0.588
Total structural lesions† (0-144)	3.4 (5.9)	0.115	2.614	0.483
Total structural lesions no sclerosis (0-104)	3.2 (5.8)	0.124	2.818	0.521
Structural lesions (X-Spine)[18]				
≥ 1 syndesmophyte	19 (5.5%)	0.037	0.841	1
mSASSS score (0-72)	0.3 (1.3)	0.043	0.977	1.162
Structural lesions (MRI-Spine)[19, 20]				
≥ 5 fatty lesions	5 (1.6%)	-0.013	0.295	1
Total structural lesions‡ (0-322)	0.4 (1.0)	0.016	0.364	1.231
Number of fatty lesions (0-92)	0.3 (0.8)	0.008	0.182	0.615
Number of corner erosions (0-92)	0.1 (0.2)	0.012	0.273	0.923
Number of corner bone spurs (0-92)	0.1 (0.3)	0.027	0.614	2.077

* Agreement of ≥2 out of 3 readers for binary variables and mean (SD) of 3 readers for continuous variables from wave 3; ¥ Estimated from a model where all independent variables (time, reader and wave) and the outcome are standardized; † fatty lesions, erosions, sclerosis, partial ankylosis/total ankylosis; ** Change of at least one grade in at least one sacroiliac joint (SIJ); †† Change of at least one grade in at least one SIJ, but with a 5-year grade ≥ 2 in the worsened joint; £ Quadratic transformation led to a better model goodness of fit (QIC: quasi-likelihood under the independence model criterion); NA, not applicable; ASAS, Assessment of SpondyloArthritis international Society; MRI, magnetic resonance imaging; X, radiograph; SIJ, sacroiliac joints; mSASSS, modified Stoke Ankylosing Spondylitis Spinal Score; sRoC, standardized rate of change.

DISCUSSION

In this prospective observational study, we have shown that, in patients with early axSpA, MRI outcomes of inflammation are more sensitive to change in the SIJ than in the spine. In addition, pelvic radiographs yield low sensitivity to change in detecting structural damage, while fatty lesions detected on MRI-SIJ emerges as a promising alternative. In contrast, MRI-spine is not better than X-spine in detecting structural changes in early axSpA patients.

In the current study, we directly compared, for the first time, inflammation outcomes on MRI-SIJ and MRI-spine and have shown that the former are more sensitive to change. Inflammation on MRI-spine remained low and essentially unchanged over a period of 5 years. Different from previous studies evaluating the sensitivity to change of imaging outcomes over shorter periods, we have applied an analytical technique ('integrated analysis') that we have previously shown to be robust for the evaluation of change over long periods of follow-up, especially with outcomes that are expected to occur infrequently over time.[8] Of note, combination algorithms (e.g. agreement between 2 out of 3 readers) are not needed when using this method. Instead

each individual reader score is analysed as it is in an assumption-free manner which, to some extent, handles across-reader variability.

The ASAS/OMERACT MRI working group has previously compared different (continuous) scores to quantify inflammation on MRI-SIJ.[22] In a multi-reader exercise the SPARCC method has been shown to be the most reliable and sensitive to change among patients with r-axSpA. The current study adds to this data by showing that both the continuous SPARCC score and the binary ASAS definition of a positive MRI-SIJ yield good sensitivity to change in the entire spectrum of axSpA (including nr-axSpA) during the early phases of the disease.

The same group performed a similar exercise for MRI-spine (also in r-axSpA).[3] This experiment has shown discrepant reliability results for the comparison between the 6- discovertebral unit (DVU) SPARCC score, the Ankylosing Spondylitis spine MRI activity (ASspiMRI-a) score and the Berlin method (SPARCC performed better when using the intraclass correlation coefficient but worse when using the smallest detectable change). All methods yielded excellent sensitivity to change according to the Guyatt's effect-size. Here, we compared the 23-DVU SPARCC to the Berlin method and 2 binary outcomes and found that all yield very poor sensitivity to change. Of note, these studies differ in several aspects, including the reading methods and population. In fact, our early axSpA population had lower baseline levels of inflammation compared to patients from the ASAS/OMERACT exercise (mean (SD) Berlin: 0.9 (2.7) vs 6 (9.0), respectively), which may hinder the detection of change, that we have shown before to be small in early axSpA.[17] Of note, in patients with nr-axSpA and high disease activity selected for RCTs, inflammation on MRI-spine performed well both in terms of sensitivity to change and in discriminating response between treatment arms.[23, 24] This confirms that the ability of the scoring methods to detect change is not only dependent on their intrinsic characteristics, but also on the population in which they are applied.

A recent study, also from DESIR, has shown that 'net' progression from mNY-negative to mNY-positive (i.e. considering measurement error) is very limited.[16] In the current study we have additionally shown that the change in the mNY (continuous) grading is as poorly sensitive to change as the mNY binary score (relative rate ≈ 1). On the other hand, the change in at least 1-grade in at least one SIJ, with or without considering the change between grade 0 and grade 1, perform better in detecting change.[16, 25]

Information on the sensitivity to change of MRI-SIJ structural outcomes is very scarce.[26] To the best of our knowledge, no previous formal comparison with X-SIJ scores has been performed thus far. We have found that ≥3 fatty lesions on MRI-SIJ largely outperform all X-SIJ outcomes. Erosions, however, performed poorly in this early population. Thus, our study yields encouraging data supporting MRI (in particular fatty lesions) as an alternative to radiographs in detecting change of structural damage at the SIJ. In contrast, in the spine, we found no evidence that MRI is better than radiographs in detecting change of structural damage. Despite the disappointing results with MRI, our results are in line with previous studies, showing that spinal radiographic progression can be detected even in early phases of the disease.[4, 27] A recent study has shown that low dose computerized tomography of the spine is more sensitive to detect new syndesmophytes than conventional radiographs promising to further expand our ability to detect change in axial damage.[28]

Our study has some limitations. First, not all available scoring systems were assessed. However, to the best of our knowledge, this is so far the largest direct comparison across scores, which includes those currently more often used in research and clinical practice. Second, we did not assess all domains of the OMERACT filter, namely validity, reliability and feasibility.[2] Thus, we cannot, and do not claim to, evoke superiority of one score over others based on our data alone. Instead, our results should be interpreted in light of the literature already informing on these aspects but falling short on direct comparisons of sensitivity to change. Third, the observed levels of inflammation, structural damage and changes over time are limited in this cohort, especially in the spine, which reduces the possibility to detect differences across methods. Finally, our data are limited to patients with early axSpA, thus our findings cannot be generalized to all patients with axSpA from clinical practice especially those with more advanced disease (i.e. with r-axSpA).

In conclusion, we have shown that MRI inflammation scores are more sensitive to change in the SIJ than in the spine. Also, X-SIJ structural outcomes are less sensitive to change compared to fatty lesions on MRI-SIJ. In contrast, MRI-spine is no better than X-spine in detecting structural changes in this early axSpA cohort. These data may help in prioritizing imaging scoring methods in subsequent observational or interventional studies in early axSpA.

SUPPLEMENTARY DATA

Supplementary data are published online on the website of Arthritis Care & Research

REFERENCES

1. Mandl P, Navarro-Compan V, Terslev L, et al. EULAR recommendations for the use of imaging in the diagnosis and management of spondyloarthritis in clinical practice. Ann Rheum Dis. 2015 Jul; 74(7):1327-1339.
2. Boers M, Kirwan JR, Wells G, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol. 2014 Jul; 67(7):745-753.
3. Lukas C, Braun J, van der Heijde D, et al. Scoring inflammatory activity of the spine by magnetic resonance imaging in ankylosing spondylitis: a multireader experiment. J Rheumatol. 2007 Apr; 34(4):862-870.
4. Ramiro S, van Tubergen A, Stolwijk C, et al. Scoring radiographic progression in ankylosing spondylitis: should we use the modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) or the Radiographic Ankylosing Spondylitis Spinal Score (RASSS)? Arthritis Res Ther. 2013 Jan 17; 15(1):R14.
5. Wanders AJ, Landewé RB, Spoorenberg A, et al. What is the most appropriate radiologic scoring method for ankylosing spondylitis? A comparison of the available methods based on the Outcome Measures in Rheumatology Clinical Trials filter. Arthritis Rheum. 2004 Aug; 50(8):2622-2632.
6. Ramiro S, Claudepierre P, Sepriano A, et al. Which scoring method depicts spinal radiographic damage in early axial spondyloarthritis best? Five-year results from the DESIR cohort. Rheumatology (Oxford). 2018 Jul 23.
7. Dougados M, d'Agostino MA, Benessiano J, et al. The DESIR cohort: a 10-year follow-up of early inflammatory back pain in France: study design and baseline characteristics of the 708 recruited patients. Joint Bone Spine. 2011 Dec; 78(6):598-603.
8. Sepriano A, , Ramiro S, van der Heijde D, Dougados M, Claudepierre P, Feydy A, Reijnierse M, Loeuille D, Landewé R. Integrated longitudinal analysis does not compromise precision and reduces bias in the study of imaging outcomes: A comparative 5-year analysis in the DESIR cohort. Semin Arthritis Rheum. 2020 Mar 3. pii: S0049-0172(20)30070-6.
9. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis research Consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. Arthritis Rheum. 2005 Oct 15; 53(5):703-709.
10. Rudwaleit M, Jurik AG, Hermann KG, 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 Oct; 68(10):1520-1527.

11. Lambert RG, Bakker PA, van der Heijde D, et al. Defining active sacroiliitis on MRI for classification of axial spondyloarthritis: update by the ASAS MRI working group. Ann Rheum Dis. 2016 Nov; 75(11):1958-1963.

12. Hermann KG, Baraliakos X, van der Heijde DM, et al. Descriptions of spinal MRI lesions and definition of a positive MRI of the spine in axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI study group. Ann Rheum Dis. 2012 Aug; 71(8):1278-1288.

13. de Hooge M, van den Berg R, Navarro-Compan V, et al. Patients with chronic back pain of short duration from the SPACE cohort: which MRI structural lesions in the sacroiliac joints and inflammatory and structural lesions in the spine are most specific for axial spondyloarthritis? Ann Rheum Dis. 2016 Jul; 75(7):1308-1314.

14. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. Arthritis Rheum. 2005 Aug 15; 53(4):502-509.

15. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum. 1984 Apr; 27(4):361-368.

16. Dougados M, Sepriano A, Molto A, et al. Sacroiliac radiographic progression in recent onset axial spondyloarthritis: the 5-year data of the DESIR cohort. Ann Rheum Dis. 2017 Jul 06.

17. Madari Q, Sepriano A, Ramiro S, et al. 5-year follow-up of spinal and sacroiliac MRI abnormalities in early axial spondyloarthritis: data from the DESIR cohort. RMD Open. 2020 Feb; 6(1).

18. Weber U, Lambert RG, Ostergaard M, et al. The diagnostic utility of magnetic resonance imaging in spondylarthritis: an international multicenter evaluation of one hundred eighty-seven subjects. Arthritis Rheum. 2010 Oct; 62(10):3048-3058.

19. Creemers MC, Franssen MJ, van't Hof MA, et al. Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system. Ann Rheum Dis. 2005 Jan; 64(1):127-129.

20. Ostergaard M, Maksymowych WP, Pedersen SJ, et al. Structural lesions detected by magnetic resonance imaging in the spine of patients with spondyloarthritis - Definitions, assessment system, and reference image set. Journal of Rheumatology. 2009; 36(SUPPL. 84):18-34.

21. Krabbe S, Sorensen IJ, Jensen B, et al. Inflammatory and structural changes in vertebral bodies and posterior elements of the spine in axial spondyloarthritis: construct validity, responsiveness and discriminatory ability of the anatomy-based CANDEN scoring system in a randomised placebo-controlled trial. RMD Open. 2018; 4(1):e000624.

22. van der Heijde DM, Landewé RB, Hermann KG, et al. Application of the OMERACT filter to scoring methods for magnetic resonance imaging of the sacroiliac joints and the spine. Recommendations for a research agenda at OMERACT 7. J Rheumatol. 2005 Oct; 32(10):2042-2047.

23. Maksymowych WP, Dougados M, van der Heijde D, et al. Clinical and MRI responses to etanercept in early non-radiographic axial spondyloarthritis: 48-week results from the EMBARK study. Ann Rheum Dis. 2016 Jul; 75(7):1328-1335.

24. Sieper J, van der Heijde D, Dougados M, 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 Jun; 72(6):815-822.

25. Poddubnyy D, Rudwaleit M, Haibel H, et al. Rates and predictors of radiographic sacroiliitis progression over 2 years in patients with axial spondyloarthritis. Ann Rheum Dis. 2011 Aug; 70(8):1369-1374.

26. Baraliakos X, van der Heijde D, Braun J, et al. OMERACT magnetic resonance imaging initiative on structural and inflammatory lesions in ankylosing spondylitis--report of a special interest group at OMERACT 10 on sacroiliac joint and spine lesions. J Rheumatol. 2011 Sep; 38(9):2051-2054.

27. Cho SK, Sakai R, Nanki T, et al. A comparison of incidence and risk factors for serious adverse events in rheumatoid arthritis patients with etanercept or adalimumab in Korea and Japan. Modern Rheumatology. 2014 Jul; 24(4):572-579.

28. de Koning A, de Bruin F, van den Berg R, et al. Low-dose CT detects more progression of bone formation in comparison to conventional radiography in patients with ankylosing spondylitis: results from the SIAS cohort. Ann Rheum Dis. 2018 Feb; 77(2):293-299.