

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/29572> holds various files of this Leiden University dissertation.

Author: Berg, Rosaline van den

Title: Spondyloarthritis : recognition, imaging, treatment

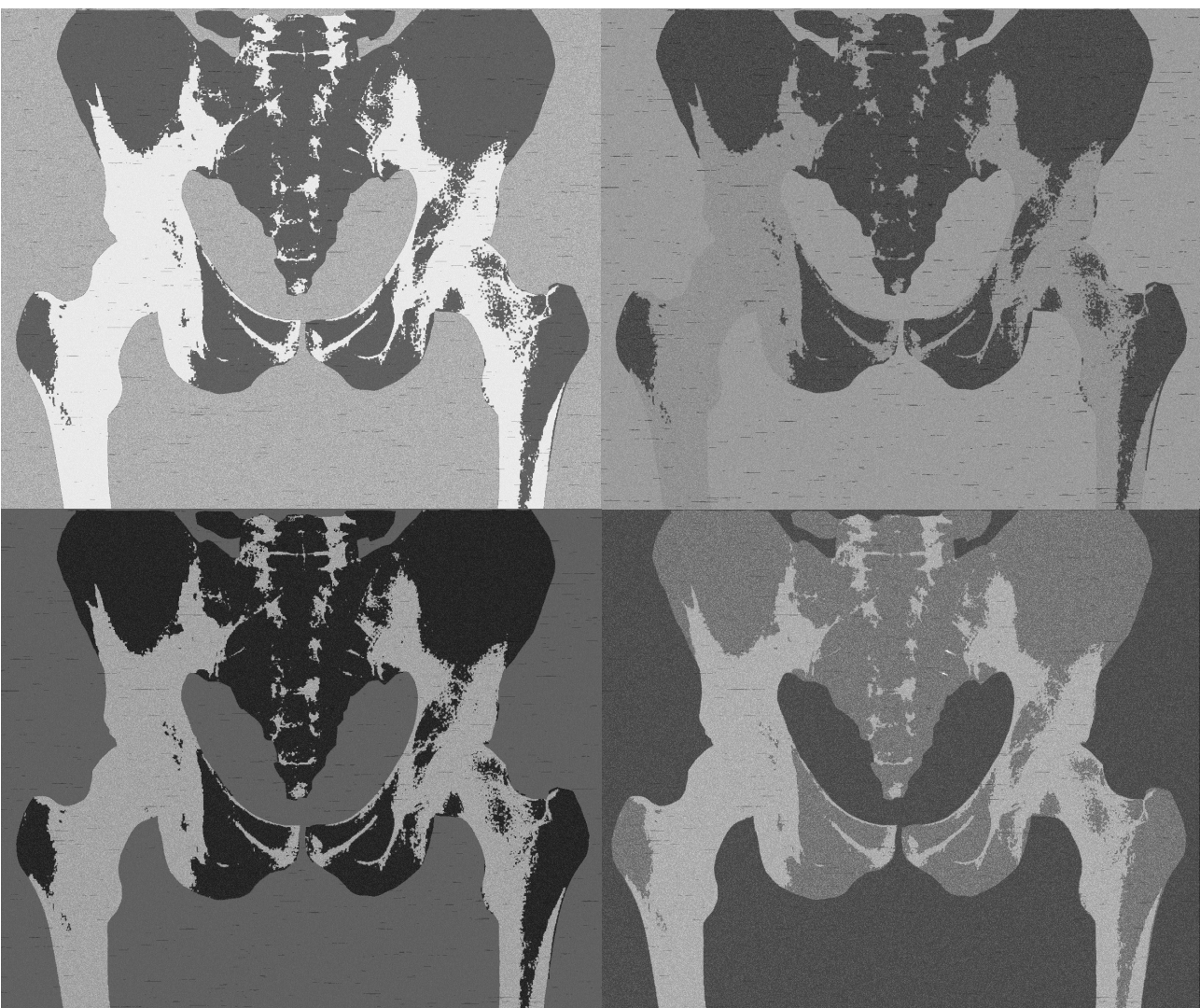
Issue Date: 2014-10-29

Classification of axial SpA based on positive imaging (radiographs and/or MRI of the sacroiliac joints) by local rheumatologists or radiologists versus central trained readers in the DESIR-cohort

Ann Rheum Dis Published Online First: [24 June 2014]

doi:10.1136/annrheumdis-2014-205432

R. van den Berg
G. Lenczner
F. Thévenin
P. Claudepierre
A. Feydy
M. Reijnierse
A. Saraux
A. Rahmouni
M. Dougados
D. van der Heijde



ABSTRACT

Objective

Investigating changes in patient classification (ASAS axSpA criteria) based on evaluation of images of the sacroiliac joints (MRI-SI and X-SI) by local and central readers.

Methods

The DESIR-cohort included patients with inflammatory back pain (IBP; ≥ 3 months, but < 3 years), suggestive of axSpA. Local radiologists/rheumatologists (local reading) and two central readers (central reading) evaluated baseline images. Agreement regarding positive MRI (pos-MRI) between central readers and between local reading and central reading was calculated (kappas). Number of patients classified differently (ASAS criteria) by using local reading instead of central reading was calculated.

Results

Interreader agreement between the two central readers and between local reading and central reading was substantial ($\kappa=0.73$ and $\kappa=0.70$, respectively). In 89/663 MRI-SIs (13.4%) local reading and central reading disagreed; 38/223 patients (17.0%) with pos-MRI (local reading) were negative by central reading; 51/440 patients (11.6%) with neg-MRI (local reading) were positive by central reading.

In 163/582 patients eligible for applying ASAS criteria (28.0%), local reading and central reading disagreed on positive imaging (MRI-SI and/or X-SI; $\kappa=0.68$). In 46/582 patients (7.9%) a different evaluation resulted in a different classification; 18/582 patients (3.1%) classified no SpA (central reading) were axSpA by local reading; 28/582 patients (4.8%) classified axSpA (central reading) were no SpA by local reading. Among axSpA patients (central reading), 16/419 patients (3.8%) fulfilling imaging arm by central reading fulfilled clinical arm by local reading; 29/419 patients (6.9%) fulfilling clinical arm by central reading fulfilled also imaging arm by local reading.

Conclusion

In patients with recent onset IBP, trained readers and local rheumatologists/radiologists agree well on recognizing a pos-MRI. While disagreeing in 28% of the patients on positive imaging (MRI-SI and/or X-SI), classification of only 7.9% of the patients changed based on a different evaluation of images, showing the ASAS axSpA criteria's robustness.

INTRODUCTION

The 2009 classification criteria for axial spondyloarthritis (axSpA) by the Assessment of SpondyloArthritis international Society (ASAS) are gaining more awareness and are increasingly being used to guide daily practice and include patients in clinical trials ¹⁻³. According to the ASAS axSpA criteria it is possible to classify patients with chronic back pain as axSpA via the clinical arm based on the presence of at least two SpA-features in addition to HLA-B27 positivity, or to classify patients via the imaging arm. In the presence of sacroiliitis on plain radiographs (modified New York (mNY) criteria) and/or MRI (ASAS definition of a positive MRI (pos-MRI)), a patient can be classified as axSpA if at least one additional SpA-feature should be present ^{1,4,5}. However, recognition of sacroiliitis on MRI and especially on plain radiographs is challenging ⁶⁻⁸. It is known that interpretation of findings vary according to the expertise of the physician interpreting the image ⁷. In daily practice, local radiologists and/or rheumatologists judge MRIs and radiographs of the SI-joints, frequently with knowledge of the clinical signs and symptoms, while in research cohorts and clinical trials ≥ 1 trained reader - blinded for clinical information - judge the images. As the classification as axSpA is heavily based on sacroiliitis, the classification of a patient could change as another reader judges the same MRI and/or radiographs differently. The ABILITY-1 trial included patients with non-radiographic axSpA (nr-axSpA), based on readings of the pelvic radiographs by local radiologists or rheumatologists. A post-hoc central reading (for another purpose) was performed and based on this reading, 37% of the patients classified as nr-axSpA by local sites were reclassified as fulfilling the mNY criteria ^{3,9}. In another trial, the RAPID-axSpA trial, a similar analysis was performed resulting in reclassification of 36% of the patients (26% reclassified as fulfilling the mNY criteria, and 10% reclassified as nr-axSpA, based on the central reading in contrast to the local reading) ^{2,10}.

As sacroiliitis by two imaging methods as well as HLA-B27 positivity play an important role in the ASAS axSpA criteria, a patient will not necessarily be classified differently based on another reading of the radiograph and/or MRI of the SI-joints. Therefore, we investigated the change in classification of patients according to the ASAS axSpA criteria based on the evaluation of local and central readers of the same set of images. We performed this investigation in the DESIR (DEvenir des Spondylarthropathies Indifférenciées Récentes)-cohort, which has information on MRIs and radiographs of the SI-joints scored by the local rheumatologist or radiologist and also by two trained central readers.

METHODS

Patients

Baseline data from the DESIR-cohort was used for this analysis. The DESIR-cohort is described extensively before ¹¹. In short, consecutive patients aged 18-50 with Inflammatory Back Pain (IBP) in the thoracic and/or lumbar spine and/or the buttock area (≥ 3 months, but < 3 years) fulfilling either the Calin (4/5 criteria) or the Berlin (2/4 criteria) for IBP and a suspicion of SpA by the rheumatologist with a score of ≥ 5 on a scale of 0 to 10 (where 0 was not suggestive of axSpA and 10 was very suggestive of axSpA) from 25 centers in France were included in this prospective longitudinal cohort ^{12,13}. In total, 708 patients were included between December 2007 and April 2010. The study is approved by the appropriate medical ethical committee and fulfilled Good Clinical Practice Guidelines. Before patients were included in the study, they gave written informed consent.

A detailed description of the study protocol is available at the website (<http://www.lacohortedesir.fr/desir-in-english/>). The research proposal for this particular analysis was approved by the scientific committee of the DESIR-cohort.

Data collection

With the use of a standardized CRF a database was built. According to the DESIR protocol the following data, among others, were collected: physical examination, ongoing treatment, co-morbidities, laboratory tests and questionnaires¹¹. The database for the baseline data used for this analysis was locked on October 30th 2012.

Images and scoring methods

In each participating center MRIs of the SI-joints (MRI-SIs) were performed at baseline, with magnetic fields between 1.0 and 1.5T, using T1-FSE and STIR sequences with 12-15 semi-coronal slices of 4 mm thickness, parallel to the long axis of the sacrum, without the use of a contrast agent. All initial MRI-SIs were checked on quality by a central reader in Montpellier, and regular calibration by the manufacturer was required. Plain radiographs of the pelvis (X-SI) were performed in anteroposterior view at baseline.

All available baseline MRI-SIs (n=663) were scored by a local radiologist/rheumatologist who might have had access to all clinical and laboratory data at each participating center (local reading)¹⁴. Each SI-joint on MRI was assessed on the presence/absence of inflammation by answering the following question on the CRF '*Are there characteristic acute/active inflammatory lesions compatible with axial spondyloarthritis of the sacroiliac joints or entheses, outside the sacroiliac joints? Normal (score 0), doubtful (score 1) or abnormal (score 2).*' On the CRF, inflammatory lesions were defined as 'Bone edema/contrast product uptake in or adjacent to the sacroiliac joints or entheses (compatible with active lesions observed in cases of ankylosis spondylitis/axial spondyloarthritis; STIR and/or T1 sequences with gadolinium injection are required)'. In this reading, a pos-MRI was defined as a score of 2 in at least one of the SI-joints.

Two central readers (RvdB and FT), experienced in scoring MRI-SIs, participated in a calibration training on reading MRI-SIs according to the ASAS definition. MRI-SIs were considered positive according to the ASAS definition if BME lesions highly suggestive of SpA were present if ≥ 1 BME lesion on ≥ 2 consecutive slices, or if several BME lesions are visible on a single slice. The presence of only synovitis, enthesitis or capsulitis without BME is not sufficient for a positive MRI-SI⁵. During the calibration session, executed by two senior radiologists (MR and AF) and two senior rheumatologists (PC and MD), supervised by an expert in AS and imaging scoring (DvdH), definitions of lesions, examples and pitfalls were discussed, followed by a supervised reading of training cases by the two readers. After this calibration session, 30 blinded MRI-SIs were read independently by the two readers ($\kappa=0.30$; positive agreement 73.7%; negative agreement 54.5%). A consensus meeting followed with the same group. Six weeks later, a second set consisting of 20 blinded MRI-SIs, were read independently by the two readers, again followed by a consensus meeting with the same group. After this second training session, agreement between the two readers was considered sufficient so the readers could start reading the DESIR-cohort ($\kappa=0.74$; positive agreement 80.0%; negative agreement 93.3%).

All available baseline MRI-SIs were read independently by the two readers, blinded for all clinical and laboratory data, the other imaging modality, as well as the local readings. Agreement on presence/absence of a pos-MRI was calculated and in case of disagreement, one of the senior radiologists involved in the calibration session (MR) served as adjudicator and scored the MRI-SI blinded to the information of the primary readers. An image was marked as pos-MRI (central reading) if 2/3 readers agreed.

The evaluation of X-SIs (n=688) by both local readers and central readers has been described before⁸. In short, the calibration of the two central readers (RvdB and GL) was performed in a similar way as for MRI-SI. Based on the mNY criteria, sacroiliitis was defined as grade ≥ 2 bilaterally or grade 3-4 unilaterally by central reading (pos-X-SI)⁴. The local readers evaluated X-SIs according to a method derived from the mNY. Since the local readers, who are working

in regular clinical practice, were not trained experts it was considered more appropriate use a scoring system that better resembles common clinical practice than the mNY criteria do. Local readers were asked to rate each SI-joint either as ‘normal’ or as ‘doubtful sacroiliitis’ or as ‘obvious sacroiliitis’ or as ‘SI-joint fusion’. In this analysis at least a unilateral rating of ‘obvious sacroiliitis’ was considered a pos-X-SI for local reading. This has been explained in more detail before ⁸.

Statistical analysis

Agreement was calculated using cross-tabulation expressed in Cohen’s Kappa (κ), agreement on positive cases (positive agreement) and on negative cases (negative agreement) for the following comparisons (online supplementary text 1) ¹⁵⁻¹⁸: interreader agreement between the two central readers, agreement between local reading and central reading and between local reading and the two individual central readers on the presence/absence of a pos-MRI. Central reading was considered the external standard.

Next, the number of patients with a different MRI-SI and/or X-SI read using local reading instead of central reading was calculated, followed by the number of patients classified differently according to the ASAS axSpA criteria. This was done both for overall fulfillment and fulfillment of the imaging versus clinical arm.

SPSS software version 20.0 was used for the statistical analysis.

RESULTS

The mean age of patients with available MRI-SI (n=663) was 31.7 (SD 8.7) years, mean symptom duration was 17.8 (SD 10.5) months, 309 (46.6%) patients were men and 387 (58.4%) were HLA-B27 positive.

Finally, in 15 patients X-SI was missing resulting in 648 patients with complete imaging. In 66/648 patients with complete imaging, IBP onset was >45 years and therefore the ASAS axSpA criteria could not be applied, leaving 582 patients (figure 1). Patient characteristics of these 582 patients were very similar to the patients with complete MRI-SI; mean age was 31.5 (SD 7.2) years, mean symptom duration was 18.3 (SD 10.6) months, 277 (47.7%) patients were men and 350 (60.1%) were HLA-B27 positive.

Agreement on a positive MRI

Interreader agreement between the two central readers regarding a pos-MRI is substantial ($\kappa=0.73$; table 1); 84/663 MRI-SIs (12.7%) were adjudicated because of disagreement.

Table 1: Interreader agreement between central reader 1 and central reader 2

Central reader 1	Total n=663	Central reader 2	
		MRI-SI+ (ASAS)	MRI-SI- (ASAS)
MRI-SI+ (ASAS)	200	56	
MRI-SI - (ASAS)	28		379
κ (95% CI): 0.73 (0.67-0.78)		Positive agreement: 82.6%	Negative agreement: 90.0%

Positive agreement is the agreement on positive cases. Negative agreement is the agreement on negative cases.

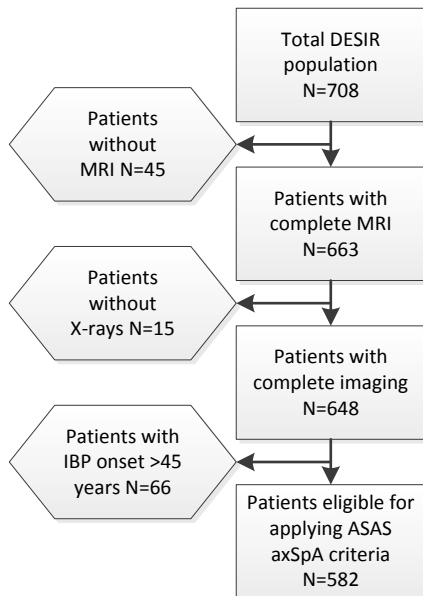


Figure 1: Flowchart of patients included in the DESIR-cohort and included in this analysis.

According to central reading, 236/663 patients (35.6%) had a pos-MRI; according to local reading, 33.6% had a pos-MRI. Agreement between local reading and central reading was also substantial ($\kappa=0.70$). In 13.4% of the MRI-SIs, local reading and central reading disagreed; 38/223 patients (17.0%) with a pos-MRI according to local reading, were read negative by central reading; 51/440 patients (11.6%) without a pos-MRI according to local reading, were read positive by central reading (online supplementary text 2). Comparisons of local reading versus the individual central readers show very similar results (table 2). There was no difference in agreement between local reading and central reading if MRI-SIs were read by local rheumatologists ($n=174$) or by local radiologists ($n=457$) ($n=32$ read by both a radiologist and a rheumatologist; data not shown).

Classification of patients according to the ASAS axSpA criteria

In this paragraph we focus only on the 582 patients in which the ASAS axSpA criteria could be applied. In 28.0% of the patients there was a disagreement on pos-imaging, MRI-SI and/or X-SI ($\kappa=0.68$). In 15.6% of the patients the disagreement was caused by a different X-SI read only (agreement on MRI-SI); in 10.1% the read of MRI-SI was different only (agreement on X-SI); and in 2.2% both X-SI and MRI-SI were read differently.

In total, 409 patients (70.2%) fulfilled the ASAS axSpA criteria based on local reading and 419 patients (72.0%) based on central reading. In 7.9% of the patients a different evaluation of imaging (MRI-SI and/or X-SI) resulted in a different classification. Eighteen patients were classified no SpA based on central reading but were classified axSpA based on local reading; in 28 patients it was the other way around (figure 2). In 14/18 and 13/28 patients, respectively, the different classification was the result of a different X-SI evaluation, consequently these patients changed from AS to no SpA and vice versa (table 3). The results of the comparison of local reading versus the individual central readers are similar (online supplementary table S2).

Table 2: Agreement between local reading and central reading, and between local reading and the individual central readers regarding presence/absence of sacroiliitis on MRI.

Local reading	Total n=663	Central reading (2/3)	
		MRI-SI+ (ASAS)	MRI-SI- (ASAS)
MRI-SI+		185	38
MRI-SI -		51	389
κ (95% CI): 0.70 (0.65-0.76)		Positive agreement: 80.6%	Negative agreement: 89.7%
Local reading	Total n=663	Central reader 1	
		MRI-SI+ (ASAS)	MRI-SI- (ASAS)
MRI-SI+		180	43
MRI-SI -		76	364
κ (95% CI): 0.61 (0.55-0.67)		Positive agreement: 75.2%	Negative agreement: 86.0%
Local reading	Total n=663	Central reader 2	
		MRI-SI+ (ASAS)	MRI-SI- (ASAS)
MRI-SI+		177	46
MRI-SI -		51	389
κ (95% CI): 0.67 (0.62-0.73)		Positive agreement: 78.5%	Negative agreement: 88.9%

Positive agreement is the agreement on positive cases. Negative agreement is the agreement on negative cases.

Additional discrepancies were seen when interested in whether patients fulfill the imaging arm or the clinical arm within the ASAS axSpA criteria. By definition, patients fulfilling the clinical arm will always fulfill the clinical arm as HLA-B27 status will not change, but could fulfill the imaging arm as well, or not anymore, if a different evaluation of the same imaging set is used. Among the patients classified as axSpA based on central reading (n=419), 16 axSpA patients fulfilled the imaging arm based on central reading but fulfilled the clinical arm only based on local reading (in 8 patients due to a different X-SI read) (figure 2). When solely interested in whether patients fulfilled the imaging arm of the ASAS axSpA criteria or not, 44 patients fulfilled the imaging arm by central reading but not by local reading. Vice versa, 29 axSpA patients fulfilled the clinical arm only based on central reading but fulfilled the imaging arm based on local reading (in 13 patients due to a different X-SI read). Again, when interested in whether patients fulfill the imaging arm or not, 47 patients fulfilled the imaging arm by local reading but not by central reading (table 3). Comparisons of local reading versus the individual readers show similar results (online supplementary table S1).

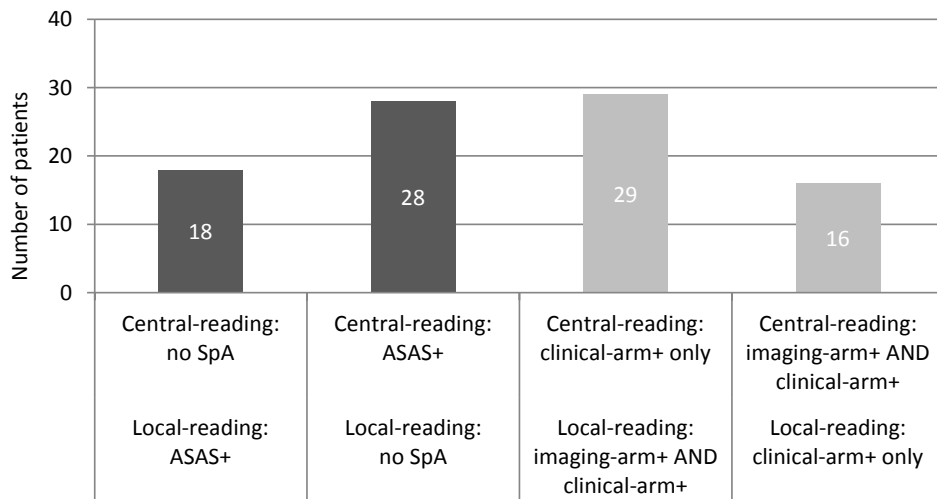


Figure 2: Bar graphs representing patients with SpA (ASAS+) according to one reading (local or central) but without SpA (dark grey) or a different arm of the ASAS criteria (light grey) according to the other reading (local or central) (total n=582, in 163 patients the imaging read was different between local and central reading).

Table 3: Classification of patients according to the ASAS axSpA criteria using Local reading instead of Central reading or the individual central readers.

Patients in which local and central reading (2/3 readers) differed, n=163		Central reading (2/3 readers)						Clinical arm+	No SpA (ASAS-)	
		Both arms +			Imaging arm+					
		mNY+ MRI+	mNY+ MRI-	mNY- MRI+	mNY+ MRI+	mNY+ MRI-	mNY- MRI+			
Local reading	Both arms +	mNY+ MRI+	0	4	29			3		
		mNY+ MRI-	8	0	0			10		
		mNY- MRI+	8	1	0			16		
	Imaging arm+	mNY+ MRI+				0	1	13		4
		mNY+ MRI-				3	0	1		10
		mNY- MRI+				4	0	0		4
	Clinical arm+	1	7	8						
No SpA (ASAS-)				3	10	15				

Both arms; patients fulfil both the imaging arm and the clinical arm. Imaging arm; patients fulfil the imaging arm only. Clinical arm; patients fulfil the clinical arm only. MRI+; sacroiliitis on MRI. mNY+; sacroiliitis on radiograph. Boxes in grey are empty as a patient fulfilling the clinical arm by definition will always fulfil the clinical arm as HLA-B27 status will not change, regardless of a different reading of images.

DISCUSSION

In the DESIR-cohort, agreement between two trained central readers as well as between central reading and local reading on pos-MRI was substantial, thereby comparable to levels of agreement reported in a study designed to test inter- and intrareader agreement between experienced radiologists on a pos-MRI ($\kappa=0.79-0.85$)¹⁹. Though, it should be noted that at the start of the DESIR-cohort, the ASAS definition of a positive MRI-SI was not published yet. The levels of agreement of pos-MRI in the DESIR-cohort were higher than levels of agreement on pos-X-SI in the same cohort ($\kappa=0.46-0.55$). In addition, where misclassification by local reading regarding X-SIs almost exclusively consisted of overclassification of positive cases, the disagreement regarding MRI-SI is more balanced (as many positive as negative misclassifications)⁸.

Our data provide interesting information of what would happen in case of testing eligibility of patients for clinical trials. Potentially 163/582 patients in which MRI-SI and/or X-SI reading was different between local reading and central reading could have a different classification according to the ASAS axSpA criteria. If patients in the DESIR-cohort would have been included in a clinical trial requiring fulfillment of mNY criteria based on local reading, 76/183 (41.5%) of the patients would not have fulfilled the mNY criteria by central reading. Similarly, 38/505 (7.5%) of the patients would be included based on central reading but not based on local reading⁸. Assuming a requirement of sacroiliitis on MRI according to local reading, 38/223 (17.0%) of the patients included would not be eligible based on central reading; the other way around, 51/440 patients (11.6%) not eligible for inclusion based on local reading would be included based on central reading. However, if inclusion would have been based on fulfillment of the imaging arm of the ASAS axSpA criteria the total percentage of reclassified patients would be 15.6% (91/582); 44 patients (7.6%) eligible based on central reading would not be included based on local reading and 47 patients (8.1%) the other way around. Based on the fulfillment of the entire axSpA criteria this percentage is 7.9% (46/582 patients); 28 patients (4.8%) would be included based on central reading but not on local reading and 18 patients (3.1%) the other way around.

The effect of local versus central reading regarding fulfillment of mNY criteria became recently evident by data provided to the Food and Drug Administration (FDA). In both the ABILITY-I and RAPID-axSpA trial, over 25% of the patients were reclassified as fulfilling mNY criteria based on central reading while they were entered as nr-axSpA based on local reading^{2, 3, 9, 10}. The DESIR-cohort confirms this disagreement between local and central readers in the largest cohort addressing this issue. Moreover, there are no data on this aspect for MRI-SI this far, so the data presented in this study are the first data on MRI-SI in a large group of patients. As X-SI reading is so unreliable, the question arises whether it would be an option to only conduct MRI-SI and leave out X-SI completely, especially if structural lesions on MRI-SI are considered as well. More data from other cohorts, including patients with long-standing disease, are necessary to address this question in more detail.

Without knowing the truth of the result of imaging, central reading based on a consensus score of 2/3 readers, is the best approximation of the truth, followed by the reading of one central reader trained in the scoring, followed by local reading, (readers not specifically trained for this purpose). The choice for local reading or central reading for inclusion in clinical trials depends also on the purpose: if the aim is to test a drug in the way it will be applied in clinical practice, local reading would be preferred; if the aim is testing efficacy in the purest population, central reading would be preferred. The latter is mostly required by registration agencies. Furthermore, the European Medical Agency has approved TNF-inhibitors for patients with nr-axSpA only if additional signs of objective inflammation such as elevated CRP and/or a pos-MRI are present, while in patients fulfilling the mNY no additional sign of objective inflammation is required. Looking at all axSpA patients (including patients fulfilling the clinical arm) in the DESIR-cohort and assuming eligibility of all patients

for treatment with TNF-inhibitors (i.e. assuming that patients in the clinical arm had signs of objective inflammation and that all patients had active disease), 18 patients could have had inappropriate treatment with TNF-inhibitors and 28 patients were not treated with TNF-inhibitors based on false classification by local reading in comparison to the external standard of central reading. It should be noted that this situation implies an intrinsic dissimilarity in requirements to start with TNF-inhibitors based on the potentially fallible judgement on the presence or absence of radiographic sacroiliitis.

This study has several strengths we would like to address. The DESIR-cohort consists of a high number of patients, and in every patient both local reading and central reading of the same baseline set of images is available, thereby offering the unique opportunity to investigate the effect of local reading versus central reading. As patients were recruited in 25 centers where several rheumatologists and radiologists are working, local reading is a wide representation of clinical practice. Furthermore, central reading was performed by two independent trained readers and included an adjudication score, ensuring the robustness of central reading.

The main limitation of this study is that the DESIR-cohort only comprises patients with short disease duration. Patients with short symptom duration usually do not show extensive lesions, thereby making recognition of lesions in patients in the DESIR-cohort probably more difficult than in patients with established disease. Thus the results regarding agreement on positive imaging presented in this study might be slightly worse than could be expected in more established diseased patients. Another limitation is the fact that all sites were in France. It is unknown if this is generalizable to other countries. However, the two RCTs with similar percentages of disagreement in X-SI scores included many international sites across the world. Lastly, the role of structural damage on MRI-SI has not been taken into account. It would be interesting to know how the agreement between local and central reading is for this aspect, and if these structural changes could be taken into account in addition to or instead of the X-SI.

In conclusion, substantial levels of agreement between the two central readers and between local reading and central reading indicate that both local rheumatologists/radiologists and trained readers performed well in recognizing a pos-MRI in patients with recent onset IBP. However, when taking into account the reading of X-SI as well, levels of agreement between local reading and central reading are decreasing, yet it is reassuring that only 7.9% of the patients in the DESIR-cohort were classified differently using the full ASAS axSpA criteria, based on a different reading of the same set of images by local reading and central reading. These results point out the robustness of the ASAS axSpA classification criteria to differences in reading of the images, showing that these criteria can be applied reliably in clinical practice.

SUPPLEMENTARY DATA

Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/annrheumdis-2014-205432>).

REFERENCES

1. Rudwaleit M, van der Heijde D, Landewé R, *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.
2. Landewé R, Braun J, Deodhar A, *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.
3. 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;72:815-22.
4. van der Linden S, Valkenburg H, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum* 1984;27:361-8.
5. Rudwaleit M, Jurik A, Hermann K, *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.
6. van Tubergen A, Heuft-Dorenbosch L, Schulpen G, *et al.* Radiographic assessment of sacroiliitis by radiologists and rheumatologists: does training improve quality? *Ann Rheum Dis* 2003;62:519-25.
7. Marzo-Ortega H, McGonagle D, Bennett A. Magnetic resonance imaging in spondyloarthritis. *Curr Opin Rheumatol* 2010;22:381-7.
8. van den Berg R, Lenczner G, Feydy A, *et al.* Reading of sacroiliac joints on plain pelvic radiographs: agreement between clinical practice and trained central reading. Results of the DESIR-cohort. *Arthritis Rheumatol* 2014;66:2403-11.
9. <http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ArthritisAdvisoryCommittee/UCM366517.pdf>. (24 December 2013, date last accessed).
10. <http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ArthritisAdvisoryCommittee/UCM366519.pdf>. (24 December 2013, date last accessed).
11. Dougados M, D'Agostino M, 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;78:598-603.
12. Calin A, Porta J, Fries J, *et al.* Clinical history as a screening test for ankylosing spondylitis. *JAMA* 1977;237:2613-4.
13. Rudwaleit M, Metter A, Listing J, *et al.* Inflammatory back pain in ankylosing spondylitis: a reassessment of the clinical history for application as classification and diagnostic criteria. *Arthritis Rheum* 2006;54:569-78.
14. Molto A, Paternotte S, van der Heijde D, *et al.* Evaluation of the validity of the different arms of the ASAS set of criteria for axial spondyloarthritis and description of the different imaging abnormalities suggestive of spondyloarthritis: data from the DESIR-cohort. *Ann Rheum Dis* [online 3-1-2014] doi: 10.1136/annrheumdis-2013-204262.
15. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70:213-20.
16. Dice L. Measures of the amount of ecological association between species. *Ecology* 1945;26:297-302.
17. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
18. Cicchetti D, Feinstein A. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990;43:551-8.
19. Arnbak B, Jensen T, Manniche C, *et al.* Spondyloarthritis-related and degenerative MRI changes in the axial skeleton - an inter- and intra-observer agreement study. *BMC Musculoskelet Disord* 2013;14:274.

