Performance of Magnetic Resonance Imaging in the Diagnosis of Axial Spondyloarthritis: A Systematic Literature Review

Alexis Jones[1†], Timothy JP Bray[2†], Peter Mandl[3], Margaret A Hall-Craggs[2], Helena Marzo-Ortega[4,5], Pedro M Machado[1,6,7]

[1]Department of Rheumatology, University College London Hospitals NHS Foundation Trust, London, UK

[2]Centre for Medical Imaging, University College London, London, UK

[3]Department of Rheumatology, Medical University of Vienna, Vienna, Austria

[4]NIHR NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds, UK

[5]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK

[6]Department of Rheumatology, Northwick Park Hospital, London North West University Healthcare NHS trust, London, UK

[7]Centre for Rheumatology and MRC Centre for Neuromuscular Diseases, University College London, London, UK

[†]These authors have contributed equally to this work.

Correspondence to

Pedro M Machado

University College London

1st Floor, Russell Square House

10-12 Russell Square, WC1B 5EH London

Email: p.machado@ucl.ac.uk

Current word count (excluding title page, abstract, references, figures and tables): 3476

Number of tables: 4

Number of figures: 1

Number of supplementary files: 1

Keywords: Spondyloarthritis, Ankylosing Spondylitis, Sacroiliitis, Magnetic Resonance Imaging, Inflammation, Diagnosis, Diagnostic Performance, Systematic Literature Review.

Abstract

Objectives: To summarise the evidence on the performance of MRI for the diagnosis of axial spondyloarthritis (axSpA).

Methods: Systematic literature review (SLR) of all studies from January 2013 to March 2017 including adult patients with clinically suspected axSpA undergoing MRI. Studies from a previously published SLR up to January 2013 were also included.

Results: Thirty-one studies were included. Six studies demonstrated good sensitivity and specificity for sacroiliac joint (SIJ) bone marrow oedema (BMO). Specificity was increased by the presence of other structural lesions alongside BMO, particularly erosions or fat infiltration. Four studies addressed the utility of SIJ fat infiltration, finding good sensitivity but poor specificity. SIJ erosions showed good specificity in five studies. Studies addressing high T1 signal in the SIJ, fluid signal in the SIJ, ankylosis, sclerosis, capsulitis, backfill and vacuum phenomenon reported limited diagnostic value. In the spine, four studies reported moderate sensitivity and specificity for corner inflammatory lesions, and four reported poor sensitivity and specificity for spinal fat infiltration. Five studies evaluated the added value of spinal MRI over SIJ MRI alone, with variable results depending on the cohort. Six studies addressed the effect of acquisition parameters on diagnostic accuracy: fat-saturated T2-weighted imaging and STIR imaging showed comparable utility in identifying BMO. Three studies showed that gadolinium was of minimal added value in the detection of BMO.

Conclusions: These results confirmed the diagnostic utility of MRI in axSpA. Performance varied according to the characteristics of the cohort and the number and combination of MRI lesions considered.

Introduction

Spondyloarthritis (SpA) encompasses a group of immune-mediated inflammatory diseases characterised by axial inflammation, peripheral arthritis, enthesitis, dactylitis and extra-articular features such as psoriasis, uveitis and inflammatory bowel disease (1). The Assessment of Spondyloarthritis International Society (ASAS) simplified the classification of SpA by dividing the group into axial (2) and peripheral SpA (3). Peripheral SpA refers to disease with predominantly peripheral features, while axial SpA (axSpA) describes patients with disease predominately affecting the axial skeleton. The ASAS classification criteria are not diagnostic criteria but aim to differentiate groups of patients within a disease spectrum, mainly for research purposes. The diagnosis of axSpA should be made by the clinician based on the combination of clinical, laboratory and imaging features, with MRI being one of the imaging components.

MRI is the only imaging technique capable of detecting both active (inflammatory) and chronic (structural) lesions as well as their anatomical distribution, contributing to the early diagnosis of axSpA. MRI correlates with histological findings in axSpA (4), is a predictor of response to therapy and can be used to monitor disease activity over time (1).

Despite the clear utility of MRI in axSpA, there remains inconsistency around its use in clinical practice. A recent survey of 269 radiologists in acute UK National Health Service trusts/health boards showed substantial variability in the use of paramagnetic contrast, sequence choice and anatomical coverage (5). This survey found that only 75% of radiologists were aware of the term axSpA, and only 31% and 25% were aware of the ASAS definitions of positive MRI of the SIJ and spine (5). Despite being widely accepted as a key diagnostic marker, bone marrow oedema (BMO) was not used as a potential diagnostic feature of axSpA by 18% of radiologists (5).

The heterogeneity of MRI protocols and image interpretation is likely to cause inconsistency in the way that axSpA is diagnosed and may lead to missed or delays in diagnosis and inadequate or unnecessary treatment for patients. As such, there is a need to standardise the use of MRI and a consensus on how MRI lesions should be interpreted in relation to axSpA.

The aim of this systematic literature review (SLR) is to summarise the available evidence on the diagnostic utility of MRI in axSpA, including the significance of specific lesions, the influence of anatomical coverage and effect of acquisition parameters. The results of this SLR will be used to inform future consensus exercises regarding the use of MRI in axSpA.

Materials and Methods

Research Questions

Members of a British Society of Spondyloarthritis (BRITSpA) MRI task force (nine musculoskeletal radiologists and nine rheumatologists with an interest in axSpA), proposed clinically relevant research questions (RQs) related to key aspects of the use of MRI in axSpA. Three final research questions (RQ1-3) were formulated and agreed upon by consensus (Table 1).

These questions were framed according to the Population, Intervention, Comparator, Outcome (PICO) format (6), as detailed in Supplementary Information Section 1 (SS1 - Tables S1, S2 and S3). For all three questions, the population of interest consisted of adult patients (≥18 years) with suspected and/or established axSpA, and the reference standard consisted of a clinical diagnosis of axSpA (optimal scenario) or global imaging criteria considered suggestive of axSpA (suboptimal scenario). The outcomes of interest were the sensitivity, specificity and likelihood ratios for the diagnosis of axSpA; for RQ2 and RQ3 additional endpoints including the prevalence of spinal inflammation in groups with and without SIJ inflammation and additional metrics relating to sequence performance (see SIS1).

Study Selection and Data Extraction

The SLR was conducted by two reviewers (TJPB and AJ) under the guidance of the methodologist (PMM). The search strategy (SIS2) from a previous European League Against Rheumatism (EULAR) systematic review, addressing the role of imaging in spondyloarthritis, was adopted (7). MEDLINE (1946), Embase (1974) and Cochrane (1993) databases were searched without language restrictions. We included all studies performed between January 2013 and March 2017, in addition to relevant studies selected from the previous EULAR SLR, which included all studies from the inception of the databases up to January 2013 (7). Each reviewer screened titles and abstracts of all citations independently, and potentially

relevant articles were reviewed in full text (SIS3 – Figure S1). Papers fulfilling the inclusion criteria underwent full data extraction (SIS4 - Tables S4, S5 and S6) and were assessed for risk of bias (RoB) (SIS5 - Tables S7, S8 and S9). Both reviewers independently retrieved data using a predefined data extraction sheet. The following data were extracted: main characteristics of study (authors, journal and year of publication), study design, number of included patients (subdivided into axSpA patients and controls), reference standard, features of interest, technical factors relating to the acquisition (magnetic field strength, slice thicknesses, use of gadolinium, acquisition planes, spine coverage and sequence parameters), and the relevant outcome data. For studies addressing the effect of acquisition parameters (Q3), we also recorded technical performance metrics including the contrast-to-noise ratio (SIS 6 – Figures S2, S3 and S4).

Quality Assessment

Each study was assessed independently for RoB by the same two reviewers who conducted the SLR (TJPB and AJ) using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool (SIS5 - Tables S7, S8 and S9). This tool involves RoB assessment in four domains (patient selection, index test, reference standard, flow and timing); the first three domains are also assessed for applicability concerns, resulting in seven separate assessments for each study. Each assessment produced a rating of ‘low’, ‘high’ or ‘unclear’ (assigned scores of 0, 1 and 2 respectively). Discrepancies between reviewers regarding study selection, data extraction and RoB assessment were solved by discussion; a third reviewer (PMM) was available in case no consensus could be achieved.

Results

Of the 8114 articles screened, 31 studies were finally included (SIS3 – Figure S1). Twenty articles related to the diagnostic accuracy of specific lesions on MRI in the diagnosis of axSpA (Q1) (8–27), five articles related to the influence of anatomical coverage on diagnostic performance (Q2), and six related to the influence of acquisition parameters (Q3).

Diagnostic Accuracy

Sensitivity, specificity and likelihood ratios for each of the studies investigating diagnostic accuracy are shown in Table 2; results for BMO and combinations of BMO with other features are also shown graphically in Figure 1. The main study and patient characteristics of these studies are summarised in Table S4, and details on the RoB assessment are described in Table S7.

Sacroiliac Joints

Six studies investigated the diagnostic utility of BMO in the sacroiliac joints (SIJ) (10,14,16,22,26,28) (Table 3). In general, these studies showed that BMO was the most sensitive individual lesion for the diagnosis of axSpA, although sensitivity (SE) (0.35–0.91) and specificity (SP) (0.75-0.90) estimates varied depending on the patient cohort, definition used for the reference standard, and number of MRI lesions used to categorise the patients (10,14,16,22,26,28) .

Defining a reference standard for axSpA is challenging. Expert clinical opinion has limitations and is frequently made with knowledge of imaging results, leading to circular interpretation. Imaging standards fail to reflect the full clinical picture of axSpA, and there is a well-known delay from disease onset to radiographic changes. Weber et al (10,26) used clinical examination and plain radiography to identify those patients with axSpA. In their earlier study, Weber et al (14) used a 'global assessment of MRI' to confirm a positive diagnosis of axSpA. Jans et al (22) used the ASAS classification criteria as their reference standard in patients undergoing MRI with inflammatory back pain. Wick et al (16) used a retrospective diagnosis of axSpA from clinical notes – it is unclear whether MRI had been used to make this diagnosis. Marzo-Ortega et al (28) used Calin's criteria for the diagnosis of inflammatory back pain at baseline and one year.

There were subtle differences in the definition of BMO among authors. Jans et al (22) defined a positive MRI SIJ for BMO if there was high T2FS/STIR signal of the ilium or sacrum typically located periarticularly. If there was only one lesion, this had to be present on at least two consecutive slices. If there was more than one signal on a single slice, this was considered adequate. Weber et al (10,14) used a relatively similar definition using the

SPARCC assessment, where the SIJ is represented as a schematic with 4 quadrants. As with the ASAS definition, BMO had to be present in ≥2 SIJs quadrants on the same slice or in the same SIJ quadrant on ≥2 consecutive slices. In an earlier study, Weber et al, 2013 (26) used a cut off of BMO in at least one quadrant. Marzo-Ortega et al (28) used the Leeds scoring system: BMO was defined as low signal on T1 with enhancement after gadolinium administration and/or high or intermediate bone marrow signal with irregular contour on a T2 SPIR image. The presence of BMO was recorded and severity ranked on a semi quantitative scale based on the percentage area covered in each quadrant: 0, absent; grade 1, mild (<25%); grade 2, moderate (25-75%); grade 3, severe (75%). An overall score of inflammatory activity was calculated as the sum of scores of BMO. A positive MRI SIJ was defined as moderate/severe BMO (score ≥2).

Both Jans et al. (22) and Weber et al. (10,14,26) found that the combinations of BMO and/or erosions could increase the sensitivity and specificity of MRI for the diagnosis of axSpA. Sensitivity and specificity were also increased by the combination of BMO and fat infiltration (20,22). Jans et al. (22) also reported an increase in specificity (but significant decrease in sensitivity) for the presence of BMO concomitantly with enthesitis, capsulitis or ankylosis.

Weber et al. investigated specific lesion-based criteria for defining a global positive sacroiliac joint MRI, and derived estimates of sensitivity and specificity for a number of different lesion cut-offs (10). It was shown that lesion-based criteria including both BMO and erosions had superior sensitivity compared to criteria including BMO alone; for example the presence of BMO in ≥3 quadrants and erosions in ≥3 quadrants produced SE 0.83 and SP 0.85 for the fulfilment of the global imaging criteria for axSpA (10). However, estimates of sensitivity and specificity again varied substantially depending on the patient cohort.

Four studies addressed the utility of fat infiltration adjacent to the SIJ (9,20,22,26). The presence of fat infiltration was found to have low/moderate sensitivity (0.15-0.70) and moderate/high specificity (0.72-0.95) for the diagnosis of axSpA, although estimates varied depending on study design, the specific axSpA population under investigation and lesions' cut-offs (9,20,22,26). Weber et al. found that fat infiltration was more specific for the diagnosis of AS than for non-radiographic axSpA (SE/SP 0.7/0.72 and 0.46/0.72 respectively) (20). De Hooge et al. showed that using a cut off of ≥3 fatty lesions correctly classified 63.6%

of AS patients, whilst a combined threshold of ≥5 fatty lesions and/or erosions performed similarly well (9).

Five studies investigated the diagnostic utility of erosions (Table 3) (9,10,16,20,26). In general, erosions demonstrated good specificity for the diagnosis of axSpA, but only poor to moderate sensitivity (Table 3). Erosions were more sensitive in AS than in non-radiographic axSpA or axSpA as a whole (9,20), and were more sensitive against a pre-specified MRI reference standard than against a clinical reference standard (26). Using both erosions and fat infiltration as a diagnostic criterion increased specificity, but reduced sensitivity, compared to criteria consisting of fat infiltration alone (20).

Three studies addressing other SIJ lesions including high T1 signal in the SIJ, fluid signal in the SIJ, ankylosis, vacuum phenomenon, sclerosis, enthesitis, capsulitis and backfill reported low to moderate diagnostic performance for these features (8,22,23).

Spine

Five studies demonstrated moderate sensitivity and specificity of spinal inflammatory lesions in the diagnosis of axSpA (Table 4) (9,11,13,21,25). In general, these studies demonstrated that lower thresholds for the number of inflammatory lesions resulted in reasonable sensitivity but poor specificity; increasing the threshold improved specificity but worsened sensitivity. Four of the five studies also investigated the diagnostic utility of spinal fatty lesions, and found poor sensitivity and high specificity, shown in Table 4 (9,11,21,25). De Hooge et al (9), however, found that ≥5 spinal inflammatory lesions and ≥5 spinal fatty lesions to be highly specific for axSpA, while still assuring an acceptable and useful level of discrimination between axSpA patients and non axSpA patients.

Effect of Anatomical Coverage

Five studies evaluated the added value of combined spinal and SIJ MRI over SIJ MRI alone (28,30–33). Study and patient characteristics for these studies are summarised in Table S5, and details on the RoB assessment are described in Table S8. Two studies found that combined spinal and SIJ MRI did not add significant value over SIJ MRI alone, either because spinal inflammation was rare in the absence of SIJ inflammation (30) or because combined MRI resulted in a high rate of false positives (31). However, three studies observed spinal

inflammation in up to half of patients without SIJ inflammation, arguing that combined MRI adds value over SIJ alone (28,32,33).

Effect of Acquisition Parameters on Diagnostic Performance

The key acquisition parameters including method of fat suppression, anatomical coverage and use of contrast for all included studies, are summarised in SIS6. Six of the included studies specifically investigated the effect of acquisition parameters (34–39). Study and patient characteristics for these studies are summarised in Table S6, and details on the RoB assessment are described in Table S9.

Of the six studies, three investigated the effect of sequence choice on diagnostic accuracy of axSpA or on the characteristics of the images themselves (34–36). Boy et al. found that sensitivity and specificity was highest for FS-T2W imaging, and progressively decreased for STIR, diffusion-weighted and dynamic-contrast enhanced images respectively (34). Dalto et al. showed good levels of agreement between FS-T2W imaging and STIR imaging, with a Lin's concordance correlation coefficient of 0.94 for reader 1 and 0.88 for reader 2 (range 0 to 1) (36). Ozgen et al. investigated the role of T2-weighted Dixon imaging in the identification of BMO, and found a superior contrast-to-noise ratio compared to FS-T2W imaging (35). Three studies investigated the role of gadolinium in the SIJs, and overall found minimal or no added value (37–39).

Discussion

We systematically reviewed the literature regarding the use of MRI in the diagnosis of axSpA, informing a task force of radiologists and rheumatologists with the aim of standardising the use of MRI in suspected axSpA.

Overall, studies investigating specific SIJ MRI lesions have shown that BMO is the most sensitive and specific individual lesion. Structural lesions including fat infiltration have moderate sensitivity and specificity, whilst erosions demonstrate good specificity but relatively poor sensitivity. An important consideration is that several of these studies use fixed specificity values; it is likely that specificity would be lower, but sensitivity higher, if these values were allowed to vary freely.

Other SIJ lesions including high T1 signal in the SIJ, fluid signal in the SIJ, ankylosis, vacuum phenomenon, sclerosis, enthesitis, capsulitis and backfill have a low to moderate diagnostic utility, and are, therefore, unlikely to be of diagnostic value in isolation. Owing to the heterogeneity of the data, with varying reference standards and patient cohorts across studies, or repeated use of the same cohort (implying an overlap in at least part of the study populations) we have been unable to create an accurate meta-analysis of lesion-based criteria in the diagnosis of axSpA.

A number of studies have assessed combinations of lesions and their diagnostic performance. These studies showed that a combination of BMO and erosions, or BMO and fat infiltration, yielded higher sensitivity and specificity than BMO alone. Pre-defined numbers of lesions or cut-offs have also been analysed and suggest that BMO in $\geq$3 quadrants and erosions in $\geq$3 quadrants show high sensitivity and specificity and presence of 3-5 fatty lesions also yield good sensitivity. However, further studies are required to validate these findings.

In the spine, studies investigating the value of spinal inflammatory lesions found moderate sensitivity and specificity, whilst spinal fatty lesions were found to have relatively poor sensitivity and specificity. Although the results suggest that spinal lesions alone are unlikely to have sufficient diagnostic performance for use in axSpA, these lesions might be useful in combination with features identified on SIJ MRI – this is an area that requires further research.

The results of studies investigating the effect of anatomical coverage on diagnosis were mixed: two studies suggested that spinal inflammation is rare in the absence of SIJ inflammation, three found the opposite. Assuming patients seen in clinical practice have variable presentations, imaging the spine would facilitate the diagnosis and management of patients with axial pain. Unfortunately, even amongst studies that have imaged the spine, there has been substantial heterogeneity in anatomical coverage (SIS6), and there is clearly scope for further work to determine the 'optimal' spinal protocol. Importantly, this research will need to consider the trade-off between scan time (and therefore also cost) and diagnostic yield, particularly as pressures on radiology departments continue to increase.

The number of studies assessing the impact of acquisition parameters on diagnostic accuracy was relatively small. The available evidence suggests that contrast adds little value, although no studies have rigorously addressed this question in the spine. Again, there is a need for further research to address this issue.

Of the studies specifically investigating sequence choice, several studies investigated methods of fat suppression other than STIR imaging. FS-T2W was shown to have superior sensitivity and specificity to STIR imaging (34), with assessments of disease severity at the MRI level agreeing closely between the two sequences (36). Similarly, Ozgen et al. demonstrated superior contrast-to-noise ratios for T2W Dixon imaging compared to STIR, but did not assess diagnostic sensitivity (35). Overall, these methods are promising alternatives to STIR and may offer improvements in image quality in the future.

There are several limitations of the studies included in this SLR. First, a number of the studies were potentially biased by the inclusion of information from MRI scans in their reference standard (Tables S7, S8 and S9). In some studies, a positive MRI scan was used as an inclusion criterion; other studies selected patients based on previous MRI scans. Even those studies that did not explicitly use MRI-based reference standards, it is unclear whether MRI had been used in the patients' prior diagnostic work-up or referral.

A true assessment of the diagnostic utility of MRI would omit any MRI imaging from the reference standard. However, in the absence of a robust biomarker for the disease, finding an accurate and reliable reference standard poses a challenge. Some studies incorporated a purely clinical reference standard with a diagnosis of axSpA made by a panel of expert physicians. An alternative approach might be to use reference standards based on follow-up and assessment at multiple time-points, to ensure high level of confidence in the diagnosis of axSpA.

The use of control groups by the included studies was suboptimal, resulting in 'unclear' or 'high' RoB for a number of studies when assessed using the QUADAS-2 tool. Healthy controls can artificially inflate the sensitivity and specificity statistics, since it is typically easier to distinguish axSpA from healthy patients than from patients with other axial problems, namely chronic non-specific low back pain.

On a similar note, there remains uncertainty about the frequency of MRI lesions in the general population. Marzo-Ortega et al (28) reported a high prevalence of BMO affecting the SIJs of up to 6/22 (27%) in a control sample of healthy volunteers and patients with mechanical back pain. Similar findings have been found in postpartum women (40) runners (40), soldiers (41) athletes (42) and the general population (43). In patients with chronic low back pain recruited from primary care without previous rheumatological assessment, 21% met the MRI classification criteria based on SIJ BMO alone, but 42% of these lesions were small and of questionable clinical relevance as they showed no association with clinical SpA features (44).

A further limitation of this SLR is that the numbers of studies included under each of the research questions (RQs) was relatively small. The number of studies was particularly small for RQ2 and RQ3, however, all relevant papers available have been included. Further work is needed to answer these questions more definitively.

Future research into the use of MRI in axSpA should assess MRI scans longitudinally in a cohort of patients with suspected axSpA, correlating lesions with symptoms, response to treatment and rate of radiographic progression. This cohort should cover the entire spectrum of axial disease. Separate studies on healthy controls should aim to assess the background noise of SIJ and spinal lesions associated with mechanical causes in a normal population, providing guidelines on minimum requirements or 'cut-offs' for lesions to determine an abnormal scan.

Advanced MRI techniques, including quantitative MRI (qMRI) hold promise for more accurate assessment of inflammation. qMRI techniques typically use a succession of scans to ‘probe’ tissue characteristics, and infer attributes such as cellularity, vascularity or fat content. Each pixel (picture element) in a qMRI image has a measurable numerical value that reflects the intrinsic properties of a tissue, rather than arbitrary signal intensity produced by standard MRI (45,46). The application of these techniques to axSpA could potentially improve the understanding and management of the disease both through improvements in precision and through a more detailed assessment of bone marrow pathophysiology.

To conclude, the results of this SLR have informed the recommendations of a consensus group aiming to standardise practice around the use of MRI scan in the UK and can inform similar exercises in other countries or at the international level.

Key messages

- This SLR summarises the evidence on the performance of MRI techniques for the diagnosis of axial spondyloarthritis.
- Results of the SLR provide information for recommendations aiming to standardise practice around the use of MRI scans.

Acknowledgements

This work was supported by the British Society of Spondyloarthritis (BRITSpA). Pedro M Machado and Margaret A Hall-Craggs are supported by the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre (BRC). Philip G Conaghan, Andrew Grainger, Helena Marzo-Ortega, Denis McGonagle and Phill O'Connor are supported by the NIHR Leeds BRC. The views expressed are those of the authors and not necessarily those of the (UK) National Health Service (NHS), the NIHR, or the (UK) Department of Health. Timothy JP Bray is supported by Arthritis Research UK Grant 21369.

Competing interests

HMO has received grants and/or honoraria from Abbvie, Celgene, Eli-Lilly, Janssen, MSD, Novartis, Pfizer and UCB. PGC has performed consultancies or speakers bureaus for Abbvie, BMS, Novartis, Pfizer and Roche. PMM has received consulting/speaker's fees from Abbvie, BMS, Celgene, Janssen, MSD, Novartis, Pfizer, Roche and UCB. None of the other authors declared competing interests.

Tables

Table 1: Research questions (RQ) generated by the BRITSpA working group.

RQ1	Which lesion, or combination of lesions, is most sensitive and specific for the diagnosis of axSpA?
RQ2	How does the choice of anatomical region influence diagnostic performance?
RQ3	How do MRI acquisition parameters influence diagnostic performance?

Table 2: Sensitivity and specificity of criteria using bone marrow oedema (BMO) and combinations in the sacroiliac joints.

Feature	Study	Criterion	n	Reference standard	Sensitivity	Specificity	LR+	LR-
BMO	Jans et al. (22)	BMO	517	Clinical diagnosis SpA	0.65	0.75	2.60	0.47
	Weber et al. (26)	BMO ≥ 2 quadrants	177	Global MRI score	0.91/0.83 *	0.90/0.90 */**	9.10/8.30 *	0.10/0.19 *
				Clinical diagnosis SpA	0.73/0.39 *	0.90/0.90 */**	7.30/3.90 *	0.30/0.68 *
	Weber et al. (10)	BMO ≥ 2 quadrants (ASAS definition)	157	Clinical diagnosis SpA	0.8/0.42*	0.76/0.73 *	3.37/1.54 *	0.26/0.80 *
	Wick et al. (16)	BMO	179	Clinical diagnosis SpA	0.35	0.78	1.59	0.83
	Weber et al. (14)	BMO ≥ 2 quadrants (ASAS definition)	187	Clinical diagnosis SpA	0.9	0.97	30.0	0.10
	Marzo-Ortega et al. (28)	BMO > 0 (Leeds scoring)	76	Clinical diagnosis SpA	0.82	0.42	1.41	0.43
BMO and erosions	Jans et al. (22)	BMO and erosion	517	Clinical diagnosis SpA	0.77	0.81	4.05	0.28
	Weber et al. (26)	BMO and/or erosion ≥ 1 quadrant	177	Global MRI score	0.98/0.96 *	0.90/0.90 */**	9.80/9.60 *	0.02/0.04 *
				Clinical diagnosis SpA	0.82/0.51 *	0.90/0.90 */**	8.20/5.10 *	0.20/0.54 *
	Weber et al. (10)	BMO ≥ 2 quadrants and ≥ 1 erosion (MORPHO definition)	157	Clinical diagnosis SpA	0.88	0.72	3.14	0.17

	Weber et al. (14)	BMO and erosion	187	Clinical diagnosis SpA	0.9	0.97	30.0	0.10
BMO and fat infiltration	Jans et al. (22)	BMO and fat	517	Clinical diagnosis SpA	0.68	0.76	2.83	0.42
	Weber et al. (20)	BMO and fat	157	Clinical diagnosis (NR axSpA)	0.39	0.91	4.33	0.67
				Clinical diagnosis (AS)	0.58	0.91	6.44	0.46

***Values for two separate cohorts. **Pre-determined specificity.**

BMO, Bone marrow oedema; LR, likelihood ratio.

Table 3: Sensitivity and specificity of criteria using fat infiltration and erosions in the SIJs.

Feature	Study	Criterion	n	Reference standard	Sensitivity	Specificity	LR+	LR-
Fat infiltration	de Hooge et al. (9)	Fat in ≥3 quadrants	287	Diagnosis AS (mNY)	0.46	0.95**	9.20	0.57
				Diagnosis nr-axSpA	0.15	0.95**	3.00	0.89
				Clinical diagnosis SpA (clinical arm)	0.15	0.95**	3.00	0.89
	Weber et al. (20)	Fat in ≥2 quadrants	157	Diagnosis AS	0.70	0.73	2.59	0.41
				Diagnosis nr-axSpA	0.44	0.73	1.63	0.77
	Jans et al. (22)	Presence of any fat	517	Clinical diagnosis SpA	0.55	0.84	3.44	0.54
	Weber et al. (26)	Lesion-based criteria for fat infiltration	177	Pre-specified positive MRI	0.34/0.74 *	0.90/0.90 *	3.40/7.40 *	0.73/0.29 *
				Clinical diagnosis	0.30/0.49 *	0.90/0.90 *	3.00/4.90	0.78/0.57
Erosions	de Hooge et al. (9)	≥3 erosions	287	Diagnosis AS (mNY)	0.64	0.95**	12.80	0.38
				Diagnosis nr-axSpA	0.47	0.95**	9.40	0.56
				Clinical diagnosis SpA (clinical arm)	0.13	0.95**	2.60	0.92
	Weber et al. (10)	≥2 erosions	157	Clinical diagnosis SpA	0.98/0.77	0.97/0.90	32.7/7.7	0.02/0.26
	Weber et al. (26)	Lesion-based erosion criteria	177	Pre-specified positive MRI	1/1*	0.90/0.90 */**	10	0
				Clinical diagnosis	0.77/0.54 *	0.90/0.90 */**	7.70/5.40	0.26/0.51
	Wick et al. (16)	Presence of any erosion	179	Clinical diagnosis	0.11	0.93	1.57	0.96

Fat infiltration and erosions	Weber et al. (20)	Fat infiltration with erosion	157	Diagnosis AS	0.68	0.98	34.00	0.33
				Diagnosis nr-axSpA	0.34	0.98	17.00	0.67

***Values for two separate cohorts. **Pre-determined specificity.**

LR, Likelihood ratio; mNY, modified New York criteria.

Table 4: Sensitivity and specificity of criteria using inflammatory lesions and fatty lesions in the spine.

Feature	Study	Criterion	n	Reference standard	Sensitivity	Specificity	LR+	LR-
Spinal inflammatory lesions	Weber et al. (25)	≥ 2 CILs	130	Clinical diagnosis	0.53/0.55 *	0.64/0.74 *	1.47/2.12*	0.73/0.61 *
		≥ 3 CILs			0.43/0.25 *	0.75/0.89 *	1.72/2.27*	0.76/0.84 *
	Weber et al. (13)	≥ 2 CILs	95	Clinical diagnosis	0.69	0.94	11.50	0.33
	de Hooge et al. (9)	Presence of spinal inflammatory lesions	287	Diagnosis AS (mNY)	0.27	0.95**	5.40	0.77
				Diagnosis nr-axSpA	0.14	0.95**	2.80	0.91
				Clinical diagnosis axSpA (clinical arm)	0.05	0.95**	1.00	1.00
	Hu et al. (21)	≥ 1 CIL	400	Diagnosis AS (mNY)	0.52	0.55	1.16	0.87
		≥ 6 CILs			0.45	0.66	1.32	0.83
		≥ 11 CILs			0.04	0.78	0.18	1.23
	Bennett et al. (11)	≥ 1 inflammatory lesion[†]	185	Clinical diagnosis	0.67	0.56	1.52	0.59
		>3 inflammatory lesion			0.45	0.81	2.37	0.68
		>3 inflammatory lesions and age <50			0.33	0.97	11.00	0.69
Spinal fatty lesions	Weber et al. (25)	≥ 6 spinal fatty lesions	130	Clinical diagnosis	0.26/0.40 *	0.82/0.81 *	1.44/2.11*	0.90/0.74 *
	de Hooge et al. (9)	Presence of spinal fatty lesions	287	Diagnosis AS (mNY)	0.18	0.95**	3.60	0.86
				Diagnosis nr-axSpA	0.22	0.95**	4.40	0.82

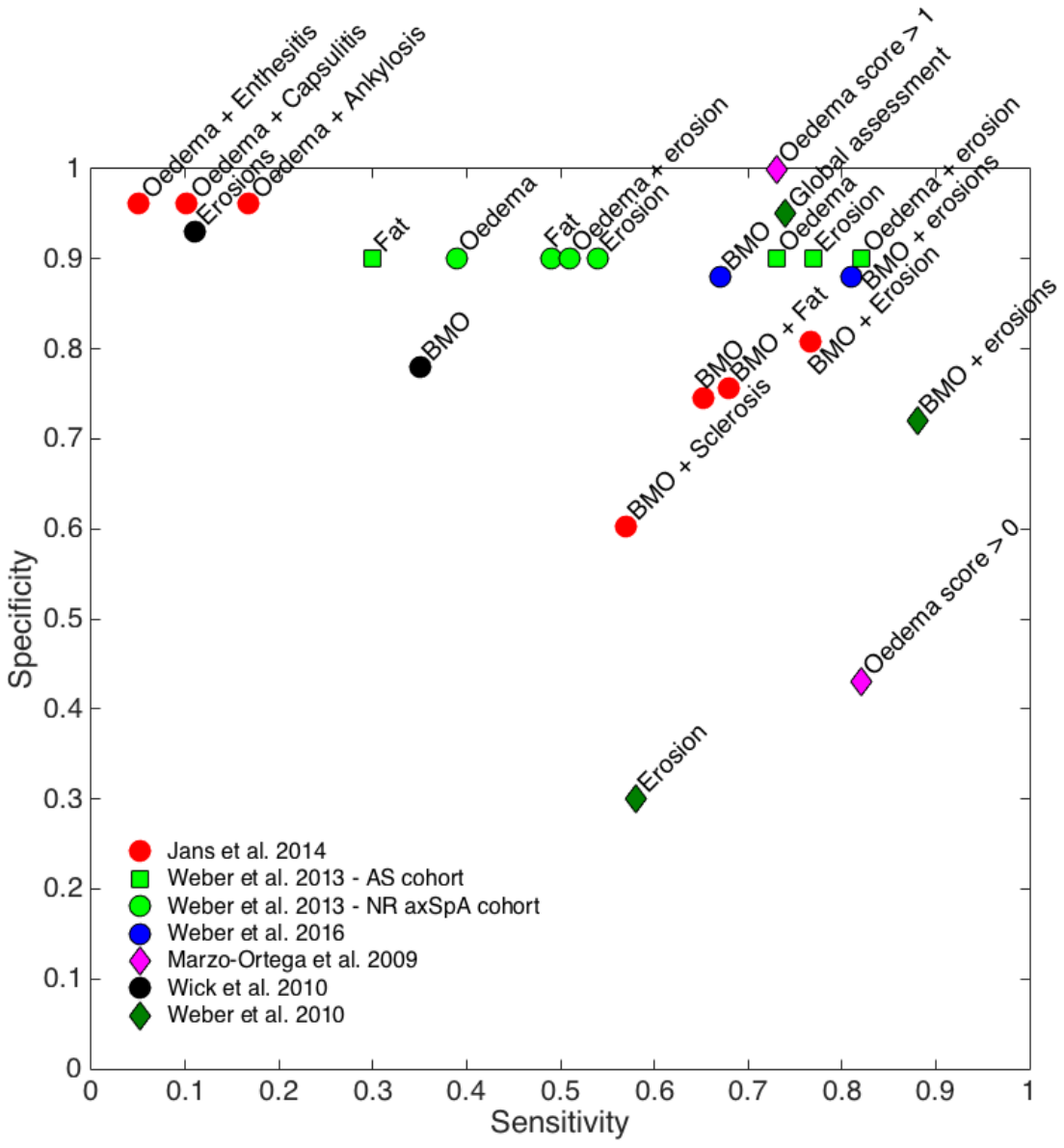
			Clinical diagnosis axSpA (clinical arm)	0.02	0.95**	0.40	1.03
Hu et al. (21)	≥ 1 spinal fatty lesion	400	Diagnosis AS (mNY)	0.13	0.94	2.17	0.93
	≥ 2 spinal fatty lesions			0.09	0.99	9.00	0.92
Bennett et al. (11)	≥ 1 spinal fatty lesion[†]	185	Clinical diagnosis	0.33	0.93	4.71	0.72
	>3 spinal fatty lesions[†]			0.22	0.98	11.00	0.80

*Values for two separate cohorts. **Pre-determined specificity. [†]These lesions were referred to as Romanus lesions and fatty Romanus lesions in (11).

CIL Corner Inflammatory Lesions; LR Likelihood ratio; mNY modified New York;

Figures

Figure 1 – Diagnostic performance of bone marrow oedema and combinations in MRI SIJs. Sensitivity and specificity values are shown on a scatterplot for all relevant studies; performance for other features include in those studies (e.g. erosions alone) is also shown.



Performance of Magnetic Resonance Imaging in the Diagnosis of Axial Spondyloarthritis: A Systematic Literature Review

Supplementary Material

Table of Contents

1. Research Questions

Table S1 – Diagnostic utility of MRI lesions (RQ1)

Which lesion on magnetic resonance imaging (MRI) **(intervention)** has the highest diagnostic utility **(outcome)** for the diagnosis of axial spondyloarthritis using clinical or other imaging criteria as a reference standard **(comparator)**?	
Population	Patients (age ≥18 years) with suspected axial spondyloarthritis (axSpA) (optimal scenario) or established axSpA (sub-optimal scenario, high risk of bias).
Intervention	Magnetic resonance imaging.
Comparator	Clinical diagnosis of axSpA by clinical assessment (for the population of suspected axSpA), control group without axSpA (for the population with established axSpA, sub-optimal scenario, high risk of bias), or global imaging reference standard suggestive of axSpA (sub-optimal scenario, high risk of bias).
Outcome	Test performance reflected in sensitivity, specificity, positive and negative likelihood ratios.

Table S2 – Effect of anatomical coverage on diagnosis (RQ2)

How does the anatomical coverage of the MRI scan **(intervention)** influence the diagnostic utility **(outcome)** for the diagnosis of spondyloarthritis in adult patients suspected of spondyloarthritis **(population)** using clinical or other imaging criteria as a reference standard **(comparator)**?	
Population	Patients (age ≥18 years) with suspected axial spondyloarthritis (axSpA) (optimal scenario) or established axSpA (sub-optimal scenario, high risk of bias).
Intervention	Magnetic resonance imaging.
Comparator	Clinical diagnosis of axSpA by clinical assessment (for the population of suspected axSpA), control group without axSpA (for the population with established axSpA, sub-optimal scenario, high risk of bias), or global imaging reference standard suggestive of axSpA (sub-optimal scenario, high risk of bias).
Outcome	Test performance reflected in sensitivity, specificity, positive and negative likelihood ratios, and/or in the proportion of patients having spinal inflammation in the presence and absence of sacroiliac joint inflammation.

Table S3 – Effect of acquisition parameters

How do the acquisition parameters of the MRI scan **(intervention)** influence the diagnostic utility **(outcome)** for the diagnosis of spondyloarthritis in adult patients suspected of spondyloarthritis **(population)** using clinical or other imaging criteria as a reference standard **(comparator)**?	
Population	Patients (age ≥18 years) with suspected axial spondyloarthritis (axSpA) (optimal scenario) or established axSpA (sub-optimal scenario, high risk of bias).
Intervention	Magnetic resonance imaging.
Comparator	Clinical diagnosis of axSpA by clinical assessment (for the population of suspected axSpA), control group without axSpA (for the population with established axSpA, sub-optimal scenario, high risk of bias), or global imaging reference standard suggestive of axSpA (sub-optimal scenario, high risk of bias).
Outcome	Test performance reflected in sensitivity, specificity, positive and negative likelihood ratios, and/or in technical metrics of performance/image quality, including contrast-to-noise ratio (CNR) and identification of additional areas of bone marrow oedema.

2. Search Strategy

The MEDLINE (via Pubmed), EMBASE (via Ovid) and Cochrane databases were searched using the following terms. Note that imaging modalities other than MRI (radiography, CT, PET and US) were included in the search to avoid missing studies of multiple imaging modalities including MRI; studies which did not involve MRI were excluded at the stage of screening by title and abstract.

MEDLINE via Pubmed

1. "spondylarthropathies"[MeSH Terms]
2. spondylart*[Text Word]
3. (Reactiv*[TI] AND Arthriti*[TI])
4. (Psoria*[TI] AND Arthriti*[TI])
5. (ankyl*[TI] AND Spondyl*[TI])
6. (((inflam*[TiAB] AND (peripher*[TIAB] OR tendon*[TIAB] or tendinop*[TIAB] OR limb*[TIAB]) AND pain [TIAB]))))
7. spondylo*[TiAB]
8. (((inflam*[TiAB] AND (back[TIAB] OR spin*[TIAB]) AND pain [TIAB])))
9. or/1-8
10. "Tomography"[Mesh]
11. "Magnetic Resonance Imaging"[Mesh]
12. "Ultrasonography"[Mesh]
13. "Tomography, X-Ray Computed"[Mesh]
14. "Positron-Emission Tomography and Computed Tomography"[Mesh]
15. "Positron-Emission Tomography"[Mesh]
16. "Tomography, Emission-Computed, Single-Photon"[Mesh]
17. ("magnetic"[All Fields] AND "resonance"[All Fields] AND "imaging"[All Fields])
18. "mri"[All Fields]
19. ultrasono*[TIAB]
20. echograph*[TIAB]
21. "CT scan*"[TIAB]

22. tomograph*[TIAB]
23. scintigraph*[TIAB]
24. (PET[Title/Abstract]) AND tomog*[Title/Abstract])
25. (SPECT[Title/Abstract]) AND photon[Title/Abstract])
26. or/10-25
27. 9 and 26
28. (animals[mh] NOT human[mh])
29. 27 not 28
30. (("case report*" [TI]) OR (case reports[Publication Type]))
31. 29 not 30

EMBASE via Ovid

1. (magnetic and resonance and imaging).mp.
2. magnetic resonance imaging.mp.
3. mri.mp.
4. Ultrasonography.mp. or exp echography/
5. magnetic resonance imaging.mp. or exp nuclear magnetic resonance imaging/
6. "ultrasono*".ti,ab.
7. Tomography, X-Ray Computed.mp. or exp computer assisted tomography/
8. "CT scan*".ti,ab.
9. "echograph*".ti,ab.
10. "tomograph*".ti,ab.
11. "scintigraph*".ti,ab.
12. Positron Emission Tomography.mp. or exp positron emission tomography/
13. (PET and tomog*).ti,ab.
14. Tomography, Emission-Computed, Single-Photon.mp. or exp single photon emission computer tomography/
15. (SPECT and photon).ti,ab.
16. or/1-15
17. exp ankylosing spondylitis/

18. exp psoriatic arthritis/
19. exp reactive arthritis/
20. exp spondyloarthropathy/
21. (inflam* and (peripher* or tendon* or tendinop* or limb*) and pain).ti,ab.
22. "spondylo*".ti,ab.
23. (inflam* and (back or spin*) and pain).ti,ab.
24. or/17-23
25. 16 and 24
26. limit 25 to (conference abstract or conference paper or "conference review" or letter or conference proceeding)
27. 25 not 26
28. limit 27 to (animals or animal studies)
29. limit 28 to human
30. 28 not 29
31. 27 not 30
32. "case report*".m_titl.
33. case study.m_titl.
34. case report/
35. or/28-30
36. 31 not 35

The Cochrane Library

1. MeSH descriptor: [Spondylitis, Ankylosing] explode all trees
2. MeSH descriptor: [Spondylarthropathies] explode all trees
3. MeSH descriptor: [Arthritis, Reactive] explode all trees
4. MeSH descriptor: [Arthritis, Psoriatic] explode all trees
5. MeSH descriptor: [Magnetic Resonance Imaging] explode all trees
6. MeSH descriptor: [Ultrasonography] explode all trees
7. MeSH descriptor: [Tomography] explode all trees
8. MeSH descriptor: [Radionuclide Imaging] explode all trees

9. MeSH descriptor: [Positron-Emission Tomography] explode all trees
10. MeSH descriptor: [Diagnostic Imaging] explode all trees
11. "ultrasound":ti,ab,kw (Word variations have been searched)
12. "sonograph":ti,ab,kw (Word variations have been searched)
13. "CT":ti,ab,kw (Word variations have been searched)
14. "positron emission tomograph":ti,ab,kw (Word variations have been searched)
15. #1 or #2 or #3 or #4
16. #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14
17. #15 and #16

3. Flowchart of Included Studies

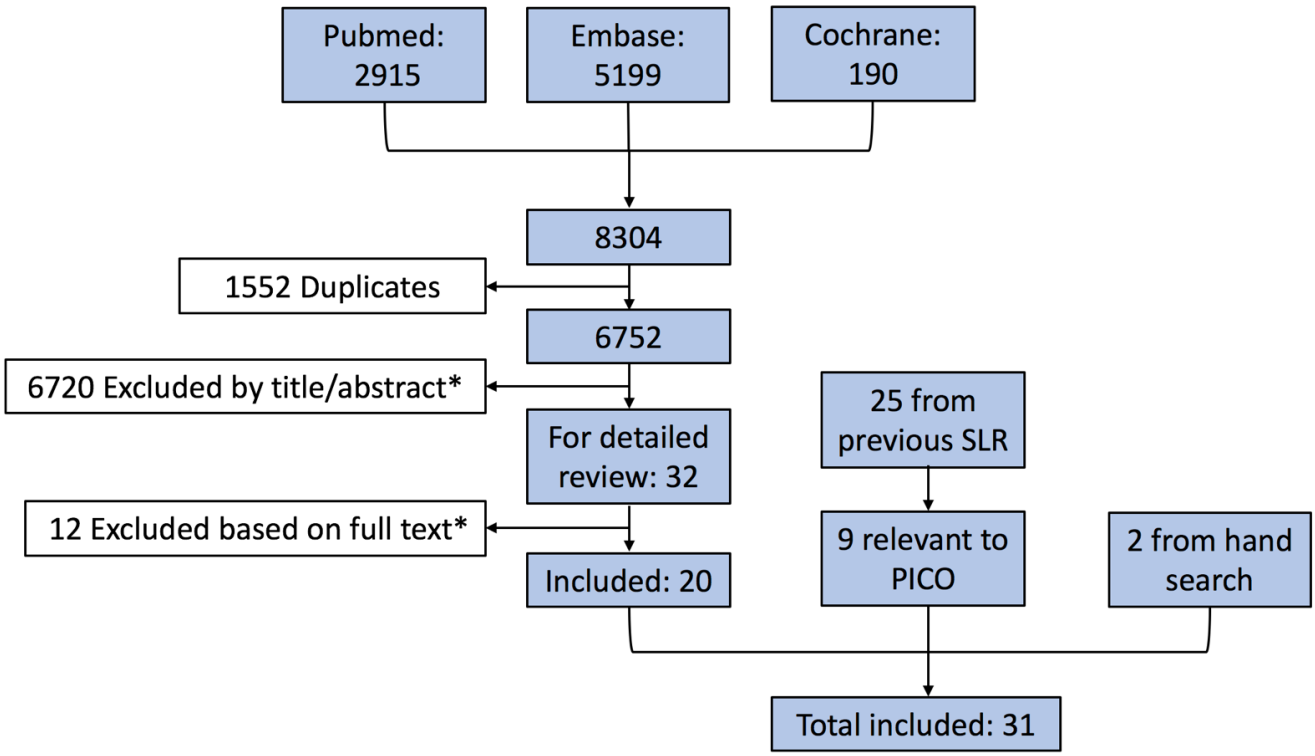


Figure S1 – Flowchart describing the process of study inclusion.

*Any studies which were excluded by title/abstract by one reviewer only also underwent detailed review. Any discrepancies between studies included for detailed review by the two reviewers were resolved by discussion. A third reviewer (PMM) was available to resolve any discrepancies which could not be resolved by discussion if needed.

4. Characteristics of Included Studies

Table S4 – Imaging features (n=18) (8–27)**. Studies are shown in alphabetical order.**

Study ID	n	n axSpA	n AS/nr-axSpA	n Controls	Design	Inclusion criteria	Reference standard	Feature of interest	Anatomical coverage	RoB /14
Bennett et al. (11) M1	185	64		121	Retro	Patients with back pain (not defined) referred for MRI spine	Clinical diagnosis (history, physical examination), non-radiologic and radiologic investigations, +/- histology. ESSG criteria for SpA prior to MRI scanning	Spinal inflammatory lesions	SIJs, whole spine	1
Bennett et al. (12) M2	185	64		121	Retro	Patients with back pain (not defined) referred for MRI spine	Clinical diagnosis (as above)	Spinal fatty lesions	SIJs, whole spine	1
de Hooge et al. (9) 2.	287	126		161	Pro	Patients ≥16 years with CBP (≥3 months, ≤2 years), onset <45years, cause unknown	Clinical diagnosis and fulfilment of ASAS criteria (including MRI)	SIJ fatty lesions, erosions, spinal inflammatory & fatty lesions	SIJs, whole spine	2
Hu et al. (21) 5.	400	192	192/0	208	Retro	Patient with back pain (not defined) referred for MRI	Clinical diagnosis and mNY criteria for axSpA patients	Spinal corner inflammatory lesions	SIJs, lumbar spine	0
Hu et al. (23) 8.	647	423	297/ 126	224	Pro	Patients with axSpA (ASAS criteria) and non-specific mechanical back pain from any cause	Clinical diagnosis and ASAS criteria (including MRI)	Backfill	SIJs	5
Jans et al. (22) 7.	517	210		307	Retro	Patients with IBP back pain (Calin criteria) referred for MRI SIJ	Clinical diagnosis and ASAS criteria (including MRI)	Bone marrow oedema, enthesitis, capsulitis, sclerosis, fat, ankylosis &	SIJs	2

							combinations		
Kim et al. (17) M8	104	52	52	Retro	mNY criteria for AS	Clinical diagnosis and pelvic radiographs (mNY criteria)	MRI corner sign	SIJs, lumbar	3
Laloo et al. (81.	353	151	212	Retro	Patients with IBP (Calin criteria) referred for MRI SIJs	Clinical diagnosis and ASAS criteria (including MRI)	High T1 signal/fluid signal in SIJ, ankylosis, vacuum phenomenon	SIJs	3

Larbi et al. (27) 22.	**110**	**28**		**82**	**Retro**	**Patients referred for SIJ MRI for screening of malignant bone marrow infiltration or for low back or sacral pain (inflammatory and mechanical)**	**Clinical diagnosis by consensus by two rheumatologists and ASAS criteria (including MRI)**	**Bone marrow oedema**	**SIJs**	**3**
Weber et al. (25) 11.	150	83	33/50	67	Pro	Cohort A: Clinically suspected sacroiliitis; Cohort B patients with anterior uveitis and a history of back pain of > 3 months on a structured questionnaire	Clinical diagnosis including structured questionnaires and laboratory tests; use of pelvic radiographs but not MRI (mNY criteria)	Corner inflammatory and corner fatty lesions in spine	Whole spine	1
Weber et al. (26) 19.	177	85	34/51	89	Pro	Suspected SpA or uveitis and CBP (as above)	Clinical diagnosis and pelvic radiographs (as above)	Bone marrow oedema, erosions and fat infiltration (number of quadrants)	SIJs	3
Weber et al. 2015 (10) 29.	157	85	34/51	72	Pro	Suspected SpA or uveitis and CBP (as above)	Clinical diagnosis and pelvic radiographs (as above)	Lesion-based criteria for diagnosis	SIJs	1
Weber et al. (20) 4.	157	85	34/51	72	Pro	Suspected SpA or uveitis and CBP (as above)	Clinical diagnosis and pelvic radiography (as above)	SIJ fatty lesions ± erosions and ± oedema	SIJs	1
Weber et al. (24) 10.	157	85	34/51	72	Pro	Suspected SpA or uveitis and CBP (as above)	Clinical diagnosis and pelvic radiography (as above)	Inflammation in ligamentous part of SIJ	SIJs	0
Weber et al. (13) M3	95	60		35	Cross-sectional	Two cohorts: a) patients fulfilling 2 of 4 IBP criteria and 1 of the following: good response to nonsteroidal anti-inflammatory drugs, peripheral arthritis, enthesitis, dactylitis, uveitis, HLA–B27; and b) patients fulfilling mNY criteria for AS	Clinical diagnosis	Spinal inflammatory lesions	Whole body	1
Weber et al. (14) M4	187	102		85	Cross-section	Patients with IBP (Calin criteria). Patients with AS	Clinical diagnosis	Oedema, overall diagnosis on MRI	SIJs	3

				al	(mNY criteria).				
Weber et al. (15) M5	187	102	85	Cross-sectional	Patients with AS (mNY criteria), IBP (Calin criteria) and clinical arm of nr-axSpA (ASAS)	Clinical diagnosis	Structural lesions on T1w images	SIJs	3
Wick et al. (16) M7	179	128	51	Retro	Recent onset of localized SI pain (duration < 1 year) with clinical suspicious for AS	Clinical diagnosis	Erosions, bone marrow oedema	SIJs	2

Abbreviations: CBP, chronic back pain; ESSG, European Spondyloarthropathy Study Group; IBP, Inflammatory back pain; mNY, modified New York; N, number; Pro, prospective; Retro, retrospective; RoB, Risk of Bias; SIJ, sacroiliac joint; SpA, Spondyloartrhitis; axSpA, axial spondyloarthritis; nr-axSpA, non-radiographic axial spondyloarthritis.

Table S5 – Anatomical coverage (n=5).

Study ID	n	n axSpA	n AS/nr-axSpA	n Controls	Design	Inclusion criteria	Reference standard	Feature of interest	Anatomical coverage	RoB /14
Weber et al. (31) 9.	130	83	33/50	47	Pro	Cohort A: Clinically suspected sacroiliitis; Cohort B: patients with anterior uveitis and a history of back pain > 3 month duration on a structured questionnaire	Clinical diagnosis and pelvic radiographs	Prevalence of inflammation on SIJ MRI and combined SIJ + spinal MRI	SIJs, whole spine	1
Althoff et al. (30) 18.	75	75	39/36	0	Retro	ASAS classification criteria for axial SpA + CBP for <5years, BASDAI ≥4 and spinal VAS ≥4, despite NSAIDS and active inflammatory lesions on MRI SIJ	SIJ/spinal inflammation on MRI	Inflammation in SIJs and spine	Whole body	2
van der Heijde et al. (32) 20.	185	185		0	Pro	nr-axSpA (ASAS criteria) with BASDAI > 4 and spinal VAS > 4 on (ABILITY 1 trial)	Positive MRI (SPARCC ≥ 2)	Presence of spinal inflammation with and without SIJ inflammation on MRI	SIJs, whole spine	4
Larbi et al. (33) 27.	112	112		0	Retro	axSpA diagnosis (ASAS criteria) ≥ 7 years and HLA-B27+	SIJ/spinal inflammation on MRI	Proportion of patients with spinal and SIJ inflammation	Thoracic, lumbar spine	0
Marzo-Ortega et al. (28) M6	76	54		22	Long	Short duration of IBP (Calin criteria), normal plain x-ray of LS and SIJs and clinical suspicion of SpA	Clinical diagnosis	Presence of SIJ/spinal lesions	SIJs, lumbar spine	1

Abbreviations: BASDAI Bath Ankylosing Spondylitis Disease Activity Index; CBP Chronic Back pain; SIJ Sacroiliac joint; SPARCC Spondyloarthritis Research Consortium of Canada; VAS Visual Analogue Score; SIJ, sacroiliac joint; SpA, Spondyloartrhitis; axSpA, axial spondyloarthritis; nr-axSpA, non-radiographic axial spondyloarthritis.

Table S6 – Technical aspects (n=4).

Study ID	n	n axSpA	n AS/nr-axSpA	n Controls	Design	Inclusion criteria	Reference standard	Feature of interest	Anatomical coverage	RoB /14
Boy et al. (34) 6.	45	31		14	Retro	Chronic LBP for > 3 months without a confirmed diagnosis of SpA. Age 18–50 years and age of onset earlier than 45 years.	Clinical diagnosis (ASAS)	Oedema (comparison of sequences)	SIJs	0
Ozgen et al. (35) H1	73	34		39	Pro	Clinically suspected sacroiliitis	Radiological diagnosis (global assessment)	Oedema (sensitivity, contrast-to-noise ratio)	SIJs	N/A
van Onna (38) 23.	68	NR		NR	Pro	Inflammatory back pain (ESPAC cohort)	Radiological diagnosis (global assessment using STIR and post-gad T1w)	Proportion of patients with bone marrow oedema on STIR/gad/both	SIJs	0
de Hooge et al. (39) 17.	127	25		102	Pro	Back pain (SPACE cohort)	Positive MRI (ASAS criteria)	Added value of gadolinium	SIJs	4

Abbreviations: ESPAC Early SpA clinical Cohort; LBP Lower back pain; SIJ sacroiliac joint; SPACE Spondyloarthritis caught early

5. Risk of Bias Assessment

The consensus RoB from the two assessors (TJPB and AJ) are shown below. Domains with high, unclear and low risk of bias are shown in green, yellow and red respectively. Where the risk of bias is unclear or high, the numbers refer to the specific reasons for the bias allocation, as detailed after the tables.

Table S7 - RoB for Q1 (sensitivity and specificity of MRI lesions)

	Patient selection		Index test		Reference standard		Flow and timing	
Study ID	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	Total RoB score
Arnbak et al. (18) 16.	1	1						3
Bennett et al. (11) M1	2							1
Bennett et al. (12) M2	2							1
de Hooge et al. (9) 2.					3			2
Hu et al. (21) 5.								0
Hu et al. (23) 8.	4			5	3			5
Jacquemin et al. (19) 28.								0
Jans et al. (22) 7.					3			2
Kim et al. (17) M8	4	6				6		3
Laloo et al. (8) 1.					3		7	3
Larbi et al. (27) 22.					3		8	3
Weber et al. (25) 11.	2							1
Weber et al. (26) 19.	2				3			3
Weber et al. 2015 (10) 29.	2							1
Weber et al. (13) M3	2							1

Weber et al. (14) M4	2	6				6		3
Weber et al. (15) M5	2	6				6		3
Weber et al. (20) 4.	2							1
Weber et al. (24) 10.								0
Wick et al. (16) M7					9		10	2

Table S8 - RoB for Q2 (effect of anatomical coverage)

	Patient selection		Index test		Reference standard		Flow and timing	
Study ID	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	Total RoB score
Weber et al. (31) 9.	2							1
Althoff et al. (30) 18.	11	11						2
van der Heijde et al. (32) 20.	11	11				11		4
Larbi et al. (33) 27.								0
Marzo-Ortega et al. (28) M6			12					1

Table S9 - RoB for Q3 (effect of acquisition parameters)

	Patient selection		Index test		Reference standard		Flow and timing	
Study ID	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	Applicability concerns	Risk of bias	Total RoB score
Boy et al. (34) 6.								0
Ozgen et al. (35) H1								N/A
Dalto et al. (36) H2								N/A
Hermann et al. (37)								N/A
van Onna (38) 23.								0
de Hooge et al. (39) 17.					2		2	4

1. ‘Selected’ data from back pain cohort based on previous MRI results may introduce bias and applicability concerns.
2. Addition of healthy controls may bias sensitivity and specificity statistics when performance diagnostic accuracy studies (for further detail see the QUADAS-2 tool (47)).
3. Reference standard not clearly blinded to index test.
4. Case control design
5. Much larger disease population than controls may bias sensitivity and specificity characteristics or be unrealistic for real-world situations
6. All patients had AS or AS-heavy cohort
7. Diagnostic pathways unclear, possibly inconsistent reference standard
8. Interval between diagnosis and MRI is unclear – long interval may lead to bias/inaccuracy
9. Unclear whether clinicians were blinded to MRI for diagnosis
10. Unclear whether clinical diagnosis was made before or after MRI scan
11. Severe cohort causing bias or applicability concerns
12. Blinding not specified
13. Simultaneous scoring of different image types may increase risk of bias

6. Acquisition Parameters

Figure S2 – Acquisition parameters. Method of fat suppression, spine coverage and use of contrast are shown for all reported studies. NB: A small number of studies also used additional methods of fat suppression such as T2W-Dixon imaging, which are not shown here.

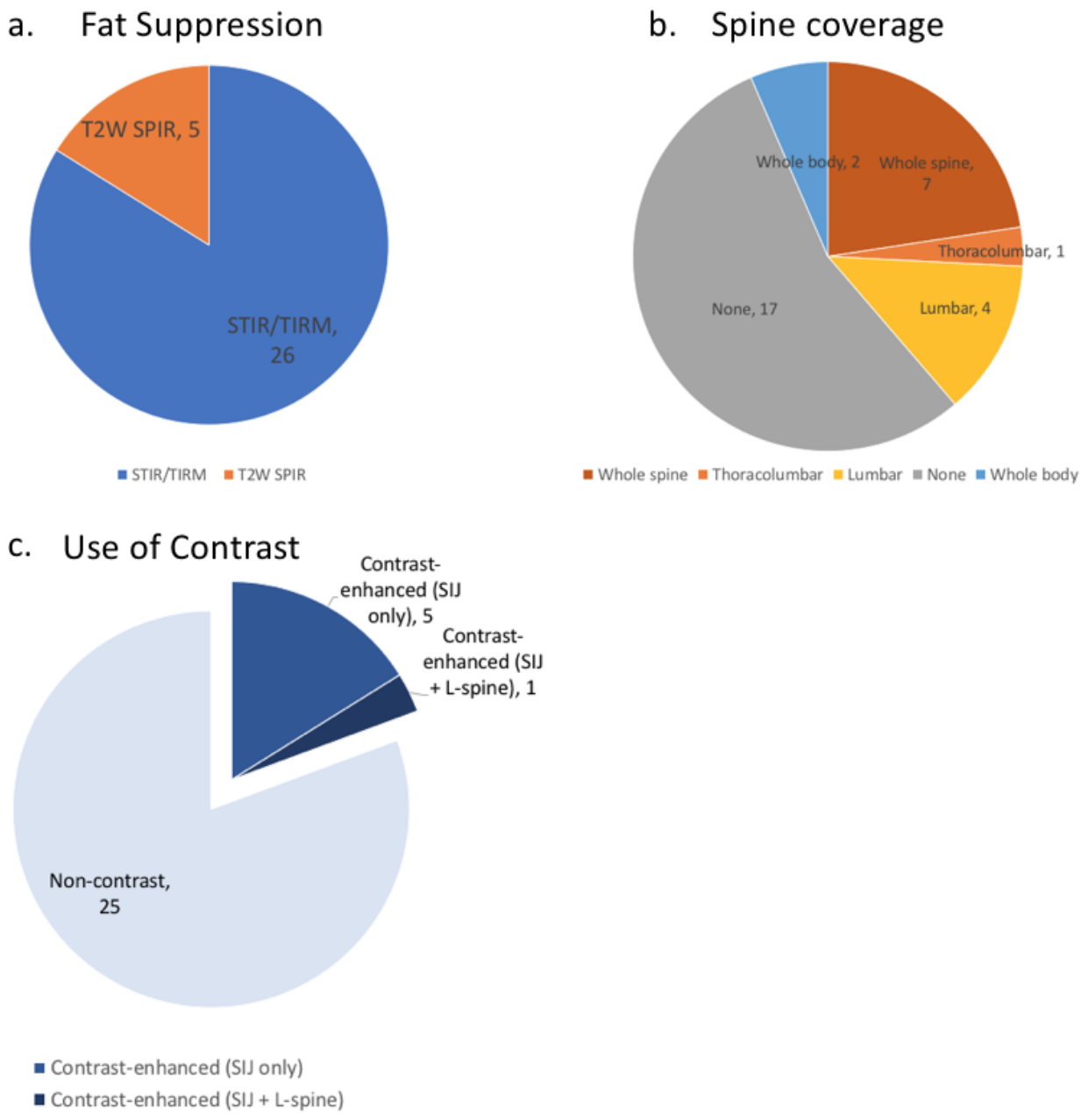


Figure S3 – Sequence parameters for STIR imaging. Inversion time, echo time, and repetition time are shown for all included studies with complete sequence information (n=16). All studies were performed on 1.5T systems.

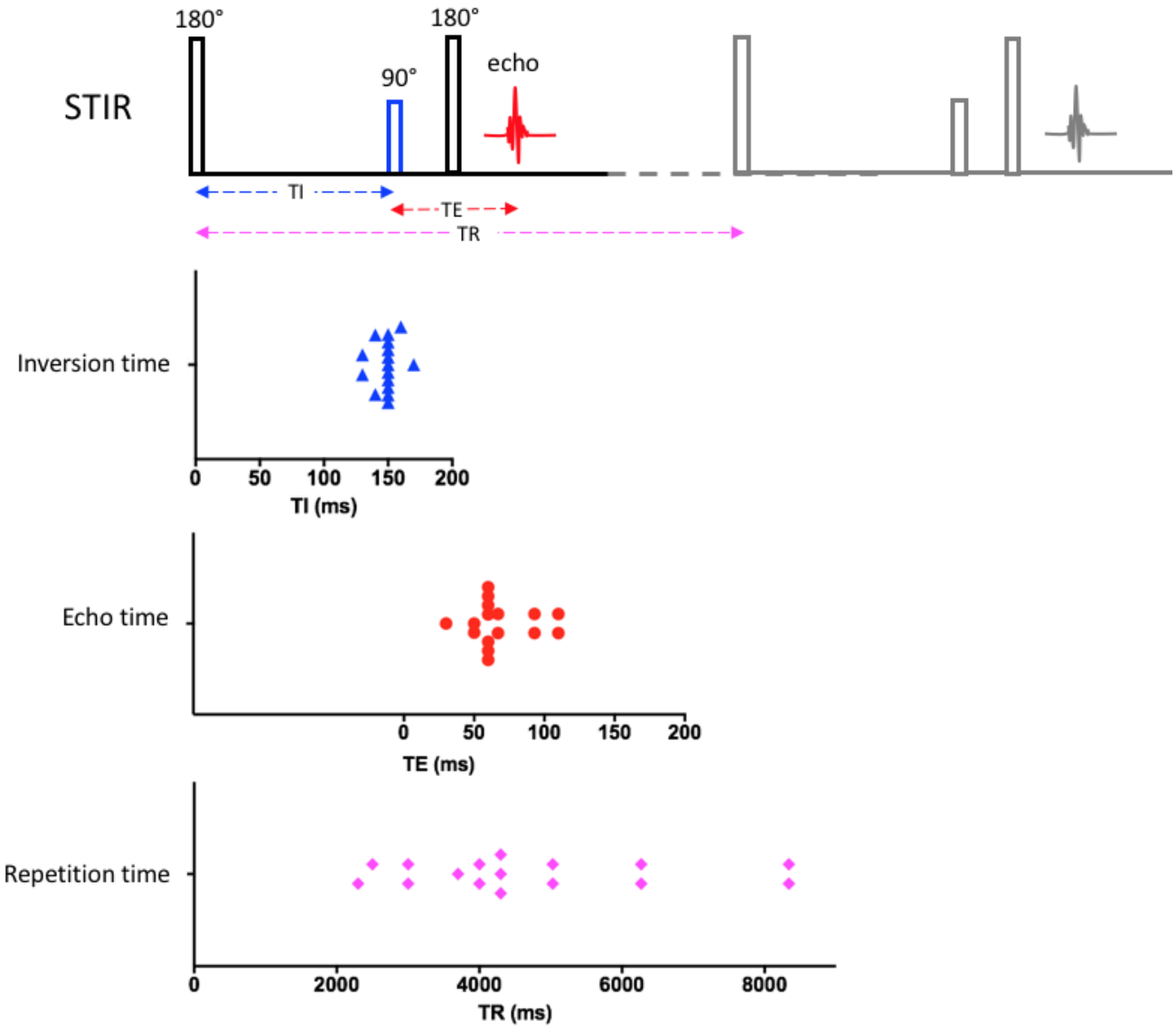
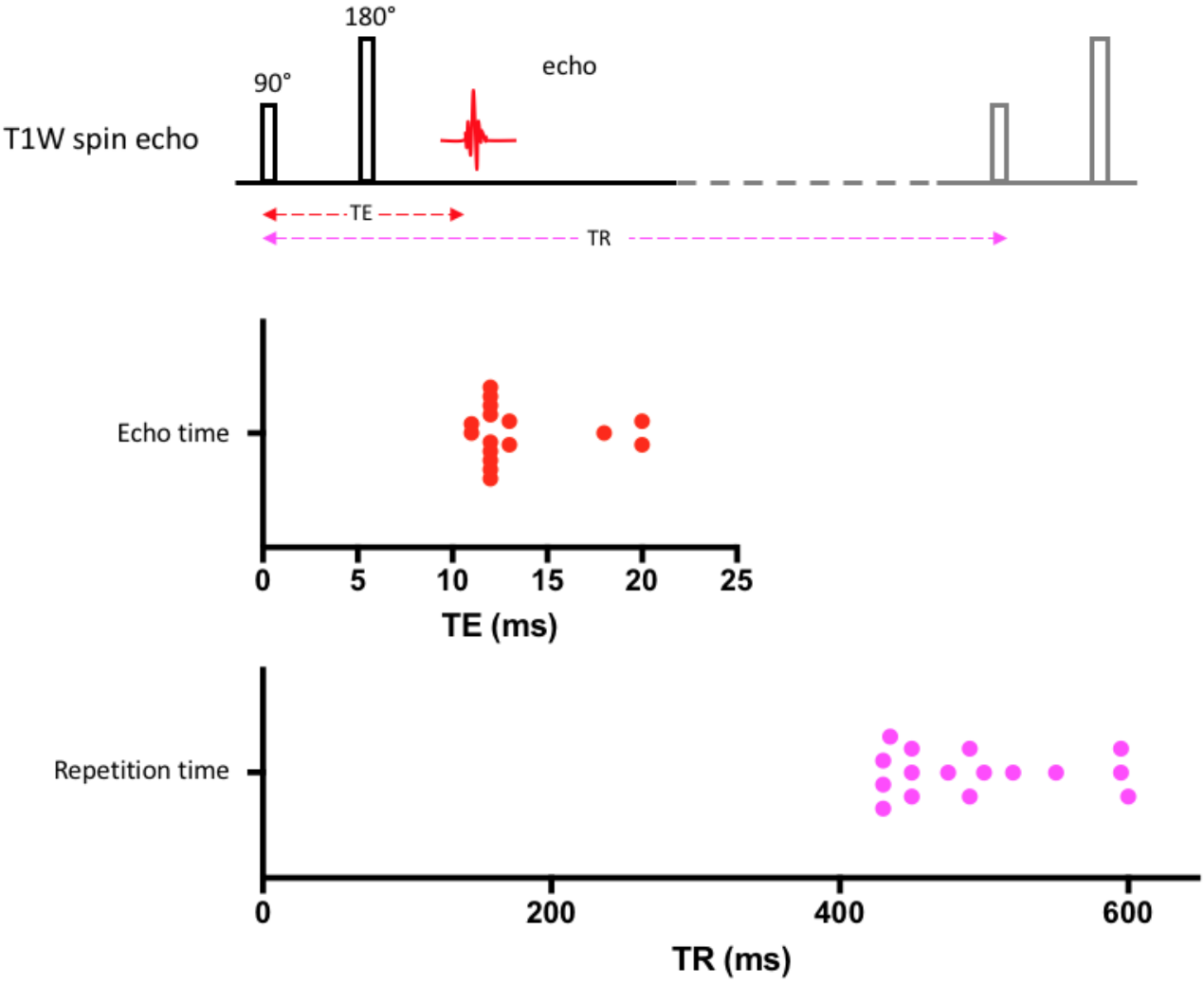


Figure S4 - Sequence parameters for T1W imaging. Echo time and repetition time are shown for all included studies with complete sequence information (n=16). All studies were performed on 1.5T systems.



References

1. Sieper J, Poddubnyy D. Axial spondyloarthritis. Lancet (London, England) [Internet]. 2017 Jan 19 [cited 2017 May 3]; Available from: http://linkinghub.elsevier.com/retrieve/pii/S0140673616315914

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 [Internet]. 2009 Jun 1 [cited 2017 May 3];68(6):777–83. Available from: http://ard.bmj.com/cgi/doi/10.1136/ard.2009.108233

3. Rudwaleit M, van der Heijde D, Landewé R, Akkoc N, Brandt J, Chou CT, et al. The Assessment of SpondyloArthritis International Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. Ann Rheum Dis [Internet]. BMJ Publishing Group Ltd; 2011 Jan [cited 2017 May 3];70(1):25–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21109520

4. Puhakka KB, Melsen F, Jurik AG, Boel LW, Vesterby A, Egund N. MR imaging of the normal sacroiliac joint with correlation to histology. Skeletal Radiol. 2004;33(1):15–28.

5. Bennett AN, Marzo-Ortega H, Kaur-Papadakis D, Rehman A. The use of magnetic resonance imaging in axial spondyloarthritis: Time to bridge the gap between radiologists and rheumatologists. J Rheumatol. 2017;44(6):780–5.

6. Sackett DL, Straus SE, Richardson WS, Rosenberg WMC, Haynes RB. Evidence-based medicine: How to teach and practice EBM. Br Med J. 2000;

7. Mandl P, Navarro-Compan V, Terslev L, Aegerter P, van der Heijde D, D'Agostino M a., et al. EULAR recommendations for the use of imaging in the diagnosis and management of spondyloarthritis in clinical practice. Ann Rheum Dis. 2015;1–13.

8. Laloo F, Herregods N, Varkas G, Jaremko JL, Baraliakos X, Elewaut D, et al. MR signal in the sacroiliac joint space in spondyloarthritis: a new sign. Eur Radiol. 2017 Sep;27(5):2024–30.

9. Hooge M De, Berg R Van Den, Navarro-compán V, Reijnierse M, Gaalen F Van, Fagerli

K, et al. Patients with chronic back pain of short duration from the SPACE cohort : which MRI structural lesions in the sacroiliac joints and in fl ammatory and structural lesions in the spine are most speci fi c for axial spondyloarthritis ? Ann Rheum Dis. 2016;0:1308–14.

10. Weber U, Ostergaard M, Lambert RGW, Pedersen SJ, Chan SM, Zubler V, et al. Candidate lesion-based criteria for defining a positive sacroiliac joint MRI in two cohorts of patients with axial spondyloarthritis. Ann Rheum Dis. England; 2015 Nov;74(11):1976–82.

11. Bennett AN, Rehman A, Hensor EMA, Marzo-Ortega H, Emery P, McGonagle D. Evaluation of the diagnostic utility of spinal magnetic resonance imaging in axial spondylarthritis. Arthritis Rheum. 2009;60(5):1331–41.

12. Bennett AN, Rehman A, Hensor EMA, Marzo-Ortega H, Emery P, McGonagle D. The fatty Romanus lesion: A non-inflammatory spinal MRI lesion specific for axial spondyloarthropathy. Ann Rheum Dis. 2010;69(5):891–4.

13. Weber U, Hodler J, Kubik RA, Rufibach K, Lambert RGW, Kissling RO, et al. Sensitivity and specificity of spinal inflammatory lesions assessed by whole-body magnetic resonance imaging in patients with ankylosing spondylitis or recent-onset inflammatory back pain. Arthritis Care Res. 2009;61(7):900–8.

14. Weber U, Lambert RGW, Østergaard M, Hodler J, Pedersen SJ, Maksymowych WP. The diagnostic utility of magnetic resonance imaging in spondylarthritis: An international multicenter evaluation of one hundred eighty-seven subjects. Arthritis Rheum. 2010;62(10):3048–58.

15. Weber U, Lambert RGW, Pedersen SJ, Hodler J, Østergaard M, Maksymowych WP. Assessment of structural lesions in sacroiliac joints enhances diagnostic utility of magnetic resonance imaging in early spondylarthritis. Arthritis Care Res (Hoboken). 2010;62(12):1763–71.

16. Wick MC, Weiss RJ, Jaschke W, Klauser AS. Erosions are the most relevant magnetic resonance imaging features in quantification of sacroiliac joints in ankylosing spondylitis. J Rheumatol. 2010;37:622–7.

17. Na RK, Choi JY, Sung HH, Woo SJ, Joon WL, Choi JA, et al. “MR corner sign”: Value for predicting presence of ankylosing spondylitis. Am J Roentgenol. 2008;191(1):124–8.

18. Arnbak B, Jensen TS, Manniche C, Zejden A, Egund N, Jurik AG. Spondyloarthritis-related and degenerative MRI changes in the axial skeleton - An inter- and intra-observer agreement study. BMC Musculoskelet Disord. 2013;14:274.

19. Jacquemin C, Rubio Vargas R, van den Berg R, Thevenin F, Lenczner G, Reijnierse M, et al. What is the reliability of non-trained investigators in recognising structural MRI lesions of sacroiliac joints in patients with recent inflammatory back pain? Results of the DESIR cohort. RMD open. England; 2016;2(2):e000303.

20. Weber U, Pedersen SJ, Zubler V, Rufibach K, Chan SM, Lambert RGW, et al. Fat infiltration on magnetic resonance imaging of the sacroiliac joints has limited diagnostic utility in nonradiographic axial spondyloarthritis. J Rheumatol. Canada; 2014 Jan;41(1):75–83.

21. Hu Z, Xu M, Lin Z, Liao Z, Deng X, Zhang Y, et al. Limitations of lumbar spine MRI in the diagnosis of ankylosing spondylitis. Clin Exp Rheumatol. Italy; 2014;32(1):22–7.

22. Jans L, Coeman L, Van Praet L, Carron P, Elewaut D, Van den Bosch F, et al. How sensitive and specific are MRI features of sacroiliitis for diagnosis of spondyloarthritis in patients with inflammatory back pain? JBR-BTR. 2014;97(4):202–5.

23. Hu Z, Wang X, Qi J, Kong Q, Zhao M, Gu J. Backfill is a specific sign of axial spondyloarthritis seen on MRI. Joint Bone Spine. France; 2016 Mar;83(2):179–83.

24. Weber U, Maksymowych WP, Chan SM, Rufibach K, Pedersen SJ, Zhao Z, et al. Does evaluation of the ligamentous compartment enhance diagnostic utility of sacroiliac joint MRI in axial spondyloarthritis? Arthritis Res Ther. England; 2015 Sep;17:246.

25. Weber U, Zhao Z, Rufibach K, Zubler V, Lambert RGW, Chan SM, et al. Diagnostic utility of candidate definitions for demonstrating axial spondyloarthritis on magnetic resonance imaging of the spine. Arthritis Rheumatol. 2015;67(4):924–33.

26. Weber U, Zubler V, Pedersen SJ, Rufibach K, Lambert RGW, Chan SM, et al. Development and validation of a magnetic resonance imaging reference criterion for defining a positive sacroiliac joint magnetic resonance imaging finding in

spondyloarthritis. Arthritis Care Res (Hoboken). United States; 2013 Jun;65(6):977–85.

27. A. L, P. V, N. M, C. L, M.P. B, P. T. Assessment of MRI abnormalities of the sacroiliac joints and their ability to predict axial spondyloarthritis: A retrospective pilot study on 110 patients. Vol. 43, Skeletal Radiology. C. Cyteval, Medical Imaging Department, Hopital Lapeyronie, 371 avenue du Doyen Gaston Giraud, 34295 Montpellier Cedex 5, France. E-mail: c-cyteval@chu-montpellier.fr: Springer Verlag (Tiergartenstrasse 17, Heidelberg D-69121, Germany); 2014. p. 351–8.

28. Marzo-Ortega H, McGonagle D, O'Connor P, Hensor EMA, Bennett AN, Green MJ, et al. Baseline and 1-year magnetic resonance imaging of the sacroiliac joint and lumbar spine in very early inflammatory back pain. Relationship between symptoms, HLA-B27 and disease extent and persistence. Ann Rheum Dis. 2009;68(11):1721–7.

29. Hu Z, Xu M, Lin Z, Liao Z, Deng X, Zhang Y, et al. Limitations of lumbar spine MRI in the diagnosis of ankylosing spondylitis. Clin Exp Rheumatol. 2014;

30. Althoff CE, Sieper J, Song I-H, Haibel H, Weiss A, Diekhoff T, et al. Active inflammation and structural change in early active axial spondyloarthritis as detected by whole-body MRI. Ann Rheum Dis. England; 2013 Jun;72(6):967–73.

31. Weber U, Zubler V, Zhao Z, Lambert RGW, Chan SM, Pedersen SJ, et al. Does spinal MRI add incremental diagnostic value to MRI of the sacroiliac joints alone in patients with non-radiographic axial spondyloarthritis? Ann Rheum Dis. England; 2015 Jun;74(6):985–92.

32. van der Heijde D, Sieper J, Maksymowych WP, Brown MA, Lambert RGW, 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 (Hoboken, NJ). United States; 2014 Mar;66(3):667–73.

33. Larbi A, Fourneret B, Lukas C, Baron M-P, Molinari N, Taourel P, et al. Prevalence and topographic distribution of spinal inflammation on MR imaging in patients recently diagnosed with axial spondyloarthritis. Diagn Interv Imaging. France; 2016 Nov;

34. Boy FN, Kayhan A, Karakas HM, Unlu-Ozkan F, Silte D, Aktas I. The role of multi-parametric MR imaging in the detection of early inflammatory sacroiliitis according to ASAS criteria. Eur J Radiol. Ireland; 2014 Jun;83(6):989–96.

35. Ozgen A. The Value of the T2-Weighted Multipoint Dixon Sequence in MRI of Sacroiliac Joints for the Diagnosis of Active and Chronic Sacroiliitis. Am J Roentgenol. 2016;

36. Dalto VF, Assad RL, Crema MD, Louzada-Junior P, Nogueira-Barbosa MH, V.F. D, et al. MRI assessment of bone marrow oedema in the sacroiliac joints of patients with spondyloarthritis: is the SPAIR T2w technique comparable to STIR? Eur Radiol. V.F. Dalto, Division of Radiology, Department of Internal Medicine, Ribeirao Preto Medical School, University of Sao Paulo, Avenida Bandeirantes, 3900 - Monte Alegre, Ribeirao Preto, SP Zip Code: 14048-900, Brazil. E-mail: fdalto@gmail.com, Germany: Springer Verlag (E-mail: service@springer.de); 2017 Feb;1–8.

37. Hermann KGA, Landewé RBM, Braun J, Van Der Heijde DMFM. Magnetic resonance imaging of inflammatory lesions in the spine in ankylosing spondylitis clinical trials: Is paramagnetic contrast medium necessary? In: Journal of Rheumatology. 2005. p. 2056–60.

38. van Onna M, van Tubergen A, van der Heijde D, Jurik AG, Landewe R. Gadolinium contrast-enhanced MRI sequence does not have an incremental value in the assessment of sacroiliitis in patients with early inflammatory back pain by using MRI in combination with pelvic radiographs: a 2-year follow-up study. Clin Exp Rheumatol. Italy; 2014;32(2):225–30.

39. de Hooge M, van den Berg R, Navarro-Compan V, van Gaalen F, van der Heijde D, Huizinga T, et al. Magnetic resonance imaging of the sacroiliac joints in the early detection of spondyloarthritis: no added value of gadolinium compared with short tau inversion recovery sequence. Rheumatology (Oxford). England; 2013 Jul;52(7):1220–4.

40. de Winter J, de Hooge M, van de Sande M, de Jong H, van Hoeven L, de Koning A, et al. Magnetic Resonance Imaging of the Sacroiliac Joints Indicating Sacroiliitis According to the Assessment of SpondyloArthritis international Society Definition in

Healthy Individuals, Runners, and Women With Postpartum Back Pain. Arthritis Rheumatol. 2018;

41. Varkas G, De Hooge M, Renson T, De Mits S, Carron P, Jacques P, et al. Effect of mechanical stress on magnetic resonance imaging of the sacroiliac joints: Assessment of military recruits by magnetic resonance imaging study. Rheumatol (United Kingdom). 2018;

42. Weber U, Jurik AG, Zejden A, Larsen E, Jørgensen SH, Rufibach K, et al. Frequency and Anatomic Distribution of Magnetic Resonance Imaging Features in the Sacroiliac Joints of Young Athletes: Exploring “Background Noise” Toward a Data-Driven Definition of Sacroiliitis in Early Spondyloarthritis. Arthritis Rheumatol. 2018;

43. Baraliakos X, Feldmann D, Ott A, Schmidt CO, Albers M, Richter A, et al. OP0243 Prevalence of inflammatory and chronic changes suggestive of axial spondyloarthritis in magnetic resonance images of the axial skeleton in individuals <45 years in the general population as part of a large community study (SHIP). Ann Rheum Dis. BMJ Publishing Group Ltd; 2018 Jun;77(Suppl 2):170–170.

44. Arnbak B, Grethe Jurik A, Hørslev-Petersen K, Hendricks O, Hermansen LT, Loft AG, et al. Associations between Spondyloarthritis Features and Magnetic Resonance Imaging Findings: A Cross-Sectional Analysis of 1,020 Patients with Persistent Low Back Pain. Arthritis Rheumatol. 2016;68(4):892–900.

45. Bray TJP, Vendhan K, Ambrose N, Atkinson D, Punwani S, Fisher C, et al. Diffusion-weighted imaging is a sensitive biomarker of response to biologic therapy in enthesitis-related arthritis. Rheumatol (United Kingdom). 2017 Dec;56(3):399–407.

46. Bray TJP, Bainbridge A, Punwani S, Ioannou Y, Hall-Craggs MA. Simultaneous Quantification of Bone Edema/Adiposity and Structure in Inflamed Bone Using Chemical Shift-Encoded MRI in Spondyloarthritis. Magn Reson Med. 2018;79(2):1031–42.

47. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36.