



Histographic analysis of oedema and fat in inflamed bone marrow based on quantitative MRI

Timothy J. P. Bray¹ · Naomi Sakai¹ · Alexandra Dudek¹ · Corinne Fisher² · Kannan Rajesparan¹ · Andre Lopes³ · Coziana Ciurtin² · Debajit Sen² · Alan Bainbridge^{1,4} · Margaret A. Hall-Craggs^{1,4}

Received: 19 September 2019 / Revised: 21 February 2020 / Accepted: 2 March 2020
© The Author(s) 2020

Abstract

Objective To demonstrate proof-of-concept for a quantitative MRI method using histographic analysis to assess bone marrow oedema and fat metaplasia in the sacroiliac joints.

Materials and methods Fifty-three adolescents aged 12–23 with known or suspected sacroiliitis were prospectively recruited and underwent quantitative MRI (qMRI) scans, consisting of chemical shift-encoded (at 3 T) and diffusion-weighted imaging (at 1.5 T), plus conventional MRI (at 1.5 T) and clinical assessment. qMRI scans produced proton-density fat fraction (PDFF) and apparent diffusion coefficient (ADC) maps of the sacroiliac joints (SIJs), which were analysed using an in-house software tool enabling partially automated ROI definition and histographic analysis. Logistic regression and receiver operating characteristic (ROC) analyses assessed the predictive performance of ADC- and PDFF-based parameters in identifying active inflammation (oedema) and structural damage (fat metaplasia).

Results ADC-based parameters were associated with increased odds of oedema (all $p < 0.05$); ROC-AUC was higher for histographic parameters representing the upper end of the ADC distribution than for simple averages. Similarly, PDFF-based parameters were associated with increased odds of fat metaplasia (all $p < 0.05$); ROC area-under-the-curve was higher for histographic parameters representing the upper end of the PDFF distribution than for simple averages. Both ADC- and PDFF-based histographic parameters demonstrated excellent inter- and intra-observer agreement (ICC > 0.9).

Conclusions ADC-based parameters can differentiate patients with bone marrow oedema from those without, whilst PDFF-based parameters can differentiate patients with fat metaplasia from those without. Histographic analysis might improve performance compared with simple averages such as the mean and median and offers excellent agreement within and between observers.

Key Points

- *Quantitative MRI with histographic analysis can identify bone marrow oedema (an active inflammatory lesion) and fat metaplasia (a 'chronic' inflammatory lesion) in patients with spondyloarthritis.*
- *The use of histographic analysis might improve the performance of quantitative MRI for detecting bone marrow oedema and fat metaplasia compared with simple averages such as the mean and median.*
- *Bone marrow oedema and fat metaplasia are known to be of diagnostic and prognostic significance, and the proposed method could support clinical decisions around biologic (and other) therapies in spondyloarthritis.*

Keywords Magnetic resonance imaging · Inflammation · Arthritis · Spine

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00330-020-06785-x>) contains supplementary material, which is available to authorized users.

✉ Margaret A. Hall-Craggs
margaret.hall-craggs@nhs.net

¹ Centre for Medical Imaging, University College London, London, UK

² Arthritis Research UK Centre for Adolescent Rheumatology, University College London, London, UK

³ Cancer Research UK & University College London Cancer Trials Centre, London, UK

⁴ Medical Physics Department, University College London Hospitals, London, UK

Abbreviations

ADC	Apparent diffusion coefficient
AUC	Area under the curve
BASDAI	Bath Ankylosing Spondylitis Disability Index
BASFI	Bath Ankylosing Spondylitis Functional Index
BEACH	Bone edema and adiposity characterisation
BRC	Biomedical Research Centre
CRP	C-reactive protein
CSE-MRI	Chemical shift-encoded MRI
DWI	Diffusion-weighted imaging
ESR	Erythrocyte sedimentation rate
MRI	Magnetic resonance imaging
NIHR	National Institute for Health Research
OR	Odds ratio
PACS	Picture archiving and communications system
PDFF	Proton density fat fraction
qMRI	Quantitative MRI
ROC	Receiver operating characteristic
ROI	Region-of-interest
SIJ	Sacroiliac joint
SpA	Spondyloarthritis
STIR	Short inversion time inversion recovery
TNFi	Tumour necrosis factor inhibitor

Introduction

Spondyloarthritis (SpA) encompasses a group of immune-mediated inflammatory diseases characterised by spinal pain, stiffness and damage which commonly affect young people and have poor long-term health outcomes [1]. Diagnosis of SpA is often difficult due to the complex nature of pain in young patients [2], and delays in diagnosis and treatment are common [3]. Identification of bone marrow oedema on MRI is of importance for showing inflammation of the sacroiliac joints and supports diagnosis of axial SpA [4–7]. This directly influences the decision to treat patients with disease-modifying or biologic drugs [5].

Unfortunately, the definition of active inflammation on MRI is based on subjective criteria and is heavily dependent on the expertise and opinion of the scan reader [8–11]. ‘Conventional’ MR images used to detect inflammation—typically short inversion time inversion recovery (STIR) and T1-weighted spin echo images—produce complex image contrast that depends on multiple tissue properties, including T1, T2, proton density, perfusion and diffusion [11–13], which may confound the identification and quantification of oedema. These factors can lead to a lack of consistency between observers and scanners/hospitals [7, 14]. Therefore, there is a need for a method which can simply and objectively assess skeletal inflammation on MRI scans to support diagnostic and therapeutic decisions.

Previous studies have investigated the use of diffusion-weighted imaging (DWI) and chemical shift-encoded MRI (CSE-MRI) as objective methods for assessing bone marrow oedema, with promising initial results [12, 13, 15]. Using DWI, it has been shown that apparent diffusion coefficient (ADC) measurements are increased in areas of marrow oedema, probably due to an expansion of the extracellular space [13, 15–17]. Using CSE-MRI, it has been shown that proton density fat fraction (PDFF) measurements are reduced in areas of oedema compared with normal marrow, due to increased water content [12]. CSE-MRI can also be used to assess the severity of fat metaplasia, defined as a focal increase in content in areas of previous inflammation (with diagnostic and prognostic significance), in a quantitative fashion [12]. Previous studies measuring ADC in subchondral bone have typically relied on manual placement of regions-of-interest (ROIs) within the subchondral bone [15, 16, 18] which introduces substantial methodological subjectivity. Furthermore, these studies have relied on *mean* ADC measurements, which may perform poorly in patients with mixed active and chronic inflammation due to neutralisation of opposing effects [19]. There is currently no validated tool for quantifying proton density fat fraction (PDFF) in the sacroiliac joints.

We describe a new analysis tool which enables a more complete and consistent assessment of subchondral bone and derives a series of histographic parameters from both ADC and PDFF maps, aiming to isolate and separately quantify the active and chronic components of the inflammatory process. We aimed to demonstrate proof-of-principle for this tool in a prospective study of young people with SpA.

Methods

This study received ethical approval from the Queen Square Research Ethics Committee, London, UK (Research Ethics Committee reference 15/LO/1475). All participants gave written informed consent prior to study entry.

Study design and participants

A prospective cross-sectional study was performed at a single specialist tertiary referral centre for adolescents and young adults with inflammatory arthritis. Fifty-three consecutive patients meeting the eligibility criteria (mean age, 18 years; age range, 12–23 years) were prospectively recruited between July 2016 and December 2018 (31 males, mean age 18 years, and 22 females, mean age 17 years). Patients were included if they were referred for an MRI scan of the sacroiliac joints for suspicion of sacroiliitis or for monitoring of known sacroiliitis and were excluded if they had a contraindication to MRI scanning. All patients with known, pre-existing sacroiliitis had a clinical diagnosis of either non-radiographic axial SpA or

enthesitis-related arthritis [18–21]. The sample size was fixed and based on logistical constraints. Patients were classified according to the presence or absence of bone marrow oedema and fat metaplasia using established criteria, based on conventional MRI scans, as described below.

Image acquisition

All subjects underwent both quantitative and conventional MRI scans on the same visit. Quantitative CSE-MR images were acquired on a 3-T Philips Ingenia scanner (Ingenia, Philips) using an investigational version of the Philips mDixon Quant acquisition and post-processing pipeline, as described previously [12]. The images were acquired using a multi-echo gradient echo acquisition with bipolar readout (TE_1 1.17 ms, ΔTE 1.6 ms, TR 25 ms, flip angle 3° , matrix size 320×320 , pixel spacing 1.76×1.76 mm, bandwidth 394 Hz/Px) and PDFF maps were generated using complex fitting incorporating $T2^*$ decay and a 10-peak model of human adipose tissue [12]. Images were acquired coronal to the long axis of the sacroiliac joint [12]. DW images were acquired on a 1.5-T Siemens Avanto scanner (Avanto, Siemens) using b values of 0, 50, 100, 300 and 600 s/mm^2 with spectrally attenuated inversion recovery (SPAIR) fat suppression and echo planar imaging readout ($TE = 89$ ms, $TR = 3600$ ms, 4 averages, 8 mm slices, matrix size 144×192 , FOV 237×316 mm, bandwidth 1447 Hz/Px), with images acquired axial to the sacroiliac joint [13, 15]. Conventional MRI consisted of T2-weighted STIR images, T1-weighted turbo spin echo images and fat-suppressed post-contrast T1-weighted turbo spin images acquired coronal to the sacroiliac joint (see [Supplementary Information](#) for sequence parameters) [11, 12].

Image analysis

Histographic parameters were obtained from the PDFF and ADC maps used an in-house software tool known as BEACH (Bone Edema and Adiposity Characterisation with Histograms) as shown in Figs. 1, 2 and 3, and as described in detail in the [Supplementary Information](#). This method generates a series of histographic parameters for both ADC and PDFF.

The BEACH tool operates as follows. The observer is prompted to define the line of the sacroiliac joint using a single series of connected straight lines—an open polygon (Fig. 1). ‘Anchor lines’ are used to define the angle made by the joint with the cortical surface, at both the top and bottom of the joint, enabling the shape of the polygonal ROIs to be closely matched to subchondral bone. The software automatically generates a pair of polygonal ROIs in the subchondral bone either side of the joint (Fig. 1, [Supplementary Figure S1](#)). This is repeated for both sacroiliac joints covering the entire fibrocartilaginous part of the joint. For the ADC maps, all slices where the fibrocartilaginous joint was visible were

included, whereas alternate slices were used for the PDFF maps due to the smaller slice thickness. For each patient, pixel values from the total volume of defined subchondral bone (i.e. from all ROIs) are analysed histographically. For both PDFF and ADC, we measured the 10th, 25th, 50th, 75th, and 90th centiles of the distribution (designated PDFF₁₀, PDFF₂₅... and ADC₁₀, ADC₂₅... etc., as shown in Figs. 2 and 3). For each quantitative score, the mean of the two observers’ measurements was used for analysis.

The BEACH analysis was performed by two radiology residents (NS and AD, with 2 and 1 year of experience in MR imaging) who received a detailed training session (from TB). Both residents were blinded to all clinical information and to the qualitative radiological scores.

Scoring of conventional MRI

Each subject’s set of conventional MR images was scored by two experienced musculoskeletal radiologists (KR and MHC) with 10 and over 25 years of MRI experience, both blinded to clinical diagnosis, to treatment and to the quantitative image data. Images were read on a research workstation. For each patient, observers assigned a qualitative score between 0 and 72 for the extent/severity of bone marrow oedema [22]. The patient was deemed to have active inflammation if the mean bone marrow oedema score from the two readers was ≥ 2 , as per the Assessment of SpondyloArthritis Internal Society (ASAS) criteria [9, 23, 24]. Structural lesions consisting of fat metaplasia, erosions and joint ankylosis were assessed using a structural visual scoring system [25]. Patients with a score of ≥ 3 were deemed to be positive for the presence of fat metaplasia [25, 26].

Clinical scores

Symptoms were assessed using a dedicated research questionnaire (see [Supplementary Information](#)). We report here the Bath Ankylosing Spondylitis Disability Index (BASDAI) and Bath Ankylosing Spondylitis Functional Index (BASFI), in addition to C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR).

Statistical analysis

Quantitative parameters derived from ADC and PDFF maps were compared between groups with and without inflammation/fat metaplasia using logistic regression ($\alpha = 0.05$) and receiver-operating characteristic analyses. The optimal operating point for the ROC analysis was defined as the cut-point with the highest positive likelihood ratio (LR+) producing sensitivity and specificity greater than 70%. ROC-AUC values for percentile measurements were compared against the median using the method of DeLong et al [27], implemented using the *roccomp* function in Stata ($\alpha = 0.05$).

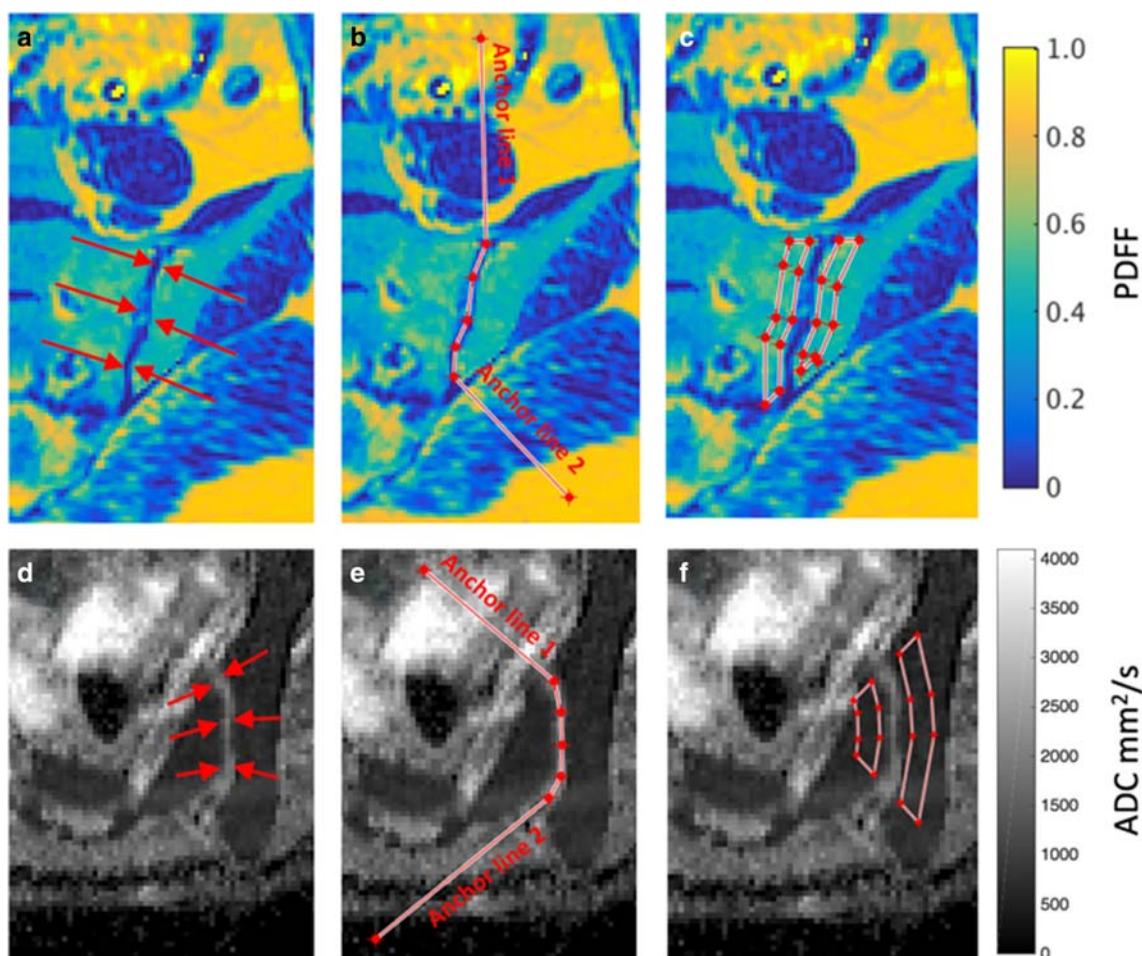


Fig. 1 Definition of polygonal ROIs on subchondral bone. The observer is asked to define the line of the sacroiliac joint and ‘anchor lines’ are added to define the angles made by the joint with the anterior and

posterior cortex of the bone, thus enabling the automatically propagated ROIs to better fit the subchondral bone

To evaluate whether combinations of parameters could improve prediction, multiple logistic regression was performed using combinations of ADC-based and PDFF-based parameters. Likelihood ratio testing was used to test whether combinations of explanatory variables provided an improved fit. Linear regression was used to evaluate the relationship between the qualitative scores and the best-performing qMRI parameters from the ROC analysis. Spearman correlation was used to evaluate the relationship between clinical scores and radiological scores. Inter- and intra-observer variability was assessed using the Bland-Altman 95% limits of agreement and the intra-class correlation coefficient.

Results

Detection of inflammation

Fifteen of 53 patients (24.5%) had sufficient bone marrow oedema to meet the ASAS criteria for active inflammation.

The inflamed group included 12 males and 3 females. The gender difference between the inflamed and uninflamed groups was not significant ($p = 0.065$). There was no significant age difference between the inflamed and uninflamed groups ($p = 0.43$).

Comparisons of quantitative parameters between inflamed and uninflamed SIJs are shown in Figs. 4 and 5, and the results of the corresponding logistic regression and ROC analyses are shown in Table 1.

All ADC-based parameters were associated with significantly increased odds of inflammation. Parameters which sampled the upper end of the ADC distribution (i.e. ADC_{75} and ADC_{90}) performed best for distinguishing inflamed from uninflamed SIJs; ADC_{90} produced an AUC value of 0.819 (0.676–0.962; $p = 0.072$ when compared with ADC_{median}). The optimal cutoff for ADC_{90} was 986 mm^2/s (sensitivity 71.4%, specificity 81.6%). Cutoffs for ADC_{mean} , ADC_{median} and ADC_{75} did not meet pre-specified performance thresholds.

PDFF-based parameters performed poorly as measures of inflammation with no significant difference between inflamed

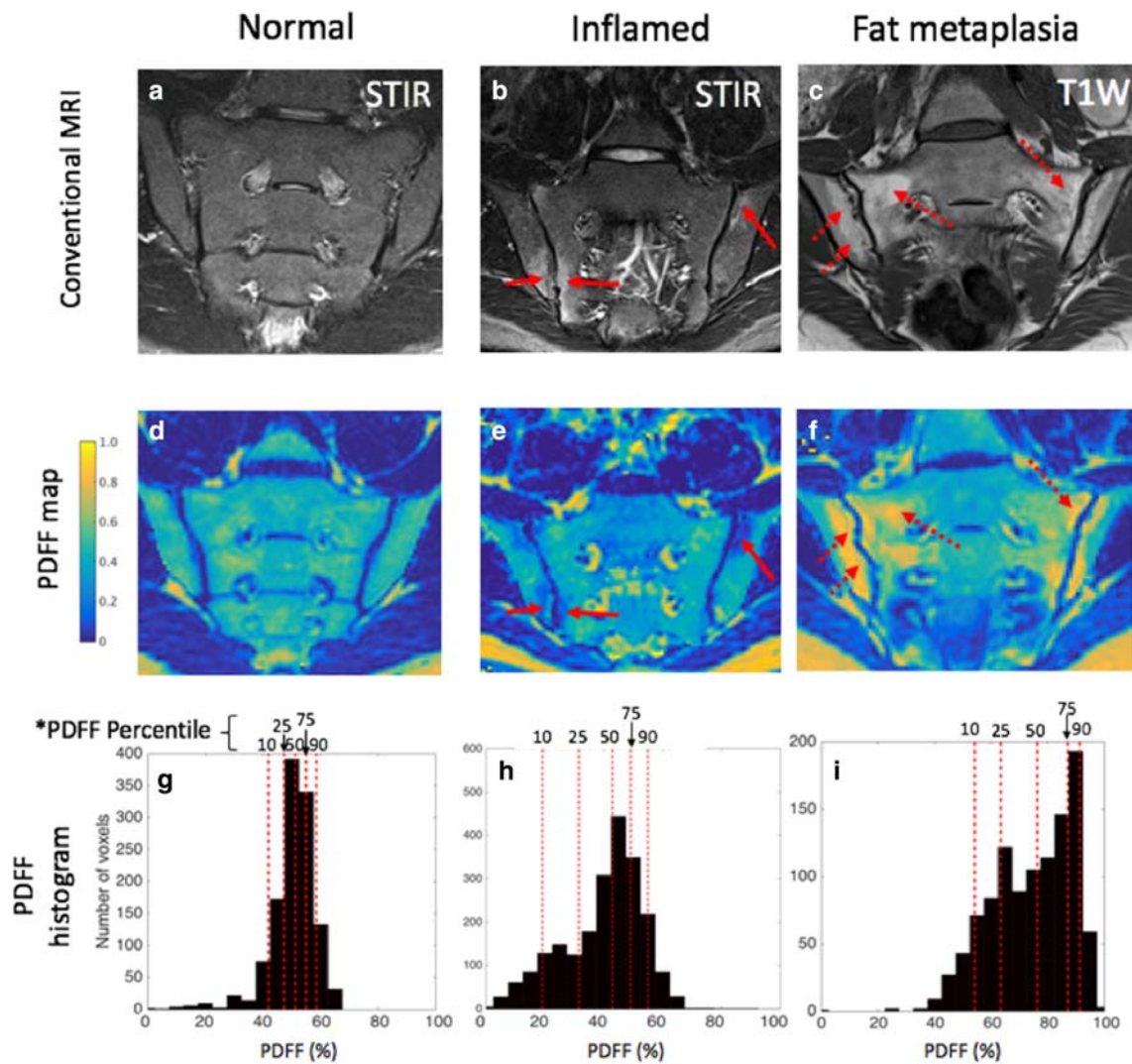


Fig. 2 Examples of histograms generated using the BEACH tool. Conventional MR images (a–c), PDFF maps (d–f) and PDFF histograms (g–i) are shown. In the normal patient’s histogram (g), PDFF values are clustered around 50%, corresponding to normal

marrow. In the patient with inflammation, a number of low-PDFF pixels have emerged in the histogram (h). In the patient with fat metaplasia, there is an upward shift in PDFF values, with a large number of high-PDFF pixels (i)

and uninflamed SIJs. Nonetheless, performance increased for parameters sampling the lower end of the distribution (AUC = 0.657 for PDFF₁₀, 0.514 for PDFF_{median}).

Detection of fat metaplasia

Thirty of 53 patients (56.6%) met the criteria for fat metaplasia. Patients with fat metaplasia were significantly older than those without fat metaplasia (mean ages (95% CI) were 19.6 (18.5–20.7) and 17.6 (16.5–18.7) respectively ($p = 0.046$)). There was no significant difference in gender between patients with and without fat metaplasia ($p = 0.56$).

Comparisons of quantitative parameters between patients with and without fat metaplasia are shown in Fig. 6, and the results of the corresponding logistic regression and ROC analyses are shown in Table 2.

PDFF-based parameters were associated with increased odds of fat metaplasia, and the separation between patients with and without fat metaplasia was improved for parameters which specifically sampled the upper end of the PDFF distribution (i.e. PDFF₇₅ and PDFF₉₀). The best performing parameter, PDFF₉₀, had an AUC of 0.780 (0.656–0.903; $p = 0.263$ when compared with PDFF_{median}). The optimal operating point for PDFF₉₀ was 55.7%, producing a sensitivity of 70% and a specificity of 73.9%.

There were no significant differences in ADC_{mean} or ADC_{median} between patients with and without fat metaplasia.

Prediction of inflammation and fat using combinations of parameters

Multiple logistic regression using both ADC₉₀ and FF₉₀ or ADC₉₀ and FF_{median} as predictor variables did not

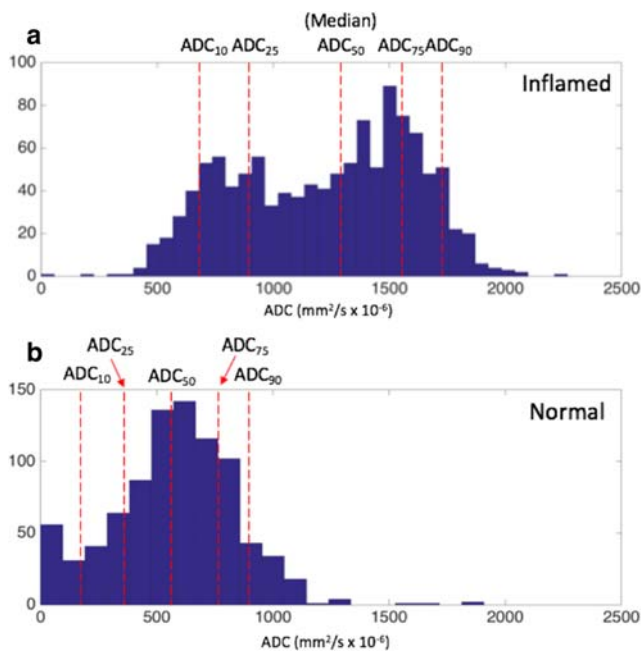
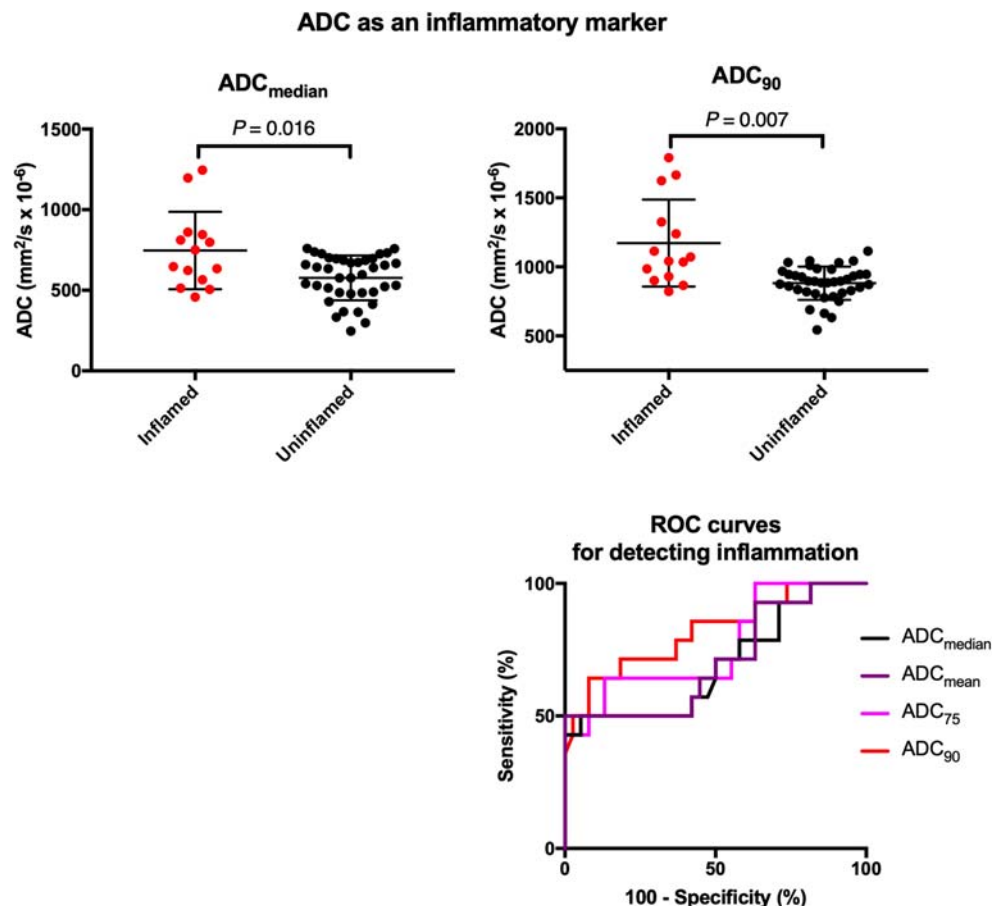


Fig. 3 Examples of ADC histograms in patients with sacroiliitis (a) and control patients (b). The red lines indicate the 10th, 25th, 50th, 75th and 90th percentiles of the ADC distribution

Fig. 4 ADC as an inflammatory marker. Representative BEACH parameters (ADC_{median} and ADC_{90}) are compared between inflamed and uninfamed groups. The displayed p values were obtained by logistic regression. ROC curves for all relevant parameters are shown in the bottom right



significantly improve the model fit compared with simple logistic regression using ADC_{90} as a single predictor ($p = 0.41$ and 0.81 , respectively). Similarly, the combination of FF_{90} and ADC_{90} or FF_{90} and ADC_{median} did not improve the model fit compared with using FF_{90} alone ($p = 0.86$ and 0.73 , respectively).

Relationship between BEACH parameters and qualitative MRI scores

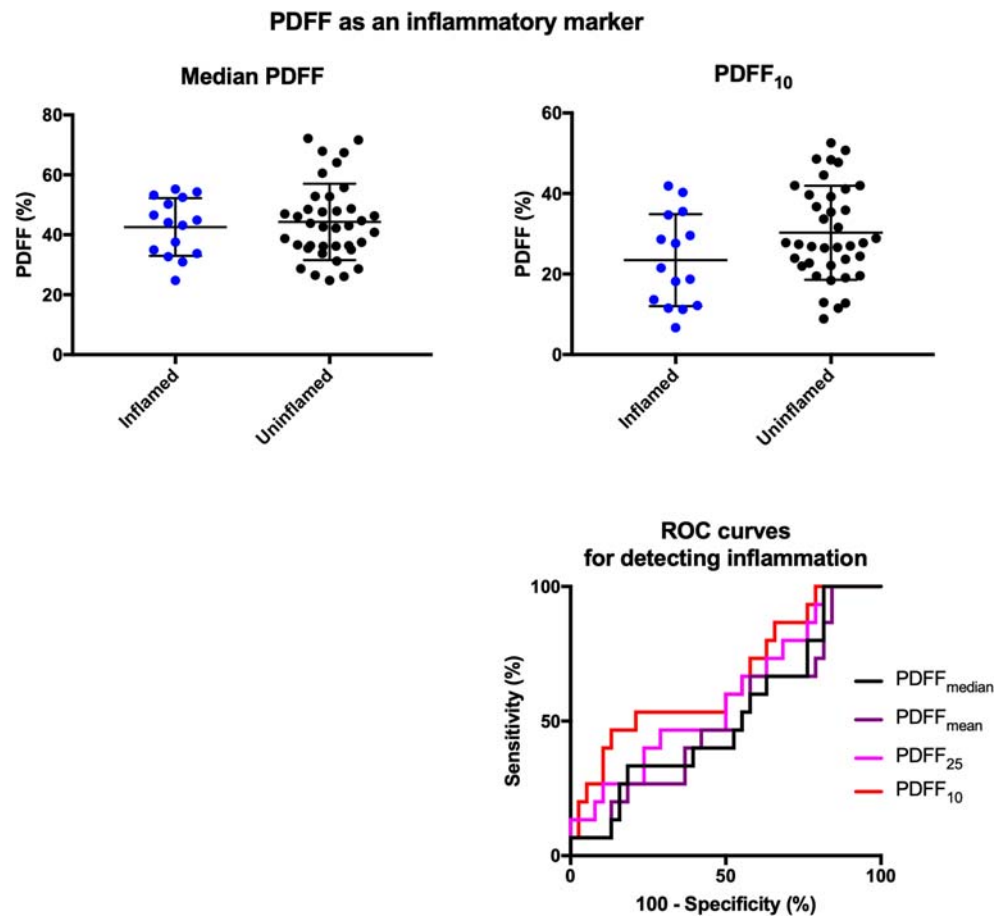
The relationship between visual scores of inflammation/fat metaplasia and qMRI parameters is shown in Supplementary Figure S2. There were significant positive relationships between ADC_{90} and the visual inflammation score (slope = 15.33 , $p < 0.0001$) and between $PDFF_{90}$ and the fat metaplasia score (slope = 1.05 , $p < 0.0001$).

Relationship between MRI and symptoms

Scatterplots showing the relationship between BASDAI scores and visual and quantitative scores of inflammation and fat metaplasia are shown in Supplementary Figure S3.

There was no significant correlation between visual scores of inflammation and any clinical score ($p = 0.45$, 0.48 , 0.14

Fig. 5 PDFF as an inflammatory marker. Representative BEACH parameters (PDFF_{median} and PDFF₁₀) are compared between inflamed and uninfamed groups. The displayed *p* values were obtained by logistic regression. ROC curves for all relevant parameters are shown in the bottom right



and 0.49 for BASDAI, BASFI, CRP and ESR) or between ADC₉₀ parameters and clinical scores (for ADC₉₀ *p* = 0.48, 0.37, 0.19 and 0.63).

There was a significant negative relationship between fat metaplasia visual scores and clinical symptoms (*p* = 0.004 and

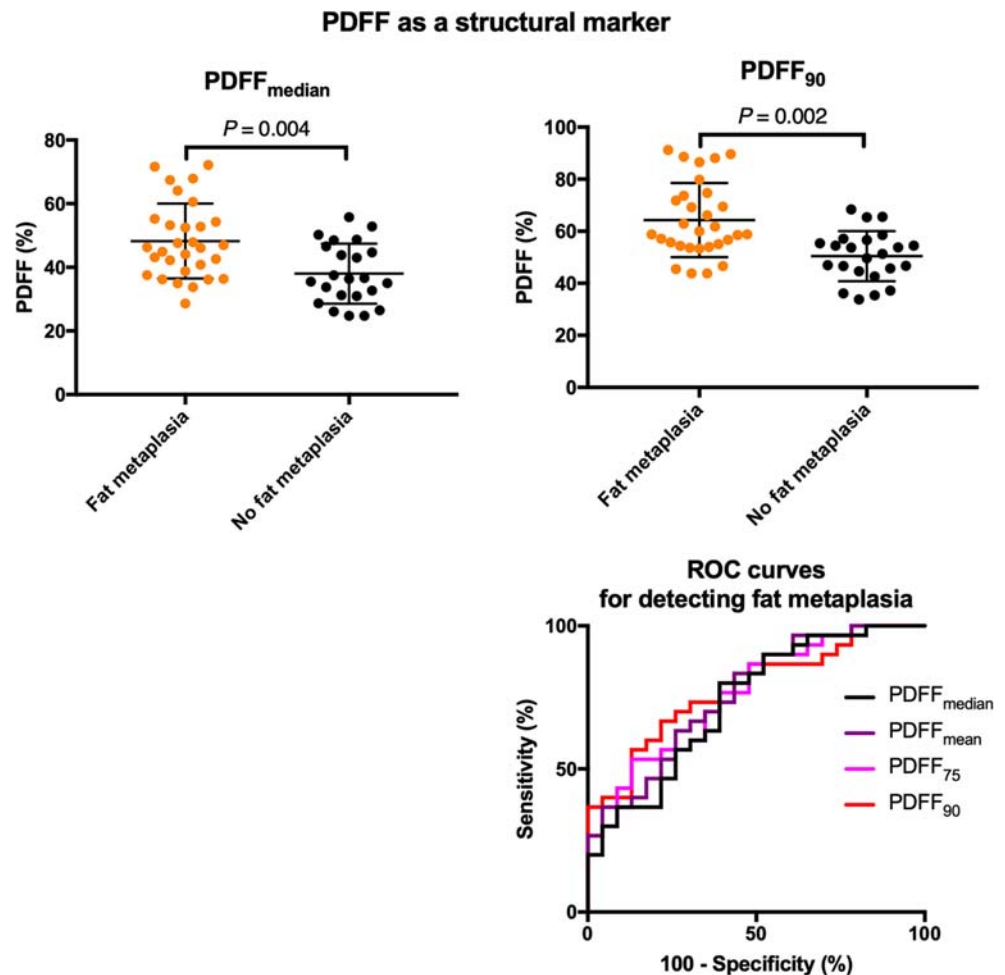
0.006 for BASDAI and BASFI), and a similar relationship was observed for the corresponding qMRI parameter PDFF₉₀ (*p* = 0.03 and 0.01 for BASDAI and BASFI). There was no significant relationship between either visual or quantitative fat metaplasia scores and CRP or ESR (all *p* > 0.05).

Table 1 Comparison of inflammatory parameters between inflamed and non-inflamed patients. ADC₇₅, ADC₉₀, etc. refer to the 75th and 90th percentiles of ADC measurements in the defined ROI. Estimates from each group are displayed as mean (95% CI). Odds ratio (OR) and *p* values (*) were derived from logistic regression. The highest ROC

AUC value for the evaluation of inflammation is shown in italics. Sensitivity and specificity values for the optimal cutoff values (far right) are provided in the main “Results” section. The right-hand *p* values (**) relate to the comparison of ROC AUC with the median value

	Inflammation		OR (95% CI)	* <i>p</i> =	ROC AUC (95% CI)	Optimal cutoff	** <i>p</i> =
	Present (<i>n</i> = 15)	Absent (<i>n</i> = 38)					
ADC _{mean}	758 (629–887)	578 (535–621)	1.006 (1.001–1.011)	0.011	0.709 (0.530–0.888)	–	0.249
ADC _{median}	747 (621–872)	577 (534–620)	1.006 (1.001–1.011)	0.016	0.695 (0.512–0.877)	–	–
ADC ₇₅	964 (813–1115)	733 (695–770)	1.008 (1.002–1.014)	0.011	0.763 (0.602–0.924)	–	0.073
ADC ₉₀	1172 (1007–1337)	882 (842–921)	1.010 (1.002–1.017)	0.007	<i>0.819</i> (0.676–0.962)	986	0.072
PDFF _{mean}	42.3 (37.2–47.4) %	43.9 (40.0–47.8) %	0.988 (0.934–1.041)	0.650	0.505 (0.329–0.682)	–	0.640
PDFF _{median}	42.6 (37.8–47.6) %	44.3 (40.2–48.4) %	0.988 (0.937–1.040)	0.639	0.514 (0.337–0.691)	–	–
PDFF ₂₅	32.8 (27.5–38.1) %	37.1 (33.4–40.8) %	0.964 (0.910–1.021)	0.215	0.589 (0.414–0.763)	–	0.010
PDFF ₁₀	23.5 (17.8–29.2) %	30.3 (26.6–34.0) %	0.948 (0.895–1.004)	0.067	0.657 (0.485–0.829)	–	0.001

Fig. 6 PDFF as a structural marker. Representative BEACH parameters (PDFF_{median} and PDFF₉₀) are compared between patients with and without fat metaplasia. The displayed *p* values were obtained by logistic regression. ROC curves for all relevant parameters are shown in the bottom right



Inter- and intra-observer agreement

Inter- and intra-observer agreement statistics for qMRI parameters and visual scores are shown in Table 3. Inter-observer and intra-observer agreement were excellent for all assessed qMRI parameters. Inter-observer

agreement was excellent for visual inflammation scores, although the 95% limits of agreement (0.6 ± 6.4) were relatively wide compared with the ASAS definition of active inflammation (score of ≥ 2 diagnostic for active inflammation). Inter-observer agreement was poorer for fat metaplasia scores with an ICC value of 0.544.

Table 2 Comparison of structural parameters between patients with and without fat metaplasia. PDFF₇₅, PDFF₉₀, etc. refer to the 75th and 90th percentiles of ADC measurements in the defined ROI. Estimates are displayed as mean (95% CI). Odds ratio (OR) and *p* values (*) were derived from logistic regression. The highest ROC AUC value for the

evaluation of fat metaplasia is shown in *italics*. Sensitivity and specificity values for the optimal cutoff values (far right) are provided in the main “Results” section. The right-hand *p* values (**) relate to the comparison of ROC AUC with the median value

	Fat metaplasia		OR (95% CI)	* <i>p</i> =	ROC AUC (95% CI)	Optimal cutoff	** <i>p</i> =
	Present (<i>n</i> = 30)	Absent (<i>n</i> = 23)					
PDFF _{mean}	48.2 (43.9–52.5) %	37.4 (33.7–41.1) %	1.110 (1.035–1.119)	0.003	0.759 (0.629–0.889)	–	0.061
PDFF _{median}	48.2 (44.1–52.3) %	38.0 (34.0–41.9) %	1.100 (1.030–1.174)	0.004	0.735 (0.597–0.872)	–	–
PDFF ₇₅	57.1 (52.0–62.2) %	44.5 (48.2–40.8) %	1.105 (1.034–1.182)	0.003	0.759 (0.630–0.888)	–	0.254
PDFF ₉₀	64.3 (59.2–69.4) %	50.4 (46.5–54.3) %	1.111 (1.037–1.189)	0.002	<i>0.780 (0.656–0.903)</i>	55.7%	0.263
ADC _{mean}	606 (541–671)	652 (571–732)	0.999 (0.996–1.002)	0.380	0.567 (0.407–0.726)	–	0.669
ADC _{median}	602 (535–669)	648 (572–724)	0.999 (0.996–1.002)	0.380	0.572 (0.413–0.731)	–	–

Discussion

We describe a quantitative, partially automated method for measurement of bone marrow oedema and fat metaplasia based on histogram analysis of quantitative MR images. We show that histogram-based qMRI parameters enable separation of patients according to the presence of oedema and fat metaplasia, both of which are of importance for the diagnosis and management of SpA. The proposed tool offers a simple and potentially repeatable means to quantify inflammation and fat and could be incorporated into picture archiving and communications system (PACS) systems relatively easily. Such a tool could be of value for monitoring inflammation over time and for guiding clinical decisions around initiation and changes of biologic and other therapies. Importantly, ADC-based and PDFF-based parameters provide discrete information regarding oedema and fat metaplasia and could therefore inform on the relative burden of active inflammation versus structural damage.

We found that ADC measurements produced superior performance to PDFF measurements for separating patients with and without inflammation. This suggests that increases in diffusivity are an important part of the inflammatory process in the bone marrow, rather than changes in water content per se. However, previous studies have shown substantial differences in PDFF between normal and inflamed marrow [12], and it may be that the discrepant observations in this study are due to the variability in the composition of normal bone marrow [28]. This could be investigated further by comparing the composition of the inflamed subchondral bone with normal bone marrow.

Our results showed that PDFF₉₀ enabled separation of patients with and without fat metaplasia. Fat metaplasia can contribute to diagnosis [6, 7] and is also a prognostic factor, since patients with fat metaplasia are more likely to fuse their sacroiliac joints [29–31].

Interestingly, the 90th percentiles of ADC and PDFF yielded more accurate separation of inflamed and non-inflamed joints and joints with and without fat metaplasia compared with simple averages, although this difference did not reach statistical significance. This suggests that percentiles measuring the extremes of the distribution might be better ‘targeted’ to areas of oedema (for ADC) or fat metaplasia (for PDFF) than mean or median measurements, which may be ‘contaminated’ by non-inflamed or non-fatty sites, respectively, to a greater extent.

Importantly, the inter- and intra-observer variability for both ADC- and PDFF-based parameters was good or excellent. Inter-observer variability was excellent for visual scoring of bone marrow oedema, but substantially poorer for scoring of fat metaplasia. Given the known inconsistencies in radiologists’ interpretation in spondyloarthritis in clinical practice [14], a more consistent measurement could be a major advantage. However, formal studies are needed to assess repeatability and reproducibility across sites and MRI vendors.

We did not find a strong relationship between inflammation on MRI and symptoms in this study, likely reflecting the complex and multidimensional nature of pain in SpA [32]. There was a negative relationship between the severity of fat metaplasia and symptom scores. This suggests that fat metaplasia, a post-inflammatory phenomenon [30, 33], is more common in patients already on treatment with well-controlled symptoms.

A strength of our study is that the control subjects (i.e. those without inflammation) were patients where MRI was clinically indicated and thus likely to have either biomechanical back pain or quiescent inflammatory arthritis. Consequently, the reported statistics for separating patients with and without inflammation are likely to be realistic in a real-world clinical setting (this point is emphasised in the QUADAS-2 quality criteria [34]). By contrast, the use of healthy controls can artificially inflate sensitivity and specificity statistics and give a misleading impression of diagnostic performance. An additional strength is that the histogrammic

Table 3 Inter-observer and intra-observer variability statistics for selected (most relevant) parameters. The intra-class correlation coefficient and Bland-Altman limits of agreement are shown

	Inter-observer variability		Intra-observer variability (observer 1)		Intra-observer variability (observer 2)	
	ICC	95% LoA	ICC	95% LoA	ICC	95% LoA
ADC _{mean}	0.962	8.4 ± 103 mm ² /s	0.949	− 27.6 ± 128 mm ² /s	0.949	− 0.19 ± 112 mm ² /s
ADC _{median}	0.943	8.4 ± 111 mm ² /s	0.954	− 20.6 ± 124 mm ² /s	0.898	11 ± 156 mm ² /s
ADC ₉₀	0.918	− 10.1 ± 188 mm ² /s	0.918	− 38.1 ± 210 mm ² /s	0.961	3.9 ± 123 mm ² /s
PDFF _{mean}	0.982	− 1.03 ± 4.42%	0.994	0.15 ± 2.40%	0.986	− 1.08 ± 3.69%
PDFF _{median}	0.974	− 0.97 ± 5.08%	0.989	0.34 ± 3.20%	0.968	− 0.87 ± 5.45%
PDFF ₉₀	0.986	− 0.31 ± 4.70%	0.990	− 0.41 ± 3.60%	0.968	0.05 ± 6.52%
Visual inflammation score	0.944	0.6 ± 6.4	–	–	–	–
Visual fat metaplasia score	0.534	− 6.8 ± 18.3	–	–	–	–

parameters used are relatively simple and likely to offer superior performance to more complex metrics based on maximum likelihood estimation. Nonetheless, future work could explore the use of more complex analysis methods, such as Gaussian mixture modeling, to identify discrete subpopulations of pixels within the ROI.

A limitation of this study is that the diagnostic performance reported is not likely to be sufficient for the current use in clinical practice. This may be partially due to the variations in the composition of normal marrow in young patients, where the marrow may be partially ossified and contains varying proportions of water and fat. This factor may bias ADC and PDFF measurements and could have weakened the separation of inflamed and non-inflamed patients. In the future, the BEACH tool could be extended to isolate ossified bone, potentially improving performance. Similarly, the proportion of red and yellow marrow in ossified bone may vary between individuals. The use of variable thresholds depending on the composition of the normal ‘background’ marrow might help to improve the technique for detecting inflammation. ADC measurements can also suffer from poor reproducibility across sites, partly due to the difficulty of achieving high-quality fat suppression [19]. A final limitation is that the proposed tool is only partially automated; further methodological development is required to achieve full automation.

In conclusion, we describe a method for quantifying bone marrow oedema and fat metaplasia in patients with SpA, based on histographic analysis. ADC-based parameters can objectively differentiate patients with bone marrow oedema from those without, whilst PDFF-based parameters can differentiate patients with fat metaplasia from those without. Histographic analysis might improve performance compared with simple averages such as the mean and median and offers excellent agreement within and between observers.

Funding information This work was undertaken at UCLH/UCL, which receives funding from the UK Department of Health’s the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) funding scheme. The views expressed in this publication are those of the authors and not necessarily those of the UK Department of Health. TJPB was supported by Arthritis Research UK Grant 21,369. CF was also supported by Arthritis Research UK. MHC is supported by the NIHR University College London Hospitals Biomedical Research Centre.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Professor Margaret A. Hall-Craggs.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval Institutional Review Board approval was obtained.

Study subjects or cohorts overlap Eighteen of the 53 subjects included in this work were included in a prior publication (developing the PDFF-mapping methodology underpinning this work, but without the BEACH analysis performed here) [Bray et al, *Magn Reson Med* 2018;79(2)] and in a separate study investigating the use of quantitative susceptibility mapping (an entirely separate method from those investigated here) [Bray et al *Magn Reson Med* 2019;81(5)].

Methodology

- Prospective
- Observational
- Performed at one institution

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Sieper J, Poddubnyy D (2017) Axial spondyloarthritis. *Lancet* 390: 73–84
2. Wedderkopp N, Leboeuf-Yde C, Bo Andersen L, Froberg K, Steen Hansen H (2003) Back pain in children: no association with objectively measured level of physical activity. *Spine (Phila Pa 1976)* 28: 2019–2024
3. Dincer U, Cakar E, Kiralp MZ, Dursun H (2008) Diagnosis delay in patients with ankylosing spondylitis: possible reasons and proposals for new diagnostic criteria. *Clin Rheumatol* 27:457–462
4. Weiss PF, Xiao R, Biko DM, Johnson AM, Chauvin NA (2015) Detection of inflammatory sacroiliitis in children with magnetic resonance imaging: is gadolinium contrast enhancement necessary? *Arthritis Rheumatol* 67:2250–2256
5. Fisher C, Ioannou Y, Hall-Craggs MA, Sen D (2012) Enthesitis related arthritis; a new era of understanding. *Ann Paediatr Rheumatol* 1:8
6. Bray TJP, Jones A, Bennett AN et al (2019) Recommendations for acquisition and interpretation of MRI of the spine and sacroiliac joints in the diagnosis of axial spondyloarthritis in the UK. *Rheumatology (Oxford)* 58:1831–1838
7. Jones A, Bray TJP, Mandl P, Hall-Craggs MA, Marzo-Ortega H, Machado PM (2019) Performance of magnetic resonance imaging in the diagnosis of axial spondyloarthritis: a systematic literature review. *Rheumatology (Oxford)* 58:1955–1965
8. Baraliakos X, Braun J (2013) Opinion: perspectives on imaging in axial spondyloarthritis. *Nat Rev Rheumatol* 9:498–502
9. Rudwaleit M, Jurik AG, Hermann KGA et al (2009) 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* 68:1520–1527
10. Rudwaleit M, van der Heijde D, Landewé R et al (2009) The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 68:777–783
 11. Sieper J, Rudwaleit M, Baraliakos X et al (2009) The Assessment of SpondyloArthritis international society (ASAS) handbook: a guide to assess spondyloarthritis. *Ann Rheum Dis* 68:ii1–i44
 12. Bray TJP, Bainbridge A, Punwani S, Ioannou Y, Hall-Craggs MA (2018) Simultaneous quantification of bone edema/adiposity and structure in inflamed bone using chemical shift-encoded MRI in spondyloarthritis. *Magn Reson Med* 79:1031–1042
 13. Bray TJP, Vendhan K, Ambrose N et al (2017) Diffusion-weighted imaging is a sensitive biomarker of response to biologic therapy in enthesitis-related arthritis. *Rheumatology (Oxford)* 56:399–407
 14. Bennett AN, Marzo-Ortega H, Kaur-Papadakis D, Rehman A (2017) The use of magnetic resonance imaging in axial spondyloarthritis: time to bridge the gap between radiologists and rheumatologists. *J Rheumatol* 44:780–785
 15. Vendhan K, Bray TJP, Atkinson D et al (2016) A diffusion-based quantification technique for assessment of sacroiliitis in adolescents with enthesitis-related arthritis. *Br J Radiol* 89:20150775
 16. Gaspersic N, Sersa I, Jevtic V, Tomsic M, Praprotnik S (2008) Monitoring ankylosing spondylitis therapy by dynamic contrast-enhanced and diffusion-weighted magnetic resonance imaging. *Skeletal Radiol* 37:123–131
 17. Gezmis E, Donmez FY, Agildere M (2013) Diagnosis of early sacroiliitis in seronegative spondyloarthropathies by DWI and correlation of clinical and laboratory findings with ADC values. *Eur J Radiol* 82:2316–2321
 18. Bozgeyik Z, Ozgocmen S, Kocakoc E (2008) Role of diffusion-weighted MRI in the detection of early active sacroiliitis. *AJR Am J Roentgenol* 191:980–986
 19. Dietrich O, Geith T, Reiser MF, Baur-Melnyk A (2017) Diffusion imaging of the vertebral bone marrow. *NMR Biomed* 30. <https://doi.org/10.1002/nbm.3333>
 20. Rudwaleit M, Landewé R, Van Der Heijde D et al (2009) The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. *Ann Rheum Dis* 68:770–776
 21. Petty RE, Southwood TR, Manners P et al (2004) International league of associations for rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. *J Rheumatol* 31:390–392
 22. Maksymowych WP, Inman RD, Salonen D et al (2005) Spondyloarthritis research consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. *Arthritis Rheum* 53:703–709
 23. Aydin SZ, Maksymowych WP, Bennett AN, McGonagle D, Emery P, Marzo-Ortega H (2012) Validation of the ASAS criteria and definition of a positive MRI of the sacroiliac joint in an inception cohort of axial spondyloarthritis followed up for 8 years. *Ann Rheum Dis* 71:56–60
 24. Lambert RGW, Bakker PAC, van der Heijde D et al (2016) Defining active sacroiliitis on MRI for classification of axial spondyloarthritis: update by the ASAS MRI working group. *Ann Rheum Dis* 75:1958–1963
 25. Maksymowych WP, Wichuk S, Chiowchanwisawakit P, Lambert RG, Pedersen S (2015) Development and preliminary validation of the spondyloarthritis research consortium of Canada magnetic resonance imaging sacroiliac joint structural score. *J Rheumatol* 42:79–86
 26. Bakker PAC, Van Den Berg R, Lenczner G et al (2017) Can we use structural lesions seen on MRI of the sacroiliac joints reliably for the classification of patients according to the ASAS axial spondyloarthritis criteria? Data from the DESIR cohort. *Ann Rheum Dis* 76:392–398
 27. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44:837
 28. Bray TJP, Vendhan K, Roberts J et al (2016) Association of the apparent diffusion coefficient with maturity in adolescent sacroiliac joints. *J Magn Reson Imaging* 44:556–564
 29. Mandl P, Navarro-Compan V, Terslev L et al (2015) EULAR recommendations for the use of imaging in the diagnosis and management of spondyloarthritis in clinical practice. *Ann Rheum Dis* 74:1327–1339
 30. Maksymowych WP, Wichuk S, Chiowchanwisawakit P, Lambert RG, Pedersen J (2014) Fat metaplasia and backfill are key intermediaries in the development of sacroiliac joint ankylosis in patients with ankylosing spondylitis. *Arthritis Rheumatol* 66:2958–2967
 31. Kang KY, Kim IJ, Yoon MA, Hong YS, Park S-H, Ju JH (2015) Fat metaplasia on sacroiliac joint magnetic resonance imaging at baseline is associated with spinal radiographic progression in patients with axial spondyloarthritis. *PLoS One* 10:e0135206
 32. Stinson JN, Luca NJC, Jibb LA (2012) Assessment and management of pain in juvenile idiopathic arthritis. *Pain Res Manag* 17:391–396
 33. Chiowchanwisawakit P, Lambert RGW, Conner-Spady B, Maksymowych WP (2011) Focal fat lesions at vertebral corners on magnetic resonance imaging predict the development of new syndesmophytes in ankylosing spondylitis. *Arthritis Rheum* 63:2215–2225
 34. Whiting PF, Rutjes AWS, Westwood ME et al (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 155:529–536

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.