

## Original article

# Diffusion-weighted imaging is a sensitive biomarker of response to biologic therapy in enthesitis-related arthritis

Timothy J. P. Bray<sup>1,2</sup>, Kanimozhi Vendhan<sup>1</sup>, Nicola Ambrose<sup>2</sup>, David Atkinson<sup>1</sup>, Shonit Punwani<sup>1</sup>, Corinne Fisher<sup>2</sup>, Debajit Sen<sup>2</sup>, Yiannis Ioannou<sup>2</sup> and Margaret A. Hall-Craggs<sup>1</sup>

## Abstract

**Objective.** The aim was to evaluate diffusion-weighted imaging (DWI) as a tool for measuring treatment response in adolescents with enthesitis-related arthropathy (ERA).

**Methods.** Twenty-two adolescents with ERA underwent routine MRI and DWI before and after TNF inhibitor therapy. Each patient's images were visually scored by two radiologists using the Spondyloarthritis Research Consortium of Canada system, and sacroiliac joint apparent diffusion coefficient (ADC) and normalized ADC (nADC) were measured for each patient. Therapeutic clinical response was defined as an improvement of  $\geq 30\%$  physician global assessment and radiological response defined as  $\geq 2.5$ -point reduction in Spondyloarthritis Research Consortium of Canada score. We compared ADC and nADC changes in responders and non-responders using the Mann–Whitney–Wilcoxon test.

**Results.** For both radiological and clinical definitions of response, reductions in ADC and nADC after treatment were greater in responders than in non-responders (for radiological response: ADC:  $P < 0.01$ ; nADC:  $P = 0.055$ ; for clinical response: ADC:  $P = 0.33$ ; nADC:  $P = 0.089$ ). ADC and nADC could predict radiological response with a high level of sensitivity and specificity and were moderately sensitive and specific predictors of clinical response (the area under the receiver operating characteristic curves were as follows: ADC: 0.97, nADC: 0.82 for radiological response; and ADC: 0.67, nADC: 0.78 for clinical response).

**Conclusion.** DWI measurements reflect the response to TNF inhibitor treatment in ERA patients with sacroiliitis as defined using radiological criteria and may also reflect clinical response. DWI is more objective than visual scoring and has the potential to be automated. ADC/nADC could be used as biomarkers of sacroiliitis in the clinic and in clinical trials.

**Key words:** magnetic resonance imaging, diffusion-weighted imaging, spondyloarthritis, juvenile idiopathic arthritis, biomarkers, apparent diffusion coefficient, adolescents, inflammation, arthritis

## Rheumatology key messages

- Diffusion-weighted imaging can be used to measure therapeutic response in enthesitis-related arthritis.
- Diffusion-weighted imaging is more objective than visual scoring as a measure of sacroiliitis in enthesitis-related arthritis.
- Quantitative imaging could be used to guide treatment decisions in the clinic in enthesitis-related arthritis.

<sup>1</sup>University College London Centre for Medical Imaging (Academic Radiology), NW1 2PG and <sup>2</sup>Arthritis Research UK Centre for Adolescent Rheumatology, University College London, London  
Submitted 17 March 2016; revised version accepted 21 October 2016  
Correspondence to: Timothy J. P. Bray, Centre for Medical Imaging, 3rd Floor East, 250 Euston Road, London. NW12PG.  
E-mail: t.bray@ucl.ac.uk

## Introduction

Enthesitis-related arthritis (ERA) is a juvenile-onset spondyloarthritis associated with severe pain, stiffness and disability, which accounts for  $\sim 20\%$  of all adolescents and young adults with childhood-onset arthritis [1, 2].

Inflammation of the sacroiliac joints (sacroiliitis), is a common feature of ERA. Unlike other subtypes of JIA, ERA almost always progresses into adult life [3]. Early treatment in spondyloarthritis has been shown to have a disease-modifying effect with consequently good outcomes [4], but if treatment is inadequate then outcomes are poor [3]. Importantly, inflammatory and biomechanical back pain may coexist in the same patient, and it may be challenging to differentiate between the two in young patients with ERA/spondyloarthritis. A reliable, quick and cheap tool that could enhance confidence of adequate treatment and inflammatory control in early disease would therefore be clinically useful, especially in childhood-onset or early adult-onset spondyloarthritis.

Clinical evaluation is helpful in assessing disease activity, but has some limitations given that standard inflammatory blood markers may be normal in active disease [5, 6]. Clinical assessment of sacroiliitis specifically is also somewhat unreliable [7, 8]. Furthermore, disease activity measures in spondyloarthritis may vary substantially over repeated measurements [9] and have not been prospectively validated in childhood-onset spondyloarthritis.

In clinical practice, radiologists typically assess bone marrow oedema on short tau inversion recovery (STIR) images. However, these scans allow for only qualitative assessment; that is, they rely on subjective assessment of the images by the interpreting radiologist. Likewise, the Spondyloarthritis Research Consortium of Canada (SPARCC) [10] system requires a specialist radiologist to assign an overall inflammation score. The SPARCC score contains a number of subjective elements, including assessment of the depth and brightness of inflammation and the number of inflamed joint quadrants. Furthermore, observers can make only binary choices for each joint quadrant, which is unsatisfactory where only early/subtle inflammatory changes are present. STIR acquisitions are also time consuming. These factors make the SPARCC system less attractive for clinical use.

Recent work has examined the use of diffusion-weighted imaging (DWI) as a fast, quantitative method for quantifying sacroiliac joint inflammation [11–14]. Increased sacroiliac joint apparent diffusion coefficient (ADC) values have been reported both in adult sacroiliitis [11–13] and in sacroiliitis in adolescents with ERA [14], and are thought to be attributable to increased extracellular fluid (exudate) and cellular infiltration [15] in the juxta-articular bone marrow. ADC measurements are intrinsically more objective than visual scoring because they are derived from pixel values in the image itself. Importantly, recent studies have examined the use of a reference region of interest (ROI) placed on normal sacral bone to normalize ADC values, with the aim of minimizing between-scan variations in measured ADC [14]. There is a good correlation between normalized ADC (nADC) measurements and SPARCC scores of inflammation in ERA [14].

However, there have been no previous studies assessing whether DWI can be used to measure the response to therapy in ERA. Establishing whether biomarker estimates

reflect biological change is an essential part of quantitative imaging biomarker validation [16, 17].

In this study, we evaluate both ADC and nADC as measures of response to TNF inhibitor (TNFi) therapy. As a primary objective, we evaluate whether the change in ADC/nADC after treatment is greater in responders to TNFi treatment than in non-responders. Secondly, we determine the extent to which change in ADC/nADC can be used to classify patients as responders/non-responders, and assess the correlation between change in nADC/ADC and change in SPARCC STIR score.

## Methods

This retrospective study was covered by Institutional Review Board approval from the National Research Ethics Service Committee London, Bentham, UK (REC ref: 11/LO/0330). Informed consent was waived because of the retrospective nature of the study.

### Subjects

A local clinical adolescent rheumatology database was used to identify all those ERA patients with sacroiliitis who had been started on biologic therapy between January 2009 and June 2015. We then performed a picture archiving and communication systems search to identify those individuals who had undergone MRI scans of the sacroiliac joints both before and after starting biologic therapy, using the imaging protocol specified below (see MRI technique). Patients who started on biologic treatments during this period and who had MRI scans both before and after starting therapy were selected for the study. All subjects fulfilled the ILAR criteria for ERA [18] and were treated with either etanercept or adalimumab. The decision to scan the patients and to treat with biologic therapy was made as part of standard clinical care in all cases. At our institution, ERA patients are typically scanned at presentation, 3 months after starting treatment (usually MTX) and again after a further 3–6 months if patients are started on biologic therapy to confirm improvement of inflammatory changes [19]. A subset of patients are also scanned at regular intervals (typically yearly) following this for disease monitoring [19]. Two patients were excluded from the study because the DWI acquisition was not performed to protocol.

### MRI technique

MRI of the sacroiliac joints was performed using a 1.5 T system (Avanto; Siemens, Berlin, Germany). In all patients, we acquired T1-weighted, STIR, post-contrast T1-weighted and diffusion DWI of the sacroiliac joints. DWI were acquired with b-values of 0, 50, 100, 300 and 600 s/mm<sup>2</sup> with fat saturation, and apparent diffusion maps were generated on the scanner using a monoexponential fit. Further scan parameters are detailed in supplementary Table S1, available at *Rheumatology* Online.

### Image analysis

nADC measurements were performed using a previously described technique [14], as follows. The central four axial slices on the ADC maps were analysed using in-house MATLAB (The MathWorks, Natick, MA, USA) code. Three linear ROIs measuring 14 mm were drawn across the synovial portion of each sacroiliac joint, with each ROI centred on the joint space. Where the anteroposterior dimensions of the joint were too small to place three ROIs, only two ROIs were placed. A further reference ROI was placed on normal sacral bone to provide internal standardization. The nADC value of each patient was defined as the ratio between the mean ADC of all joint line profiles and the mean reference ADC.

For each scan, both the uncorrected ADC and the nADC were recorded. The measurements were performed independently by two radiologists (K.V. and T.J.P.B., with 7 and 4 years of musculoskeletal MRI experience, respectively); the mean of the two radiologists' scores was used for the analysis.

The change in ADC after therapy ( $\Delta\text{ADC}$ ) was defined as follows:

$$\Delta\text{ADC} = \text{ADC}_{\text{pre}} - \text{ADC}_{\text{post}}$$

and the change in nADC after therapy was as follows:

$$\Delta\text{nADC} = \text{nADC}_{\text{pre}} - \text{nADC}_{\text{post}}$$

Note that positive  $\Delta\text{ADC}$  and  $\Delta\text{nADC}$  values represent a reduction in the post-treatment nADC (i.e. improving inflammation).

The SPARCC STIR scoring technique [10] was modified for use on axial rather than coronal images, to facilitate comparison between STIR images and ADC maps as previously described [14]. On each of the central six axial slices, the sacroiliac joint was divided into four quadrants. Increased STIR signal was given a score of one per quadrant, and normal signal was scored zero. For each slice, an additional score of one per joint was given for deep or intense lesions [10]. Each patient received a maximal score of 12 per slice and a maximal total of 72. Scoring was performed independently by two radiologists (M.H.C. and T.J.P.B.) with more than 20 and 4 years of musculoskeletal MRI experience, respectively, who were blinded to clinical data and the diffusion scores. The mean score from the two sets of measurements was used for the analysis. The change in STIR score after therapy was defined as follows:

$$\Delta\text{STIR} = \text{STIR}_{\text{pre}} - \text{STIR}_{\text{post}}$$

### Response classification

Radiological response classification was based on changes in SPARCC STIR score after treatment. Specifically, based on previous studies defining a minimally important change for SPARCC scores of sacroiliac joint inflammation [20], patients were classified as radiological responders if the mean STIR score from the two observers improved by 2.5 or more, and radiological non-

responders otherwise. This threshold was derived using a receiver operating characteristic (ROC) analysis to determine the STIR score change that would predict minimally important changes in inflammation (as determined by a radiologist) with the highest degree of sensitivity and specificity [20].

Each patient was classified as either a clinical responder or a clinical non-responder to TNFi therapy using a retrospective physician global assessment (PGA). Specifically, a specialist consultant adolescent rheumatologist (N.A.), who was blinded to SPARCC, ADC and nADC scores, reviewed the electronic medical record to determine clinical symptoms at the time each scan was acquired (i.e. both the pre-TNFi and post-TNFi scans) to define a composite global assessment of response to treatment. Patients who required emergency steroid treatment (defined as a course of systemic steroids to treat a flare, i.e. oral prednisolone, i.m. or i.v. methylprednisolone) or a switch to an alternative TNFi at the time of the second scan were classified as non-responders. Patients who had improved only marginally (defined as an improvement in PGA of < 30%, mirroring the PGA component of the ACR Pedi 30 criteria [21, 22]) were also classified as non-responders.

In cases where clinical and radiological response classifications were discordant, we reviewed the individual STIR, nADC and ADC scores to determine the reasons for disagreement. Biochemical markers of inflammation (specifically CRP and ESR) were not used as response classifiers because there are no accepted criteria and because these markers are insensitive as measures of inflammation [23].

### Statistical analysis

The Mann-Whitney-Wilcoxon test was used for between-group comparisons. The correlation between change in  $\Delta\text{ADC}/\Delta\text{nADC}$  and  $\Delta\text{STIR}$  was evaluated using Spearman's  $\rho$ . ROC analyses were performed using perfcurve function in MATLAB to assess sensitivity and specificity for determining response using both  $\Delta\text{ADC}$  and  $\Delta\text{nADC}$  (using both clinical and radiological response classifications). Repeatability was assessed separately for pretreatment and post-treatment nADC and ADC measurements using intraclass correlation coefficient (ICC) and Bland-Altman 95% limits of agreement.

## Results

### Demographics

Twenty-two patients were recruited, with a mean age at biologic initiation of 17 years 4 months. Eighteen subjects were male (mean age 17 years 3 months) and four female (mean age 17 years 6 months). The mean interval from the pretreatment scan to the initiation of biologic therapy was 4 months (range 1–8 months). The mean interval from the start of biologic treatment to post-treatment scan was 1 year 1 month (range 5 months to 2 years 6 months).

**TABLE 1** Apparent diffusion coefficient values in radiological responders and non-responders

	Radiological responders	Radiological non-responders	P-values
ADC pretreatment	912 (185)	654 (317)	0.03
ADC post-treatment	696 (226)	962 (296)	0.16
$\Delta$ ADC	+240 (407)	-308 (253)	<0.01
nADC pretreatment	1.67 (0.53)	1.29 (0.80)	0.31
nADC post-treatment	1.42 (0.43)	1.75 (0.51)	0.22
$\Delta$ nADC	+0.21 (0.27)	-0.25 (0.37)	0.055

Median (interquartile range) nADC, uncorrected ADC,  $\Delta$ nADC and  $\Delta$ ADC values are shown for responding and non-responding groups. The two groups were compared using the Mann-Whitney-Wilcoxon test; P-values are shown in the right-hand column. Positive  $\Delta$ nADC and  $\Delta$ ADC values represent an improvement in inflammation; negative values represent a worsening of inflammation. ADC: apparent diffusion coefficient; nADC: normalized

**Disease response: radiological classification**

Of the 22 patients in the cohort, 18/22 patients (82%) demonstrated a STIR score improvement of  $\geq 2.5$  and were classified as radiological responders; 4/22 (18%) were classified as radiological non-responders. Uncorrected ADC and nADC values before and after treatment, classified by radiological response, are shown in Table 1 and Fig. 1.

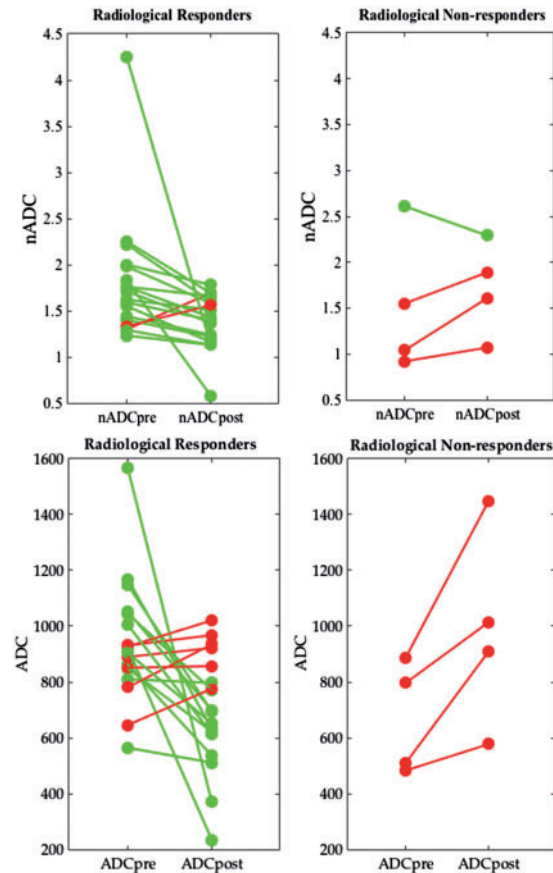
Baseline ADC values were significantly higher pretreatment in radiological responders compared with non-responders ( $P=0.03$ ). After treatment, there was a decrease in ADC values in responders and an increase in non-responders, such that post-treatment ADC values were higher in non-responders than in responders ( $P=0.16$ ). Furthermore, the change in ADC values ( $\Delta$ ADC) was significantly greater in responders than in non-responders ( $P < 0.01$ ).

Baseline nADC values were also higher in responders than in non-responders, although this difference was non-significant ( $P=0.31$ ). Following treatment, nADC values were marginally lower in responders than in non-responders, again non-significant ( $P=0.22$ ). There was a reduction in nADC values after treatment in both responders (median  $\Delta$ nADC=0.27) and non-responders (median  $\Delta$ nADC=0.10). The change in nADC values (i.e.  $\Delta$ nADC) was greater in responders than in non-responders; this difference was borderline significant ( $P=0.055$ ).

**Disease response: clinical classification**

Of the 22 patients in the cohort, 18/22 patients (82%) were classified as clinical responders and 4/22 (18%) as clinical non-responders. Of the clinical non-responders, two patients had also been classified as radiological non-

**Fig. 1** Response plots for normalized apparent diffusion coefficient and apparent diffusion coefficient by radiological response



Pre- and post-treatment nADC and ADC values are shown for all 22 patients, classified according to the radiological response. Patients whose nADC/ADC reduced after treatment are shown in green, while patients whose nADC/ADC increased are shown in red. ADC values have units of square millimetres per second  $\times 10^{-6}$ . ADC: apparent diffusion coefficient; nADC: normalized apparent diffusion coefficient.

responders. ADC and nADC values before and after treatment, classified by clinical response, are shown in Table 2 and Fig. 2.

There was no significant difference in baseline ADC values between responders and non-responders ( $P=0.90$ ). Post-treatment ADC values were higher in non-responders, although this difference was again non-significant ( $P=0.097$ ). There was no significant difference in the change in ADC values (i.e.  $\Delta$ ADC) between responders and non-responders ( $P=0.33$ ).

There was no significant difference in baseline nADC values between responders and non-responders ( $P=0.77$ ). Following treatment, nADC values were significantly lower in responders than in non-responders ( $P < 0.01$ ). Accordingly, there was a reduction in nADC



**TABLE 2** Apparent diffusion coefficient values in clinical responders and non-responders

	Clinical responders	Clinical non-responders	P-values
ADC pretreatment	897 (189)	849 (146)	0.90
ADC post-treatment	675 (310)	905 (351)	0.097
$\Delta$ ADC	+139 (426)	-102 (426)	0.33
nADC pretreatment	1.61 (0.60)	1.65 (0.49)	0.77
nADC post-treatment	1.39 (0.44)	1.79 (0.31)	<0.01
$\Delta$ nADC	+0.21 (0.27)	-0.12 (0.51)	0.089

Median (interquartile range) nADC, uncorrected ADC,  $\Delta$ nADC and  $\Delta$ ADC values are shown for responding and non-responding groups. The two groups were compared using the Mann-Whitney-Wilcoxon test; P-values are shown in the right-hand column. Positive  $\Delta$ nADC and  $\Delta$ ADC values represent an improvement in inflammation; negative values represent a worsening of inflammation. ADC: apparent diffusion coefficient; nADC: normalized apparent diffusion coefficient;  $\Delta$ ADC: change in ADC after therapy;  $\Delta$ nADC: change in nADC after therapy.

values in clinical responders (median  $\Delta$ nADC=0.21) and an increase in nADC values in non-responders (median  $\Delta$ nADC=-0.12); the difference between these groups was borderline significant (P=0.089).

### ROC analysis

#### Using radiological criteria for response classification

Any decrease in nADC after treatment was 89% sensitive and 75% specific for distinguishing radiological responders (i.e. those with a reduction in SPARCC score  $\geq 2.5$ ) from radiological non-responders. The area under the ROC curve (ROC AUC) was 0.82, with a sensitivity of 89% and specificity of 75% at the optimal operating point (Fig. 3).

Any decrease in ADC after treatment was 67% sensitive and 100% specific for distinguishing radiological responders from radiological non-responders. The ROC AUC was 0.97, with a sensitivity of 100% and specificity of 75% at the optimal operating point.

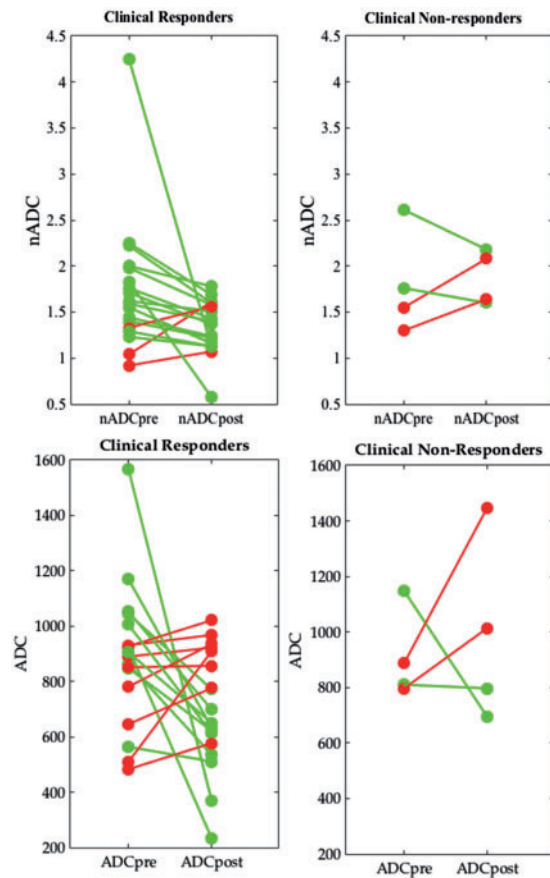
#### Using clinical criteria for response classification

Any decrease in nADC after treatment was 83% sensitive and 50% specific for distinguishing clinical responders from clinical non-responders. The ROC AUC was 0.78, with a sensitivity of 95% and a specificity of 50% at the optimal operating point.

Any decrease in ADC after treatment was 50% sensitive and 50% specific for distinguishing clinical responders from clinical non-responders. The ROC AUC was 0.67, with a sensitivity of 95% and a specificity of 50% at the optimal operating point (Fig. 3).

### nADC as a continuous response measure

Figure 4 shows the relationship between change in ADC/ $\Delta$ nADC ( $\Delta$ ADC and  $\Delta$ nADC, respectively) and change in

**FIG. 2** Response plots for normalized apparent diffusion coefficient and apparent diffusion coefficient by clinical response

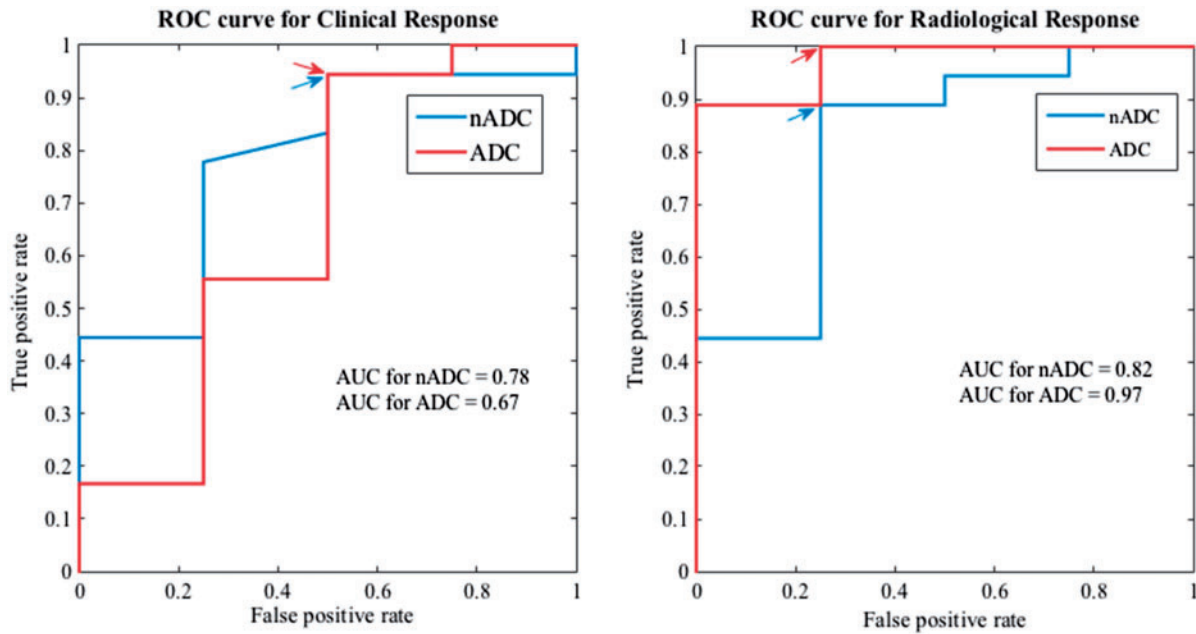
Pre- and post-treatment nADC and ADC values are shown for all 22 patients, classified according to the clinical response. Patients whose nADC/ADC reduced after treatment are shown in green, while patients whose nADC/ADC increased are shown in red. ADC values have units of square millimetres per second  $\times 10^{-6}$ . ADC: apparent diffusion coefficient; nADC: normalized apparent diffusion coefficient.

SPARCC STIR score ( $\Delta$ STIR) after TNFi treatment. There was a significant, positive correlation between  $\Delta$ ADC and  $\Delta$ STIR (R=0.60, P=0.031) and between  $\Delta$ nADC and  $\Delta$ STIR (R=0.55, P<0.01).

### Repeatability

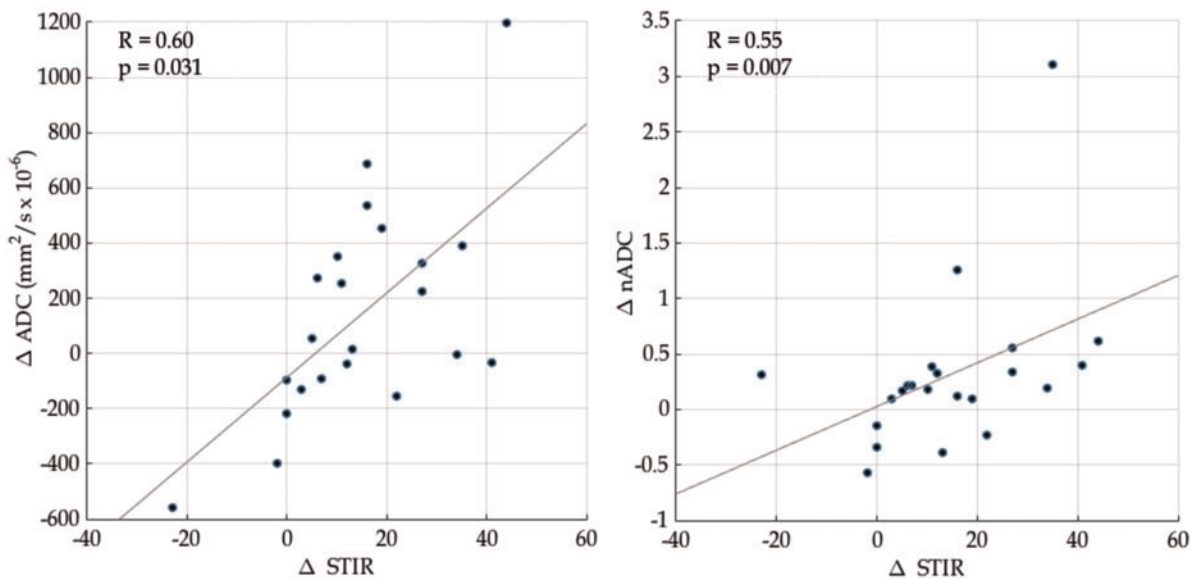
For ADC, the pretreatment ICC was 0.98 (Bland-Altman 95% limits of agreement:  $\pm 110 \times 10^{-6} \text{ mm}^2/\text{s}$ ) and the post-treatment ICC was 0.96 (Bland-Altman 95% limits of agreement:  $\pm 145 \times 10^{-6} \text{ mm}^2/\text{s}$ ). For nADC, the pretreatment ICC was 0.93 (Bland-Altman 95% limits of agreement:  $\pm 0.52$ ) and the post-treatment ICC was 0.81 (Bland-Altman 95% limits of agreement:  $\pm 0.43$ ).

**FIG. 3** Receiver operating characteristic analysis



Changes in ADC and nADC (i.e.  $\Delta$ ADC and  $\Delta$ nADC) were used to discriminate responders from non-responders. Separate curves are shown for response classification using clinical and radiological criteria. The optimal operating points are arrowed in blue for nADC and in red for ADC. ADC: apparent diffusion coefficient; nADC: normalized apparent diffusion coefficient; ROC: receiver operating characteristic;  $\Delta$ ADC: change in ADC after therapy;  $\Delta$ nADC: change in nADC after therapy.

**FIG. 4** Continuous response measurements



Positive values for  $\Delta$ ADC,  $\Delta$ nADC and  $\Delta$ STIR represent improving inflammation, whereas negative values represent worsening inflammation. ADC: apparent diffusion coefficient; nADC: normalized apparent diffusion coefficient; STIR: short tau inversion recovery;  $\Delta$ ADC: change in ADC after therapy;  $\Delta$ nADC: change in nADC after therapy;  $\Delta$ STIR: change in STIR score after therapy.

## Discussion

Several previous studies have investigated DWI as a tool for monitoring sacroiliac joint inflammation, both in adults with AS [11–13] and in adolescents with ERA [14]. However, to our knowledge, there are no previous studies evaluating the change in ADC/nADC after TNFi therapy in ERA. The results of this study suggest that changes in ADC/nADC after TNFi therapy reflect the response to treatment as defined using radiological criteria and may also reflect the response to treatment as defined clinically. Accordingly, changes in ADC/nADC could predict response with a high degree of sensitivity and specificity, and were positively correlated with changes in SPARCC STIR score.

DWI is an attractive tool for measuring response because it is more objective than STIR scoring or clinical assessment as it relies on pixel values in the image itself. Unlike STIR images, ADC maps could potentially be analysed automatically without the need for interpretation by a radiologist, making quantitative measurements of inflammation severity more readily available in the clinic. DWI is also faster than STIR imaging (typically 3 min compared with 6 min) and could help to minimize scan time for patients with stiff, painful joints. Although serial MRI scans of the sacroiliac joints are not used routinely in all rheumatological centres, images that can be acquired and analysed quickly and objectively may lower the threshold for introduction into clinical practice, thereby facilitating patient-specific therapeutic decision making. These methods could also be used to evaluate adult spondyloarthritis and to image other joints.

An interesting result of our study is that pretreatment ADC scores were significantly higher in radiological responders compared with non-responders ( $P = 0.03$ ). This raises the possibility that baseline ADC measurement could be used to determine the likelihood of response to treatment in individual patients. It may be that TNFi therapy is intrinsically more effective in patients with severe inflammation as opposed to those with lower-grade, more indolent disease. Further work in a larger cohort is needed to verify this finding.

In the present study, ADC and nADC were more accurate predictors of radiological response than of clinical response. This may be because clinical assessment is only an indirect measure of inflammation; by contrast, STIR scoring and ADC/nADC directly measure inflammation (against which TNFi treatment is directed) and are not influenced by psychological, social or biomechanical factors. Accordingly, previous studies have found clinical assessment to be an insensitive tool for diagnosing sacroiliitis [8] and that JIA patients in clinical remission frequently have evidence of ongoing inflammation on MRI scans [24]. Occult inflammation could be prognostically important because of the potential for structural damage and fusion [25, 26], which contribute to disability [27].

Here, we performed separate measurements and analyses for uncorrected ADC and for nADC values, both of which have previously been used to measure sacroiliac

joint inflammation [11–14]. Our results suggest that these measurements may have different characteristics. For example, ADC measurements demonstrated superior inter-observer repeatability compared with nADC measurements. Repeatability is clearly a desirable biomarker characteristic [16], but this study has not evaluated or compared the reproducibility of ADC and nADC. Reproducibility might be expected to be greater for nADC because normalization is designed to account for variations in image intensity between different scans and between imaging platforms [28–30]. This issue could be addressed by performing repeat scans in the same individual in different sessions and across different imaging platforms.

Establishing that biomarkers reflect biological change is a key step in imaging biomarker validation; in the present study, visual STIR scoring was used as a reference standard for biologic change [16]. However, visual STIR scoring cannot be regarded as a true gold standard for validation of ADC/nADC because ADC measurements reflect a variety of biological processes that are not assessed using STIR scoring. ADC measurements are influenced by cell membrane permeability, macromolecular packing and viscosity [31]. Additionally, fat metaplasia in areas of resolved inflammation after biologic therapy produces areas of low signal on ADC maps that are not measured using STIR scoring. ADC histogram analysis [32–34] could potentially be used to quantify active inflammation and fat metaplasia separately.

Some limitations in this work have arisen owing to the retrospective nature of the study. Firstly, this was a retrospective study, and the sample size was relatively small. Secondly, we had no control over the time interval between patients starting TNFi therapy and the second scan. ERA is a chronic disease, and sacroiliac joint changes are expected to occur over a long time scale, but it would be desirable to scan the patients at a fixed interval after starting treatment (preferably 6 months post-TNFi). Thirdly, clinical response classification was performed retrospectively, which limits the accuracy of response measurement. Nonetheless, clinical response measurements are intrinsically susceptible to a multitude of physical, psychological and social factors, and therefore, a degree of discrepancy with radiological measures of inflammation is expected. Ideally, we would collect a variety of clinical scores (including PGA, BAS Index and quality-of-life measures) to allow for a more complete assessment of clinical response [35, 36]. Although our results are promising, we aim to perform definitive biological validation in a larger, prospective study. Finally, ADC measurements in the sacroiliac joint vary with maturity [37], and further work will be required to develop strategies to account for maturity-related ADC changes.

In conclusion, we have demonstrated that DWI measurements reflect the response to treatment in adolescent ERA patients with sacroiliitis as defined using both clinical and radiological criteria. DWI is fast and objective and may facilitate patient-specific therapeutic decision making.

## Acknowledgements

This work was undertaken at University College London Hospitals (UCLH)/University College London, which receives funding from the 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. M.H.C., S.P., T.J.P.B. and Y.I. are supported by the NIHR UCLH BRC, and Y.I. is also supported by Arthritis Research UK Grant 20164.

**Funding:** No specific funding was received from any bodies in the public, commercial or not-for-profit sectors to carry out the work described in this manuscript.

**Disclosure statement:** The authors have declared no conflicts of interest.

## Supplementary data

Supplementary data are available at *Rheumatology* Online.

## References

- Solau-Gervais E, Robin C, Gambert C *et al.* Prevalence and distribution of juvenile idiopathic arthritis in a region of Western France. *Joint Bone Spine* 2010;77:47–9.
- Yilmaz M, Kendirli SG, Altintas DU *et al.* Juvenile idiopathic arthritis profile in Turkish children. *Pediatr Int* 2008;50:154–8.
- Flatø B, Hoffmann-Vold A-M, Reiff A *et al.* Long-term outcome and prognostic factors in enthesitis-related arthritis: a case-control study. *Arthritis Rheum* 2006;54:3573–82.
- Maksymowych WP, Dougados M, van der Heijde D *et al.* Clinical and MRI responses to etanercept in early non-radiographic axial spondyloarthritis: 48-week results from the EMBARK study. *Ann Rheum Dis* 2016;75:1328–35.
- Rudwaleit M, Haibel H, Baraliakos X *et al.* The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. *Arthritis Rheum* 2009;60:717–27.
- Kay J, Morgacheva O, Messing SP *et al.* Clinical disease activity and acute phase reactant levels are discordant among patients with active rheumatoid arthritis: acute phase reactant levels contribute separately to predicting outcome at one year. *Arthritis Res Ther* 2014;16:R40.
- Spadaro A, Iagnocco A, Baccano G *et al.* Sonographic-detected joint effusion compared with physical examination in the assessment of sacroiliac joints in spondyloarthritis. *Ann Rheum Dis* 2009;68:1559–63.
- Williamson L, Dockerty JL, Dalbeth N *et al.* Clinical assessment of sacroiliitis and HLA-B27 are poor predictors of sacroiliitis diagnosed by magnetic resonance imaging in psoriatic arthritis. *Rheumatology* 2004;43:85–8.
- Madsen OR, Rytter A, Hansen LB, Suetta C, Egsmose C. Reproducibility of the Bath Ankylosing Spondylitis indices of Disease Activity (BASDAI), functional status (BASFI) and overall well-being (BAS-G) in anti-tumour necrosis factor-treated spondyloarthritis patients. *Clin Rheumatol* 2010;29:849–54.
- Maksymowych WP, Inman RD, Salonen D *et al.* Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. *Arthritis Rheum* 2005;53:703–9.
- Bozgeyik Z, Ozgocmen S, Kocakoc E. Role of diffusion-weighted MRI in the detection of early active sacroiliitis. *AJR Am J Roentgenol* 2008;191:980–6.
- Gezmis E, Donmez FY, Agildere M. Diagnosis of early sacroiliitis in seronegative spondyloarthropathies by DWI and correlation of clinical and laboratory findings with ADC values. *Eur J Radiol* 2013;82:2316–21.
- Gaspersic N, Sersa I, Jevtic V, Tomsic M, Praprotnik S. Monitoring ankylosing spondylitis therapy by dynamic contrast-enhanced and diffusion-weighted magnetic resonance imaging. *Skeletal Radiol* 2008;37:123–31.
- Vendhan K, Bray TJP, Atkinson D *et al.* A diffusion-based quantification technique for assessment of sacroiliitis in adolescents with enthesitis-related arthritis. *Br J Radiol* 2015;89:20150775.
- Bollow M, Fischer T, Reissshauer H *et al.* Quantitative analyses of sacroiliac biopsies in spondyloarthropathies: T cells and macrophages predominate in early and active sacroiliitis—cellularity correlates with the degree of enhancement detected by magnetic resonance imaging. *Ann Rheum Dis* 2000;59:135–40.
- Sullivan DC, Obuchowski NA, Kessler LG *et al.* Metrology standards for quantitative imaging. *Radiology* 2015;277:813–25.
- Obuchowski NA, Reeves AP, Huang EP *et al.* Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. *Stat Methods Med Res* 2014;24:68–106.
- Petty RE, Southwood TR, Manners P *et al.* International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. *J Rheumatol* 2004;31:390–2.
- Fisher C, Ioannou Y, Hall-Craggs MA, Sen D. Enthesitis related arthritis; a new era of understanding. *Ann Paediatr Rheumatol* 2012;1:8–16.
- Maksymowych WP, Lambert RG, Brown LS, Pangan AL. Defining the minimally important change for the Spondyloarthritis Research Consortium of Canada spine and sacroiliac joint magnetic resonance imaging indices for ankylosing spondylitis. *J Rheumatol* 2012;39:1666–74.
- Giannini EH, Ruperto N, Ravelli A *et al.* Preliminary definition of improvement in juvenile arthritis. *Arthritis Rheum* 1997;40:1202–9.
- Lurati A, Pontikaki I, Teruzzi B *et al.* A comparison of response criteria to evaluate therapeutic response in patients with juvenile idiopathic arthritis treated with methotrexate and/or anti-tumor necrosis factor  $\alpha$  agents. *Arthritis Rheum* 2006;54:1602–7.
- Duurland CL, Wedderburn LR. Current developments in the use of biomarkers for juvenile idiopathic arthritis. *Curr Rheumatol Rep* 2014;16:406.



- 24 Van Veenendaal M, Hemke R, Bos MI *et al.* Magnetic resonance imaging in follow-up of clinical remission in juvenile idiopathic arthritis. *Arthritis Rheum* 2012;64:S442.
- 25 Maksymowych WP, Wichuk S, Chiowchanwisawakit P, Lambert RG, Pedersen SJ. Fat metaplasia and backfill are key intermediaries in the development of sacroiliac joint ankylosis in patients with ankylosing spondylitis. *Arthritis Rheumatol* 2014;66:2958–67.
- 26 Lories RJU, Luyten FP, de Vlam K. Progress in spondyloarthritis. Mechanisms of new bone formation in spondyloarthritis. *Arthritis Res Ther* 2009;11:221.
- 27 Machado P, Landewe R, Braun J *et al.* Both structural damage and inflammation of the spine contribute to impairment of spinal mobility in patients with ankylosing spondylitis. *Ann Rheum Dis* 2010;69:1465–70.
- 28 Wang HJ, Pui MH, Guo Y *et al.* Value of normalized apparent diffusion coefficient for estimating histological grade of vesical urothelial carcinoma. *Clin Radiol* 2014;69:727–31.
- 29 Do RKG, Chandanara H, Felker E *et al.* Diagnosis of liver fibrosis and cirrhosis with diffusion-weighted imaging: value of normalized apparent diffusion coefficient using the spleen as reference organ. *Am J Roentgenol* 2010;195:671–6.
- 30 Rosenkrantz AB, Khalef V, Xu W *et al.* Does normalisation improve the diagnostic performance of apparent diffusion coefficient values for prostate cancer assessment? A blinded independent-observer evaluation. *Clin Radiol* 2015;70:1032–7.
- 31 Le Bihan D. Apparent diffusion coefficient and beyond: what diffusion MR imaging can tell us about tissue structure. *Radiology* 2013;268:318–22.
- 32 Pope WB, Kim HJ, Huo J *et al.* Recurrent glioblastoma multiforme: ADC histogram analysis predicts response to bevacizumab treatment. *Radiology* 2009;252:182–9.
- 33 Nowosielski M, Recheis W, Goebel G *et al.* ADC histograms predict response to anti-angiogenic therapy in patients with recurrent high-grade glioma. *Neuroradiology* 2011;53:291–302.
- 34 Kim H, Choi SH, Kim JH *et al.* Gliomas: application of cumulative histogram analysis of normalized cerebral blood volume on 3 T MRI to tumor grading. *PLoS One* 2013;8:e63462.
- 35 Garrett S, Jenkinson T, Kennedy LG *et al.* A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. *J Rheumatol* 1994;21:2286–91.
- 36 Burgos-Vargas R, Tse SM, Horneff G *et al.* A randomized, double-blind, placebo-controlled multicenter study of adalimumab in pediatric patients with enthesitis-related arthritis. *Arthritis Care Res* 2015;67:1503–12.
- 37 Bray TJ, Vendhan K, Roberts J *et al.* Association of the apparent diffusion coefficient with maturity in adolescent sacroiliac joints. *J Magn Reson Imaging* 2016;44:556–64.