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

Original Research

[1,2]Timothy JP Bray

[1]Kanimozhi Vendhan

[2]Nicola Ambrose

[1]David Atkinson

[1]Shonit Punwani

[2]Corinne Fisher

[2]Debajit Sen MBBS

[2]Yiannis Ioannou

[1]Margaret A Hall-Craggs

[1]UCL Centre for Medical Imaging (Academic Radiology), 3rd floor East 250 Euston Road, London. NW1 2PG

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

Corresponding author:

Timothy JP Bray t.bray@ucl.ac.uk 0203 447 9324

Keywords: Magnetic resonance imaging, Diffusion-weighted Imaging, Spondyloarthritis, Juvenile Idiopathic Arthritis, Biomarkers

This work was undertaken at UCLH/UCL, 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. MHC, SP, TJPB, YI are supported by the NIHR University College London Hospitals BRC and YI is also supported by Arthritis Research UK Grant 20164.

Word count: 3,200

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

ABSTRACT

Objective: To evaluate diffusion-weighted imaging (DWI) as a tool for measuring treatment response in adolescents with enthesitis-related arthropathy (ERA).

Methods: 22 adolescents with ERA underwent routine MRI and DWI before and after tumour necrosis factor inhibitor (TNFi) therapy. Each patient's images were visually scored by two radiologists using the Spondyloarthritis Research Consortium of Canada (SPARCC) 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 ≥30% physician global assessment and radiological response defined as ≥2.5-point drop in SPARCC 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 ADC: 0.97, nADC: 0.82 for radiological response and ADC: 0.67, nADC: 0.78 for clinical response).

Conclusion: DWI measurements reflect response to TNFi 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.

Keywords

Diffusion-weighted imaging

Apparent diffusion coefficient

Adolescents

Inflammation

Arthritis

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

Running title

Diffusion-weighted imaging as a response biomarker in enthesitis-related arthritis

INTRODUCTION

Enthesitis-related arthritis (ERA) is a juvenile-onset spondyloarthritis associated with severe pain, stiffness and disability, which accounts for ~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 juvenile idiopathic arthritis (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 co-exist 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 are 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 only allow for *qualitative* assessment – i.e. they rely on subjective assessment of the images by the interpreting radiologist. Similarly, 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 only make 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 SIJ inflammation (11–14). Increased sacroiliac joint apparent diffusion coefficient (ADC) values have been reported in both adult sacroiliitis (11–13) and sacroiliitis in adolescents with ERA (14), and is thought to be due to increased extracellular fluid (exudate) and cellular infiltration (15) in the juxta-articular bone marrow. ADC measurements are intrinsically more objective than visual scoring since 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 *normalise* 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 response to therapy in ERA. Establishing whether biomarker estimates reflect *biological change* is an essential part of quantitative imaging biomarker (QIB) validation (16,17).

In this study, we evaluate both ADC and nADC as measures of response to tumour necrosis factor 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. Secondarily, 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.

MATERIALS AND METHODS

This retrospective study was covered by IRB approval from the National Research Ethics Service (NRES) Committee London – Bentham, England (REC ref: 11/LO/0330). Informed consent was waived due to 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 (PACS) search to identify those individuals who had undergone MRI scans of the SIJs between 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 International League of Associations for Rheumatology (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 methotrexate) 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 SIJs was performed using a 1.5T system (Avanto; Siemens, Germany). Scan parameters were as follows:

T1 turbo spin echo (TSE) coronal: TR/TE 610/11ms, slices 18, slice thickness 3mm, FOV 200mm; T1 TSE axial – TR/TE 610/11ms, slices 18, slice thickness 3mm, FOV 200mm, matrix size 256 x 256, pixel size 1mm.

T1 TSE axial: TR/TE 475/11ms, slices 20, slice thickness 5mm, FOV 200mm; T1 TSE axial – TR/TE 610/11ms, slices 18, slice thickness 3mm, FOV 200mm, matrix size 256 x 256, pixel size 1mm.

Short tau inversion recovery (STIR) axial: TR/TE 6070/83ms, inversion time 150ms, slices 18, slice thickness 5mm, FOV 200mm, matrix sixe 256 x 256, pixel size 1mm.

T1 Turbo Inversion Recovery Magnitude coronal: TR/TE 4340/83ms, inversion time 150ms, slices 14, slice thickness 4mm, FOV 200mm, matrix size 256 x 256, pixel size 1mm.

Post-contrast T1 TSE with fat saturation axial: TR/TE 619/11ms, slices 20, slice thickness 5mm, FOV 200mm, matrix size 256 x 256, pixel size 1mm.
Post-contrast T1 TSE with fat saturation coronal: T1 TSE fat sat coronal - TR/TE 795/11ms, slices 18, slice thickness 3mm, FOV 200mm, matrix size 256 x 256, pixel size 1mm.

Diffusion-weighted images axial: single-shot DWI with EPI readout. TR/TE 3500/87, FOV 316mm, matrix size 128 x 128, pixel size 2.5mm, slice thickness 8mm, averages 4, slices 17, EPI factor 120, b-values 0, 50, 100, 300 and 600s/mm^2 with fat saturation. ADC maps were generated on vendor software using a standard monoexponential fit.

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] code. Three linear regions-of-interest (ROIs)

measuring 14mm were drawn across the synovial portion of each SIJ, with each ROI centred on the joint space. Where the AP 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 normalised ADC (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 nADC were recorded. The measurements were performed independently by two radiologists (KV and TB, with seven and four years of musculoskeletal MRI experience respectively); the mean of the two radiologists' scores was used for the analysis.

The change in ADC after therapy (ΔADC) was defined as:

$$\Delta ADC = ADC_{pre} - ADC_{post}$$

and the change in nADC after therapy was:

$$\Delta nADC = nADC_{pre} - nADC_{post}$$

Note that positive ΔADC and ΔnADC 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 SIJ was divided into four quadrants. Increased STIR signal was given a score of 1 per quadrant and normal signal was scored 0. For each slice, an additional score of 1 per joint was given for deep or intense lesions (10).

Each patient received a maximum score of 12 per slice, and a maximum total of 72. Scoring was performed independently by two radiologists (MHC and TB) with over twenty years and four 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:

$$\Delta STIR = STIR_{pre} - STIR_{post}$$

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 which 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 *clinical non-responder* to TNFi therapy using a retrospective physician global assessment (PGA). Specifically, a specialist consultant adolescent rheumatologist (NA) 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, intramuscular or intravenous methylprednisolone) or a switch to an alternative TNFi at the time of the second scan were classified as non-responders. Patients who had only marginally improved

[defined as an improvement in PGA of less than 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 C-reactive protein and erythrocyte sedimentation rate) were not used as response classifiers since 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 ΔADC/ΔnADC and ΔSTIR was evaluated using Spearman's rho. Receiver operating characteristic analyses were performed using MATLAB's *perfcurve* function to assess sensitivity and specificity for determining response using both ΔADC and ΔnADC (using both clinical and radiological response classifications). Repeatability was assessed separately for pre-treatment 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 17y4m. Eighteen subjects were male (mean age 17y3m), and four were female (mean age 17y6m). The mean interval from pre-treatment scan to the initiation of biologic therapy was 4m (range 1m to 8m). The mean interval from start of biologic to post-treatment scan was 1y1m (range 5m to 2y6m).

Disease Response: Radiological Classification

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

Baseline ADC values were significantly higher pre-treatment in radiological responders compared to 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 (ΔADC) 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 ΔnADC = 0.27) and non-responders (median ΔnADC = 0.10). The change in nADC values (i.e. ΔnADC) was greater in responders than 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%) were classified as clinical non-responders. Of the clinical non-responders, two patients had also been classified as radiological non-responders. ADC and nADC values before and after treatment, classified by clinical response, are shown in Table 2 and Figure 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. ΔADC) 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 values in clinical responders (median ΔnADC = 0.21) and an increase in nADC values in non-responders (median ΔnADC = -0.12); the difference between these groups was borderline-significant (p=0.089).

Receiver operating characteristic analysis

Using radiological criteria for response classification:

nADC: 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 ≥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 (Figure 3).

ADC: Any decrease in ADC after treatment was 67% sensitive and 100% specific for distinguishing radiological responders from radiological non-responders. 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:

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

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

nADC as a Continuous Response Measure

Figure 4 shows the relationship between change in ADC/nADC (ΔADC and ΔnADC respectively) and change in SPARCC STIR score (ΔSTIR) after TNFi treatment. There was a significant, positive correlation between ΔADC and ΔSTIR (R=0.60, p=0.031) and between ΔnADC and ΔSTIR (R=0.55, p<0.01).

Repeatability

For ADC, the pre-treatment ICC was 0.98 (Bland-Altman 95% limits of agreement ± 110 x 10^{-6} mm^2/s) and the post-treatment ICC was 0.96 (Bland-Altman 95% limits of agreement ± 145 x 10^{-6} mm^2/s). For nADC, the pre-treatment ICC was 0.93 (Bland-Altman 95% limits of agreement ± 0.52) and the post-treatment ICC was 0.81 (Bland-Altman 95% limits of agreement ± 0.43).

DISCUSSION

Several previous studies have investigated DWI as a tool for monitoring sacroiliac joint inflammation, both in adults with ankylosing spondylitis (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 response to treatment as defined using radiological criteria, and may also reflect 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 since 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 three minutes

compared to six minutes), 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 be 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 pre-treatment ADC scores were significantly higher in radiological responders compared to 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 this 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 normalised ADC (nADC) values, both of which have previously been used to measure sacroiliiac joint inflammation (11–14). Our results suggest that these measurements may have different characteristics. For example, ADC measurements demonstrated superior inter-

observer repeatability compared to 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 since 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 this 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 which 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 which are not measured using STIR scoring. ADC histogram analysis (32–34) could potentially be used to separately quantify active inflammation and fat metaplasia.

Some limitations in this work have arisen due to the retrospective nature of the study. Firstly, this was a retrospective study and the sample size was small relatively small. We also had no control over the time interval between patients starting TNFi therapy and the second scan. ERA is a chronic disease and SIJ changes are expected to occur over a long timescale, but it would be desirable to scan the patients at a fixed interval after starting treatment (preferably six 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

Physician Global Assessment, Bath Ankylosing Spondylitis Index, and quality of life measures) to allow for a more complete assessment of clinical response (35,36). Whilst 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 response to treatment in adolescent ERA patients with sacroiliitis as defined using both clinical and radiological criteria. DWI is fast, objective and may facilitate patient-specific therapeutic decision making.

Funding:
This work was undertaken at UCLH/UCL, 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. MHC, SP, TJPB, YI are supported by the NIHR University College London Hospitals BRC and YI is also supported by Arthritis Research UK Grant 20164.

Disclosure Statement: The authors declare no conflicts of interest.

REFERENCES

1. Solau-Gervais E, Robin C, Gambert C, Troller S, Danner S, Gombert B, et al. Prevalence and distribution of juvenile idiopathic arthritis in a region of Western France. Jt Bone Spine. 2010;77(1):47–9.

2. Yilmaz M, Kendirli SG, Altintas DU, Karakoc GB, Inal A, Kilic M. Juvenile idiopathic arthritis profile in Turkish children. Pediatr Int. 2008;50:154–8.

3. Flatø B, Hoffmann-Vold A-M, Reiff A, Førre Ø, Lien G, Vinje O. Long-term outcome and prognostic factors in enthesitis-related arthritis: a case-control study. Arthritis Rheum. 2006;54:3573–82.

4. Maksymowych WP, Dougados M, van der Heijde D, Sieper J, Braun J, Citera G, et al. Clinical and MRI responses to etanercept in early non-radiographic axial spondyloarthritis: 48-week results from the EMBARK study. Ann Rheum Dis. 2015;annrheumdis – 2015–207596.

5. Rudwaleit M, Haibel H, Baraliakos X, Listing J, Märker-Hermann E, Zeidler H, et al. The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. Arthritis Rheum. 2009;60(3):717–27.

6. Kay J, Morgacheva O, Messing SP, Kremer JM, Greenberg JD, Reed GW, 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(1):R40.

7. Spadaro A, Iagnocco A, Baccano G, Ceccarelli F, Sabatini E, Valesini G. Sonographic-detected joint effusion compared with physical examination in the assessment of sacroiliac joints in spondyloarthritis. Ann Rheum Dis. 2009;68(10):1559–63.

8. Williamson L, Dockerty JL, Dalbeth N, McNally E, Ostlere S, Wordsworth BP. Clinical assessment of sacroiliitis and HLA-B27 are poor predictors of sacroiliitis diagnosed by magnetic resonance imaging in psoriatic arthritis. Rheumatology. 2004;43(1):85–8.

9. 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 spondyloarthropathy patients. Clin Rheumatol. 2010;29(8):849–54.

10. Maksymowych WP, Inman RD, Salonen D, Dhillon SS, Williams M, Stone M, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index

for assessment of sacroiliac joint inflammation in ankylosing spondylitis. Arthritis Care Res. 2005;53:703–9.

11. Bozgeyik Z, Ozgocmen S, Kocakoc E. Role of diffusion-weighted MRI in the detection of early active sacroiliitis. Am J Roentgenol. 2008 Oct;191(4):980–6.

12. 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.

13. 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 Feb;37(2):123–31.

14. Vendhan, K, Bray, TJP, Atkinson, D, Punwani, S, Fisher, C, Ioannou, Y, Sen, D, Hall-Craggs M. A Diffusion-based Quantification Technique for Assessment of Sacroiliitis in Adolescents with Enthesitis-related Arthritis. Br J Radiol. 2015;18:20150775.

15. Bollow M, Fischer T, Reisshauer H, Backhaus M, Sieper J, Hamm B BJ. 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(2):135–40.

16. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Kondratovich M, et al. Metrology Standards for Quantitative Imaging. Radiology. 2015;277(3):813–25.

17. Obuchowski NA, Reeves AP, Huang EP, Wang X, Buckler AJ, Kim HJG, et al. Quantitative imaging biomarkers: A review of statistical methods for computer algorithm comparisons. Stat Methods Med Res. 2014;24(1):68–106.

18. Petty RE, Southwood TR, Manners P, Baum J, Glass DN, Goldenberg J, et al. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. J Rheumatol. 2004;31:390–2.

19. Fisher, C, Ioannou, Y, Hall-Craggs, MA, Sen D. Enthesitis Related Arthritis; a New Era of Understanding. Ann Paediatr Rheumatol. 2012;1(8-16).

20. 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(8):1666–74.

21. Giannini EH, Ruperto N, Ravelli a, Lovell DJ, Felson DT, Martini a. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum. 1997;40(7):1202–9.

22. Lurati A, Pontikaki I, Teruzzi B, Desiati F, Gerloni V, Gattinara M, 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(5):1602–7.

23. Duurland CL, Wedderburn LR. Current developments in the use of biomarkers for juvenile idiopathic arthritis. Curr Rheumatol Rep. 2014;16(3):406.

24. Van Veenendaal M, Hemke R, Bos MI, Maas M, Van Rossum MAJ, Kuijpers TW. Magnetic resonance imaging in follow-up of clinical remission in juvenile idiopathic arthritis. Arthritis Rheum. 2012;64:S442–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 (Hoboken, NJ). 2014;66(11):2958–67.

26. Lories RJU, Luyten FP, de Vlam K. Progress in spondylarthritis. Mechanisms of new bone formation in spondyloarthritis. Arthritis Res Ther. 2009;11:221.

27. Machado, P, Landewe, R, Braun, J, Hermann, K-GA, Baker, D, van der Heijde D. 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, Li SR, Liu MJ, Guan J, et al. Value of normalized apparent diffusion coefficient for estimating histological grade of vesical urothelial carcinoma. Clin Radiol. 2014;69(7):727–31.

29. Do RKG, Chandanara H, Felker E, Hajdu CH, Babb JS, Kim D, 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(September):671–6.

30. Rosenkrantz AB, Khalef V, Xu W, Babb JS, Taneja SS, Doshi AM. Does normalisation improve the diagnostic performance of apparent diffusion coefficient values for prostate cancer assessment? A blinded independent-observer evaluation. Clin Radiol. 2015;

31. Le Bihan D. Apparent Diffusion Coefficient and Beyond: What Diffusion MR Imaging Can Tell Us about Tissue Structure. Radiology. 2013;268(2):318–22.

32. Pope WB, Kim HJ, Huo J, Alger J, Brown MS, Gjertson D, 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, Güler O, Tinkhauser G, Kostron H, 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, Ryoo I, Kim SC, Yeom JA, et al. Gliomas: Application of Cumulative Histogram Analysis of Normalized Cerebral Blood Volume on 3 T MRI to Tumor Grading. PLoS One. 2013;8(5).

35. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. 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, Pangan, AL, Kalabic, J, Goss, S, Unnebrink, K,

Anderson J. A Randomized, Double-Blind, Placebo-Controlled Multicenter Study of Adalimumab in Pediatric Patients with Enthesitis-Related Arthritis. Arthritis Care Res. 2015;67(11):1503–12.

37. Bray, TJ, Vendhan, K, Roberts, J, Atkinson, D, Punwani, S, Sen, D, Ioannou, Y, Hall-Craggs M. Association of the apparent diffusion coefficient with maturity in adolescent sacroiliac joints. J Magn Reson Imaging. 2016;

FIGURE CAPTIONS

Figure 1 – Response plots for nADC and ADC by radiological response. Pre- and post-treatment nADC and ADC values are shown for all 22 patients, classified according to radiological response. Patients whose nADC/ADC reduced between after treatment are shown in green, whilst patients whose nADC/ADC increased are shown in red. ADC values have units $mm^2/s \times 10^{-6}$.

Figure 2 – Response plots for nADC and ADC by clinical response. Pre- and post-treatment nADC and ADC values are shown for all 22 patients, classified according to clinical response. Patients whose nADC/ADC reduced between after treatment are shown in green, whilst patients whose nADC/ADC increased are shown in red. ADC values have units $mm^2/s \times 10^{-6}$.

Figure 3 – Receiver operating characteristic (ROC) analysis. Changes in ADC and nADC (i.e. ΔADC and ΔnADC) 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.

Figure 4 – Scatterplot showing the relationship between change in ADC and nADC (ΔADC and ΔnADC respectively) and change in SPARCC STIR score (ΔSTIR) after TNFi treatment. Positive values for ΔADC, ΔnADC and ΔSTIR represent improving inflammation, while negative values represent worsening inflammation.