Comparing the accuracy and reliability of detecting intensity of spinal inflammation on STIR sequence with ADC values in axial spondyloarthritis

Ho Yin Chung (✉ jameschunghoyin@gmail.com)

University of Hong Kong <https://orcid.org/0000-0002-0175-1346>

Tommy Tsang Cheung

University of Hong Kong

Vince Wing Hang Lau

Queen Mary Hospital

Kam Ho Lee

Queen Mary Hospital

Florence King Pui Chan

Queen Mary Hospital

Helen Hoi Lun Tsang

Queen Mary Hospital

Shirley Chiu Wai Chan

Queen Mary Hospital

Chak Sing Lau

University of Hong Kong

Research article

Keywords: Spondyloarthritis, MRI, Short tau inversion recovery sequence, Diffusion weighted imaging, intensity

Posted Date: August 13th, 2019

DOI: <https://doi.org/10.21203/rs.2.12774/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Objective To compare the accuracy and reliability of detecting the intensity of spinal inflammation on short tau inversion recovery (STIR) with the apparent diffusion coefficient (ADC) values of the active MRI lesions in axial spondyloarthritis (axSpA). **Materials and methods** Fifty active lesions in STIR sequence of spinal MRI were identified. With reference to sites of active lesions in STIR, the corresponding region of interest (ROI) on ADC map was drawn to determine the maximum ADC (ADC_{max}), mean ADC (ADC_{mean}), normalized maximum (nADC_{max}) and mean (nADC_{mean}). Four independent readers scored the identified active lesions as “intense” or “non-intense” according to the SPARCC MRI index. They were compared to various ADC parameters for assessment of accuracy and reliability. Regression analyses were used to adjust potential factors that could affect ADC. **Results** Significant differences were found in ADC_{max} between “intense” and “non-intense” lesions scored by 3 of the 4 readers (1405.7±271.4 vs 1165.8±223.8, $p=0.01$; 1420.7±272.1 vs 1209.0±248.5, $p=0.01$; 1438.0±307.2 vs 1213.6±231.0, $p=0.01$). Only 1 reader could differentiate a difference in “intense” and “non-intense” lesions with respect to ADC_{mean} (899.2±248.3 vs 711.0±222.6, $p=0.01$) and nADC_{mean} (4.4±2.1 vs 3.4±1.4, $p=0.05$). Inter-reader agreements were slight to moderate (Kappa=0.07-0.45). Reliability substantially improved when only the lowest and highest 25th percentiles of ADC values were included (Kappa=0.17-0.75). Regression analyses showed the “intense” lesions were associated with higher ADC values after adjustment for confounders. **Conclusion** Reading of STIR MRI is limited by the lack of ability in differentiating subtle differences of spinal inflammation. ADC could be an alternative method.

Introduction

Magnetic Resonance Imaging (MRI) is considered an objective method for assessment of disease activity in axial spondyloarthritis (axSpA). Spinal inflammation on MRI is the only parameter in axSpA that correlates with inflammatory cellularity in tissue biopsy (1), and is also a positive predictor of response to biologic therapy (2). These provide strong evidence that MRI could be used in disease monitoring.

Short Tau Inversion Recovery (STIR) sequence, a fat suppression sequence in MRI, is the most commonly used imaging technique in axSpA. Its ability of assessing the extent of spinal inflammation has been evaluated by various scoring methods (3–6). In contrast, apparent diffusion coefficients (ADC) of diffusion weighted imaging (DWI) is a new MRI sequence for axSpA assessment. In contrast to STIR sequence, it measures the intensity of inflammatory lesions. It reflects the magnitude of water diffusion at the tissue level (7) and is shown to be useful in measuring temporal changes of intensity of spinal inflammation in ankylosing spondylitis (AS) (8). ADC has been shown to be associated with disease activity, functional impairment and patient global assessment in axSpA in our previous publication (9). In this study, we compared the accuracy and reliability of detecting the intensity of spinal inflammation on STIR with the ADC values of the active MRI lesions in patients with axSpA according to the SPARCC MRI index method. This was done by 4 independent readers in a scoring exercise.

Materials And Methods

The data and MRI images were from an on-going multicenter cohort evaluating the utility of DWI in axSpA. The cohort has been registered in the clinical trial registry of the University of Hong Kong (HKUCTR–2087). Detailed methods have been described in our previous publications (9, 10). A brief description of the cohort is given here.

Patient recruitment

Patients with expert-diagnosed axSpA and older than 18 years of age with back pain of greater than 3 months duration were consecutively recruited from 7 rheumatology centers (Queen Mary Hospital, Grantham Hospital, Tung Wah Hospital, Pamela Youde Nethersole Eastern Hospital, Tseung Kwan O Hospital, Caritas Medical Center, and Kwong Wah Hospital) and one ophthalmology center (Hong Kong Eye Hospital). Written consent was obtained from all recruited patients. Patients pregnant or unable to undergo MRI examination were excluded from the study.

Clinical data

Clinical data collected included basic demographics, characteristics and severity of back pain, and extra-articular features. Physical examination was performed to determine the tender and swollen joint count, and spinal mobility as represented in the Bath Ankylosing Spondylitis Metrology Index (BASMI) (11). Blood tests including C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), and

human leucocyte antigen B27 (HLA-B27) were performed. Self-assessment questionnaires including Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) (12), Bath Ankylosing Spondylitis Functional Index (BASFI) (13), and Bath Ankylosing Spondylitis Global Index (BASGI) (14) were done and Ankylosing Spondylitis Disease Activity Score (ASDAS) (15) was calculated.

Radiographs

Radiographs of lumbosacral (anteroposterior views) spine were performed for the modified New York criteria for AS (16). Severity of radiographic sacroiliitis were defined as follow: grade 0, normal; grade 1, suspicious; grade 2, erosion/ sclerosis without joint space change; grade 3, sclerosis/ erosion with change in joint space or partial ankylosis; and grade 4, complete fusion. The X-rays were read by a single reader (CWSC).

MRI and interpretations

Whole spine and sacroiliac (SI) joint MRIs STIR and DWI sequences were performed consecutively in the same MRI examination in all patients using a 3T Achieva scanner (Philips Healthcare, Best, the Netherlands). The spinal MRI images were from cervical to lumbosacral levels. SI joint images were not used in this study. Free-breathing DWI with fat suppression was performed using a single-shot spin-echo echo-planar imaging sequence with 4 b-value (0, 100, 600 and 1000 sec/mm²). Details of the imaging parameters have been described in our previous publication (9, 10), with technical summary as follow: TR/TE 5000/80 (STIR), 4000/90 (DWI); field-of-view 150x240 mm² (STIR), 300x241 mm² (DWI); Slice thickness 3.5mm (STIR), 4mm (DWI).

Two readers (HHLT, CWSC) independently identified 50 active lesions for the scoring exercise. Discrepancies were resolved by consensus. Active lesions were defined as hyper-intensities in the vertebral body with no associated features of adjacent disc degeneration. With reference to sites of active lesions in STIR, the corresponding region of interest (ROI) on ADC map was drawn by a radiologist (KHL) to determine the maximum ADC (ADC_{max}) and mean ADC (ADC_{mean}). In addition, the background ADC (ADC_{bg}) was determined by drawing another ROI in a normal appearing lumbar vertebra excluding cortical endplate to calculate the mean ADC. Normalized maximum ADC (nADC_{max}) and mean ADC (nADC_{mean}) were defined by ADC_{max}/ADC_{bg} and ADC_{mean}/ADC_{bg} respectively. Figure 1 showed the lesion identified in STIR sequence and the corresponding ROI on ADC map. All MRI images and ADC values were visualized and determined using OsiriX MD v 9.5.2.

Scoring exercise

Four independent readers, blinded to clinical, radiological and ADC parameters scored the previously identified active lesions. All the 50 lesions identified were readable. Lesions were graded as “intense” when the signal intensity was similar to that of cerebrospinal fluid (3). Otherwise, they were graded as “non-intense”. The 4 readers included one musculoskeletal radiologist (VWHL) (reader 1), one rheumatologist (HYC) with 8 years of experience in reading MRI (reader 2), and a rheumatologist (TTC) (reader 3) and a medicine trainee (FKPC) (reader 4) both inexperienced in MRI interpretation. The latter two readers had received training in scoring intensity of MRI lesions according to the Spondyloarthritis research Consortium of Canada (SPARCC) MRI index (3, 4) prior to the scoring exercise and 3 lesions (not included in the analyses) were used for a validation exercises of the 4 scorers.

Statistical analyses

Clinical, radiological and MRI data were described in mean ± standard deviation (SD) or percentage. Intra-class correlation coefficient was used to determine the interobserver agreement for SPARCC MRI index scores. Reliability of the “intense” lesions scored by different readers were calculated by Cohen’s kappa coefficient (K). Overall reliability by Fleiss Kappa coefficient. Subgroup analyses were performed using 1) data included the lowest and highest 25th percentiles of maximum ADC, and 2) data included the lowest and highest 25th percentiles of mean ADC. The degree of reliability was interpreted as follows: 0.00–0.20 as slight; 0.21–0.40 as fair; 0.41–0.60 as moderate; 0.61–0.80 as substantial and 0.81–1.00 as almost perfect agreement.

We used ADC values as the “gold standard” to assess whether STIR MRI lesions could predict the true degree of inflammation. These were done by using t-tests, univariate regressions and multivariate regressions. Independent t-test was first used to compare the difference in

ADC_{max} , ADC_{mean} , $nADC_{max}$, and $nADC_{mean}$ between lesions graded as “intense” and “non-intense” by different readers. The “intense” lesions with a p-value less than 0.1 were used as independent variables in univariate linear regression analyses to determine their associations between ADC values. Independent multivariate regression models were built using “intense” lesions with a p-value less than 0.1 in univariate analyses as independent variables. In addition to the “intense” lesions, factors known or expected to be associated with a change of ADC values were also tested as regressors in univariate linear regression analyses. These included: age and male gender. Independent variables with a p-value less than 0.1 in univariate analyses were re-tested in multivariate regression analyses using “enter” mode. Results were reported as standard coefficient (SC) and regression coefficient (RC) with 95% confident interval (CI) stated. All statistics were performed using the commercial software Statistical Package for Social Sciences (SPSS) version 25. A p-value of less than 0.05 was defined as statistically significant.

Ethics approval and consent to participate

The cohort was approved by the Institutional Review Boards of The University of Hong Kong/ Hospital Authority Hong Kong West Cluster (institutional review board reference no. UW 14–085) and ethics committees of regional hospitals. It was conducted in accordance with the Declaration of Helsinki and the guidance of Good Clinical Practice, November 30, 2006.

Results

Clinical and demographic data is described in table 1. All selected patients fulfilled the Assessment in Spondyloarthritis International Society (ASAS) classification criteria for axial spondyloarthritis (axSpA). Most were also HLA-B27 positive (82.0%) and had established AS (68.0%). Our cohort was characterized by male predominance (72%), long disease duration (14.9 years), moderate to high disease activity (BASDAI 4.7; ASDAS-ESR 3.2), moderate functional impairment (BASFI 3.6), increased skeletal inflammation on MRI spine (SPARCC MRI index 17.1), and increased radiological damage (mSASSS 18.8).

Accuracy of intensity of lesions as defined by SPARCC MRI index

Table 2 shows the differences in ADC_{max} , ADC_{mean} , $nADC_{max}$, and $nADC_{mean}$ between lesions graded as “intense” and “non-intense” on STIR. There were differences in ADC_{max} between “intense” and “non-intense” lesions scored by the rheumatologist experienced in reading MRI (reader 2) (1405.7271.4 vs 1165.8223.8, $p = 0.01$), the rheumatologist inexperienced in reading MRI (reader 3) (1420.7272.1 vs 1209.0248.5, $p = 0.01$) and the medicine trainee (reader 4) (1438.0307.2 vs 1213.6231.0, $p = 0.01$). ADC_{max} between “intense” and “non-intense” lesions scored by the musculoskeletal radiologist (Reader 1) also tended to be different statistically (1343.2279.0 vs 1213.5256.0, $p = 0.09$). In contrast, $nADC_{max}$ were not different between “intense” and “non-intense” lesions scored by the 4 readers.

For ADC_{mean} of the lesions, differences were observed in “intense” and “non-intense” lesions scored by the rheumatologist inexperienced in reading MRI (reader 3) (889.6221.2 vs 704.6234.0, $p = 0.01$) and the medicine trainee (reader 4) (899.2248.3 vs 711.0222.6, $p = 0.01$). There was also tendency to be different in the scorings by the musculoskeletal radiologist (reader 1) (829.0224.1 vs 700.9251.8, $p = 0.06$). The medicine trainee was the only one who scored a difference in $nADC_{mean}$ (4.42.1 vs 3.41.4; $p = 0.05$).

The number of “intense” lesions identified by different readers according to the maximum and mean ADC values are shown in table 3. Readers 1, 2 and 3 identified more “intense” lesions in the highest 25th percentile of ADC_{max} while readers 2 and 4 identified more “intense” lesions in the highest 25th percentile of ADC_{mean} .

Agreement among readers

Agreements on intensity of lesions among readers ranged from slight to moderate (table 4). The best agreement was between the medicine trainee (reader 4) and rheumatologist inexperienced in reading MRI (reader 3) ($K = 0.45$). The worst agreement was between the 2 rheumatologists—one experienced (reader 2) and the other inexperienced (reader 3) in reading MRI ($K = 0.07$). Overall agreement by Fleiss Kappa was 0.26. Intra-class correlation coefficient between SPARCC scores of the 2 readers was 0.855.

Subgroup analyses

In the subgroup including the lowest and highest 25th percentiles of maximum ADC only, the musculoskeletal radiologist (reader 1) and rheumatologist experienced in reading MRI (reader 2) had the best agreement ($k = 0.75$). The worst agreement was found between the musculoskeletal radiologist (reader 1) and rheumatologist not experienced in reading MRI (reader 3) ($k = 0.17$). Overall agreement by Fleiss Kappa was 0.39.

In the subgroup including the lowest and highest 25th percentiles of mean ADC only, the musculoskeletal radiologist (reader 1) and rheumatologist experienced in reading MRI (reader 2) had the best agreement ($k = 0.59$). The worst agreement was found between the rheumatologist experienced in reading MRI (reader 2) and rheumatologist not experienced in reading MRI (reader 3) ($k = 0.05$). Overall agreement by Fleiss Kappa was 0.31. Results are shown in table 4.

Univariate and multivariate regression models using ADC values as dependent variables

Independent variables tested in univariate linear regression analyses included: age, male gender, and “intense” lesions. ADC_{mean} was positively associated with “intense” lesions by the musculoskeletal radiologist (reader 1) (RC 128.16[95%CI -7.17; 263.50], $p = 0.06$), “intense” lesions by the rheumatologist inexperienced in reading MRI (reader 3) (RC 185.02[95% CI 47.08; 322.96], $p = 0.01$) and “intense” lesions by the medicine trainee (reader 4) (RC 188.22[95% CI 45.29; 331.15], $p = 0.01$). ADC_{max} was negatively associated with male sex (RC -150.16[95% CI -319.61; 19.29], $p = 0.08$) and positively associated with “intense” lesions by the musculoskeletal radiologist (reader 1) (RC 129.71[95% CI -22.96; 282.38], $p = 0.09$), “intense” lesions by the rheumatologist experienced in reading MRI (reader 2) (RC 239.92[95% CI 98.94; 380.91], $p = 0.01$), “intense” lesions by the rheumatologist inexperienced in reading MRI (reader 3) (RC 211.69[95% CI 57.65; 365.73], $p = 0.05$), and “intense” lesions by the medicine trainee (reader 4) (RC 224.36[95% CI 65.78; 382.94], $p = 0.01$). $nADC_{mean}$ was positively associated with male gender (RC 1.14[95% CI 0.12; 2.15], $p = 0.03$), and “intense” lesions by the medicine trainee (reader 4) (RC 1.00[95% CI -0.002; 2.00], $p = 0.05$). $nADC_{max}$ was not associated with “intense” lesions by any readers.

Multivariate regressions showed ADC_{mean} is positively associated with “intense” lesions by the rheumatologist inexperienced in reading MRI (reader 3), and “intense” lesions by medicine trainee (reader 4). ADC_{max} is positively associated with “intense” lesions by the rheumatologist experienced in reading MRI (reader 2), “intense” lesions by the rheumatologist inexperienced in reading MRI (reader 3), and “intense” lesions by the medicine trainee (reader 4) after potential associated factors adjustment. $nADC_{mean}$ had no association with “intense” lesions by any of the readers. Results are shown in table 5.

Discussion

In this study, we compared the accuracy and reliability of grading intensity in STIR sequence in patients with axSpA with the computer-generated ADC parameters.

DWI and ADC are new MRI sequences and measurements in spinal inflammation in axSpA. They have been validated in previous studies (8–9, 17). In contrast to STIR sequence, they allow quantitative assessment (1, 8) of disease activity. Measurement of ADC however, is not without limitation. ADC has a wide degree of variability as a result of instrumental variation and errors, and biological variations. Therefore, a proposed solution is the normalized ADC, which calculated the ratio between the abnormal ADC and normal ADC values to eliminate the variations. At the present moment, there is still a lack of validation data on the two methods in axSpA. In this study, a higher ADC or normalized ADC is assumed to represent higher degree of inflammation.

We used the SPARCC MRI index as a reference method to score the intensity of MRI inflammatory lesions. The original definition was “The signal from cerebrospinal fluid constituted the reference for designating an inflammatory lesion as intense” (3). Our data shows human eye has ability in differentiating lesions with greater degree of inflammation from those with less degree of inflammation but different readers have different ways of MRI interpretation. Using this method, most (3 out of 4) readers were able to differentiate the image intensity of maximally inflamed areas. Two of the readers were also able to differentiate the intensity of the mean degrees of inflammation within the lesions. This suggests that readers tended to use the most inflamed area as the reference. As ADC_{mean} would depend on the way the ROI was drawn, ADC_{max} could represent a more objective measurement.

Overall intensity grading of STIR MRI inflammation has poor reliability. Inter-readers agreement on intensity of lesions were only slight to fair. Significant difference in ADC_{max} and ADC_{mean} were only observed in the intensity grading by the rheumatologist inexperienced in MRI

reading (reader 3). There were also significant discrepancies in the number of “intense” lesions identified by different readers. When only the most and least inflamed lesions were included in the subgroup analyses, the reliability significantly improved. In the subgroup analyses including the lowest and highest 25th percentiles of maximum ADC only, “Intense” lesions identified by the musculoskeletal radiologist (reader 1) and rheumatologist experienced in reading MRI (reader 2) achieved substantial reliability. The number of “intense” lesions graded correctly also increased in the lowest and highest 25th percentiles. As the differences in intensity of inflammation (as reflected by the ADC parameters) between the “intense” and “non-intense” lesions were small, human eye would be inferior to computer in differentiating subtle differences in intensity. Our results showed STIR MRI could be inferior to ADC in identifying lesion intensity and is compatible with another international study (18). Experience of readers is a factor to improve the reliability of MRI interpretation.

ADC could be affected by a number of factors including the way the ROI was drawn, health of the spine and skeletal maturity (19). Age and osteoporosis have also been reported to affect the ADC (20). Although we did not directly evaluate the effect of osteoporosis in our analyses, we adjusted the ADC values for age and sex, two known risk factors for osteoporosis (21–23). Upon adjustments, we still found positive associations between ADC_{max} and “intense” lesions identified by 3 of the readers. Positive associations were also found between ADC_{mean} and “intense” lesions identified by the rheumatologist inexperienced in reading MRI (reader 3), as well as nADC_{mean} and “intense” lesions identified by the medicine trainee (reader 4). The results suggested STIR MRI could differentiate the degree of inflammation despite the effect of age and sex.

STIR MRI showed poor ability to differentiate different nADC values. As a matter of fact, no difference in nADC_{max} was observed between the “intense” and “non-intense” lesions by different readers. nADC is defined as the ratio of lesion ADC to normal spine ADC. The value allows comparison between different machines. At present, we are still not sure the best way to perform the normalization. However, our study only involved one MRI machine, normalization would not be absolutely necessary. Our data also showed ADC acquired from a single MRI machine outperformed nADC as the former value appeared to be less affected by variability in interpretation.

Conclusion And Future Direction

By comparing with ADC, we showed that the STIR MRI has the ability to differentiate degree of spinal inflammation. However, it is limited by the inability to differentiate subtle differences. Moreover, different readers have different ways of MRI interpretation. ADC is an alternate method. With technological advances and development of artificial intelligence in the future of radiology (24), we believe that ADC may eventually be automatically computer-generated to replace the intensity grading in STIR sequence of patients with axSpA.

Abbreviations

MRI = Magnetic resonance imaging; STIR = short tau inversion recovery; DWI = diffusion weighted imaging; ADC = Apparent Diffusion Coefficient; HLA = Human Leucocyte Antigen; SI = sacroiliac; ASAS = Assessment in Spondyloarthritis International Society; axSpA = axial spondyloarthritis; MNY = modified New York; AS = ankylosing spondylitis; BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; BASFI = Bath Ankylosing Spondylitis Functional Index; BASMI = Bath Ankylosing Spondylitis Metrology Index; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein; ASDAS = Ankylosing Spondylitis Disease Activity Score; mSASSS = modified Stoke Ankylosing Spondylitis Spine Score; nADC = normalized Apparent Diffusion Coefficient; ADC_{max} = maximum Apparent Diffusion Coefficient; ADC_{mean} = mean Apparent Diffusion Coefficient; nADC_{max} = normalized maximum Apparent Diffusion Coefficient; nADC_{mean} = normalized mean Apparent Diffusion Coefficient.

Declarations

Ethics approval and consent to participate

The cohort was approved by the Institutional Review Boards of The University of Hong Kong/ Hospital Authority Hong Kong West Cluster (institutional review board reference no. UW 14–085) and ethics committees of regional hospitals. It was conducted in accordance with the Declaration of Helsinki and the guidance of Good Clinical Practice, November 30, 2006.

Consent for publication

Not applicable

Availability of data and material

All data generated or analysed during this study are included in this published article.

Competing interests

The authors declare no conflict of interest.

Funding

This work is supported by Novartis research grant and the Hong Kong Society of Rheumatology Project Fund.

Authors' contributions

Ho Yin Chung is responsible for study design, patients recruitment, MRI scorings, data analyses and manuscript preparation.

Tommy Tsang Cheung is responsible for MRI scorings and manuscript preparation.

Vince Wing Hang Lau is responsible for MRI scorings and manuscript preparation.

Kam Ho Lee is responsible for MRI scorings and manuscript preparation.

Florence King Pui Chan is responsible for MRI scorings and manuscript preparation.

Helen Hoi Lun Tsang is responsible for x-ray scorings, patients recruitment and manuscript preparation.

Shirley Chiu Wai Chan is responsible for x-ray scorings, patients recruitment and manuscript preparation.

Chak Sing Lau is responsible for study design, patients recruitment and manuscript preparation.

References

1. Bollow M, Fischer T, Reisschauer H, Backhaus M, Sieper J, Hamm B, et al. Quantitative analyses of sacroiliac biopsies in spondyloarthropathies: T cells and macrophages predominate in early and active sacroiliitis: cellularity correlates with the degrees of enhancement detected by magnetic resonance imaging. *Ann Rheum Dis* 2000; 59: 135-40.
2. Rudwaleit M, Schwarzlose S, Hilgert ES, Listing J, Braun J, Sieper J. MRI in predicting a major clinical response to anti-tumour necrosis factor treatment in ankylosing spondylitis. *Ann Rheum Dis* 2008; 67: 1276-81.
3. Maksymowych WP, Inman RD, Salonen D, Dhillon SS, Krishnananthan R, Stone M, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. *Arthritis Rheum* 2005; 53: 502-9.
4. 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 Rheum* 2005; 53: 703-9.
5. Braun J, Baraliakos X, Golder W, Brandt J, Rudwaleit M, Listing J, et al. Magnetic resonance imaging examinations of the spine in patients with ankylosing spondylitis, before and after successful therapy with infliximab: evaluation of a new scoring system. *Arthritis Rheum*. 2003 48(4): 1126-36.
6. Rudwaleit M, Schwarzlose S, Hilgert ES, Listing J, Braun J, Sieper J. MRI in predicting a major clinical response to anti-tumour necrosis factor treatment in ankylosing spondylitis. *Ann Rheum Dis* 2008; 67(9): 1276-1281.
7. Koh DM, Collins DJ, Orton MR. Intravoxel incoherent motion in body diffusion-weighted MRI: reality and challenges. *AJR Am J Roentgenol*. 2011; 196: 1351-61.
8. 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.

9. Lee KH, Chung HY, Xu X, Lau VWH, Lau CS. Apparent diffusion coefficient as an imaging biomarker for spinal disease activity in axial spondyloarthritis. *Radiology*. 2019; 291: 121-128.
10. Chan CWS, Tsang HHL, Li PH, Lee KH, Lau CS, Wong PYS, et al. Diffusion weighted imaging versus short tau inversion recovery sequence: Usefulness in detection of active sacroiliitis and early diagnosis of axial spondyloarthritis. *PLoS one*. 2018; 13: e0201040.
11. Jones SD, Porter J, Garrett SL, Kennedy LG, Whitelock H, Calin A. A new scoring system for the Bath Ankylosing Spondylitis Metrology Index (BASMI). *J Rheumatol* 1995;22:1609.
12. 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.
13. Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. *J Rheumatol* 1994;21:2281-5.
14. Jones SD, Steiner A, Garrett SL, Calin A. The Bath Ankylosing Spondylitis Patient Global Score (BAS-G). *Br J Rheumatol* 1996;35:66-71.
15. Lukas C, Landewe R, Sieper J, Dougados M, Davis J, Braun J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. *Ann Rheum Dis* 2009;68:18-24.
16. Van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum* 1984; 27: 361-8.
17. Gezmis E, Donmez FY, Aglidere 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.
18. Wang F, Chu C, Zhu L, Zhao C, Wei Y, Chen W et al. Whole-lesion ADC histogram analysis and the spondyloarthritis research consortium of Canada (SPARCC) MRI index in evaluating the disease activity of ankylosing spondylitis. *J Magn Reson Imaging*. 2018.
19. Bray TJ, Vendhan K, Roberts J, Atkinson D, Punwani S, Sen D, et al. Association of the apparent diffusion coefficient with maturity in adolescent sacroiliac joints. *J Magn Reason Imaging*. 2016; 44: 555-64.
20. Yeung DK, Wong SY, Griffith JF, Lau EM. Bone marrow diffusion in osteoporosis: evaluation with quantitative MRI imaging. *J Magn Reson Imaging* 2004; 19: 222-228.
21. Patsch JM, Deutschmann J, Pietschmann P. Gender aspects of osteoporosis and bone strength. *Wien Med Wochenschr*. 2011; 161: 117-23.
22. Jackson RD, Mysiw WJ. Insights into the epidemiology of postmenopausal osteoporosis: the Women's Health Initiative. *Semin Reprod Med*. 2014; 32: 454-62.
23. Akkawi I, Zmerly H. Osteoporosis: current concepts. *Joints*. 2018; 6: 122-127.
24. Wong SH, Al-Hasani H, Alam Z, Alam A. Artificial intelligence in radiology: how will we be affected? *Eur Radiol*. 2019: 141-143.

Tables

Table 1. Clinical and demographic characteristics of the studied patients

	Mean $\pm$ SD or percentage
Age (n=50)	46.9 $\pm$ 14.3
Male gender	36/50 (72%)
Disease duration (n=49)	14.9 $\pm$ 11.6
Smoker	14/49 (28.0%)
Drinker	18/47 (36.0%)
HLA-B27 positivity	41/49 (82.0%)
Fulfilled ASAS axSpA criteria	50/50 (100.0%)
Fulfilled MNY criteria for AS	34/47 (68.0%)
BASDAI (n=48)	4.7 $\pm$ 2.0
BASFI (n=48)	3.6 $\pm$ 2.7
BASMI (n=48)	4.2 $\pm$ 1.5
ESR (n=50)	38.3 $\pm$ 24.4
CRP (n=50)	1.3 $\pm$ 1.3
ASDAS-CRP (n=48)	1.9 $\pm$ 0.9
ASDAS-ESR (n=48)	3.2 $\pm$ 0.9
mSASSS (n=45)	18.8 $\pm$ 20.2
ADC _{max} (n=50)	1280.9 $\pm$ 270.4
nADC _{max} (n=50)	6.4 $\pm$ 2.7

SD=standard deviation; HLA=Human Leucocyte Antigen; ASAS= Assessment in Spondyloarthritis International Society; axSpA=axial spondyloarthritis; MNY=modified New York; AS=ankylosing spondylitis; BASDAI=Bath Ankylosing Spondylitis Disease Activity Index; BASFI=Bath Ankylosing Spondylitis Functional Index; BASMI=Bath Ankylosing Spondylitis Metrology Index; ESR=erythrocyte sedimentation rate; CRP=C-reactive protein; ASDAS=Ankylosing Spondylitis Disease Activity Score; mSASSS=modified Stoke Ankylosing Spondylitis Spine Score; ADC=Apparent Diffusion Coefficient; nADC=normalized Apparent Diffusion Coefficient

Table 2. Differences in ADC_{max}, ADC_{mean}, nADC_{max}, and nADC_{mean} between lesions graded as “intense” and “non-intense”

	Reader 1 (musculoskeletal radiologist)			Reader 2 (Rheumatologist experienced in reading MRI)			Reader 3 (Rheumatologist not experienced in reading MRI)			Reader 4 (Medicine trainee)		
	Intense (n=26)	Non-intense (n=24)	p-value	Intense (n=24)	Non-intense (n=26)	p-value	Intense (n=17)	Non-intense (n=33)	p-value	Intense (n=15)	Non-intense (n=35)	p-value
ax	1343.2 ± 279.0	1213.5 ± 256.0	0.09	1405.7 ± 271.4	1165.8 ± 223.8	0.01	1420.7 ± 272.1	1209.0 ± 248.5	0.01	1438.0 ± 307.2	1213.6 ± 231.0	0.01
ean	829.0 ± 224.1	700.9 ± 251.8	0.06	832.4 ± 256.9	707.6 ± 219.7	0.07	889.6 ± 221.2	704.6 ± 234.0	0.01	899.2 ± 248.3	711.0 ± 222.6	0.01
max	6.8 ± 2.6	5.9 ± 2.8	0.25	7.0 ± 2.8	5.8 ± 2.6	0.13	6.2 ± 2.4	6.5 ± 2.9	0.72	7.0 ± 2.9	6.1 ± 2.6	0.28
mean	4.1 ± 1.5	3.3 ± 1.8	0.10	4.1 ± 1.8	3.4 ± 1.5	0.14	3.9 ± 1.6	3.7 ± 1.7	0.66	4.4 ± 2.1	3.4 ± 1.4	0.05

ADC_{max} =maximum Apparent Diffusion Coefficient; ADC_{mean} =mean Apparent Diffusion Coefficient;
 $nADC_{max}$ =normalized maximum Apparent Diffusion Coefficient; $nADC_{mean}$ =normalized mean Apparent Diffusion Coefficient; n=number of subjects

Table 3. Comparing no of “intense” lesions with different percentiles of ADC values

Comparing number of "intense" lesions with the maximum ADC values				
	First 25.0 percentile	25.0-49.9 percentile	50.0-74.9 percentile	Last 75.0 percentile
"intense" lesions by reader 1	4/12 (33.3%)	7/13 (53.8%)	7/13 (53.8%)	8/12 (66.7%)
"intense" lesions by reader 2	2/12 (16.7%)	6/13 (46.2%)	7/13 (53.8%)	9/12 (75.0%)
"intense" lesions by reader 3	2/12 (16.7%)	4/13 (30.8%)	5/13 (38.5%)	6/12 (50.0%)
"intense" lesions by reader 4	2/12 (16.7%)	2/13 (15.4%)	6/13 (46.2%)	5/12 (42.7%)
Comparing number of "intense" lesions with the mean ADC values				
"intense" lesions by reader 1	4/12 (33.3%)	4/13 (30.8%)	11/13 (84.6%)	7/12 (58.3%)
"intense" lesions by reader 2	5/12 (41.7%)	5/13 (38.5%)	5/13 (38.5%)	9/12 (75.0%)
"intense" lesions by reader 3	2/12 (16.7%)	2/13 (15.4%)	7/13 (53.8%)	6/12 (50.5%)
"intense" lesions by reader 4	2/12 (16.7%)	2/13 (15.4%)	5/13 (38.5%)	6/12 (50.5%)

ADC= apparent diffusion coefficient

Table 4. Cohen's kappa coefficient between readers using lowest and highest 25% of ADC parameters only

All ADC values included				
	Reader 1 (musculoskeletal radiologist)	Reader 2 (Rheumatologist experienced in reading MRI)	Reader 3 (Rheumatologist not experienced in reading MRI)	Reader 4 (Medicine trainee)
Reader 1 (musculoskeletal radiologist)	--	0.36	0.17	0.25
Reader 2 (Rheumatologist experienced in reading MRI)	0.36	--	0.07	0.31
Reader 3 (Rheumatologist not experienced in reading MRI)	0.17	0.07	--	0.45
Reader 4 (Medicine trainee)	0.25	0.31	0.45	--
Including the lowest and highest 25% of maximum ADC values only				
	Reader 1 (musculoskeletal radiologist)	Reader 2 (Rheumatologist experienced in reading MRI)	Reader 3 (Rheumatologist not experienced in reading MRI)	Reader 4 (Medicine trainee)
Reader 1 (musculoskeletal radiologist)	--	0.75	0.17	0.42
Reader 2 (Rheumatologist experienced in reading MRI)	0.75	--	0.23	0.48
Reader 3 (Rheumatologist not experienced in reading MRI)	0.17	0.23	--	0.32
Reader 4 (Medicine trainee)	0.42	0.48	0.32	--
Including the lowest and highest 25% of mean values ADC only				
Reader 1 (musculoskeletal radiologist)	--	0.59	0.06	0.23
Reader 2 (Rheumatologist experienced in reading MRI)	0.59	--	0.05	0.53
Reader 3 (Rheumatologist not experienced in reading MRI)	0.06	0.05	--	0.44

Reader 4 (Medicine trainee)	0.23	0.53	0.44	--
-----------------------------	------	------	------	----

MRI=magnetic resonance image; ADC=apparent diffusion coefficient

Table 5. Multivariate linear regression analyses using ADC_{mean}, ADC_{max}, and nADC_{mean} as dependent variables and “intense” lesions identified by the 4 readers as independent variables

	ADC _{mean}			ADC _{max}			nADC _{mean}		
	SC	RC (95% CI)	p-value	SC	RC (95% CI)	p-value	SC	RC (95% CI)	p-value
Reader 1 (musculoskeletal radiologist)	N=50			N=50			--		
Male gender	--	--	--	-0.24	-141.48 (-308.52; 25.56)	0.10	--	--	--
“Intense” lesions	0.27	128.16 (-7.17; 263.50)	0.06	0.22	121.55 (-28.58; 271.67)	0.11	--	--	--
Reader 2 (Rheumatologist experienced in reading MRI)	--			N=50			--		
Male gender	--	--	--	-0.20	-121.27 (-276.92; 34.38)	0.12	--	--	--
“Intense” lesions	--	--	--	0.42	227.48 (87.60; 367.37)	0.002	--	--	--
Reader 3 (Rheumatologist not experienced in reading MRI)	N=50			N=50			--		
Male gender	--	--	--	-0.28	-166.97 (-324.27; -9.68)	0.04	--	--	--
“Intense” lesions	0.36	185.02 (47.08; 322.96)	0.01	0.39	223.00 (73.91; 372.09)	0.004	--	--	--
Reader 4 (Medicine trainee)	N=50			N=50			N=50		
Male gender	--	--	--	-0.30	-179.31 (-335.46; -23.16)	0.02	0.28	1.03 (0.03; 2.03)	0.04
“Intense” lesions	0.36	188.22 (45.29; 331.15)	0.01	0.42	244.85 (91.86; 397.85)	0.03	0.25	0.88 (-0.09; 1.86)	0.08

ADC_{max} =maximum Apparent Diffusion Coefficient; ADC_{mean}=mean Apparent Diffusion Coefficient;
nADC_{mean}=normalized mean Apparent Diffusion Coefficient; SC=standard coefficient; RC=regression coefficient;
CI=confident interval; N=number of subjects

Figures

Figure 1

Lesion identified in STIR sequence and ROI of the lesion on ADC map. STIR=short tau inversion recovery; ROI=region of interest; ADC=apparent diffusion coefficient