Article:
Wetterslev, M, Lambert, RGW, Maksymowych, WP et al. (2021) Arthritis and enthesitis in the hip and pelvis region in spondyloarthritis – OMERACT validation of two whole-body MRI methods. Seminars in Arthritis and Rheumatism, 51 (4). pp. 940-945. ISSN: 0049-0172

https://doi.org/10.1016/j.semarthrit.2021.05.006

Abstract

Objective

To validate reliability, correlation and responsiveness of two whole-body MRI scores for the hip/pelvis region in spondyloarthritis.

Methods

Assessment of hip/pelvis inflammation in 4 multi-reader exercises using the OMERACT MRI Whole-body score for Inflammation in Peripheral joints and Entheses(MRI-WIPE) and Hip Inflammation Magnetic Resonance Imaging Scoring System(HIMRISS).

Results

In exercises 3-4 (11/20 cases, respectively; 9 readers) reliability was mostly good for the 3 best calibrated readers. Median pairwise single-measure ICC for status were 0.58-0.65(WIPE-osteitis), 0.10-0.88(HIMRISS-osteitis) and for status/change 0.38-0.72/0.52-0.60(WIPE-synovitis/effusion) and 0.68-0.89/0.78-0.85(HIMRISS-synovitis/effusion). SRM was 1.23 for WIPE-osteitis, while lower for WIPE-synovitis/effusion and HIMRISS.

Conclusion

MRI-WIPE and HIMRISS may after further validation be useful in future spondyloarthritis trials.

Keywords

OMERACT, spondyloarthritis, hip, Whole-body MRI, MRI-WIPE, HIMRISS

Abbreviations[1]

1. Introduction

Inflammation in patients with axial spondyloarthritis (axSpA) and psoriatic arthritis (PsA) affects joints and entheses. Magnetic resonance imaging (MRI) can capture inflammation in both bone (osteitis/bone marrow oedema) and soft tissues[1, 2], traditionally in a limited anatomical area. Whole-body MRI (WB-MRI) allows assessment of the overall inflammatory status in joints and entheses in arthritis patients[3-5]. To enhance the use of WB-MRI in clinical trials, the Outcome Measures in Rheumatology (OMERACT) MRI in Arthritis Working Group developed the OMERACT MRI whole-body score for inflammation in peripheral joints and entheses in inflammatory arthritis (MRI-WIPE) based on definitions of core pathologies, with preliminary validation for the total body, including the hip and pelvis region[6, 7, 8]. More detailed scoring systems for assessing inflammation have been developed and validated for heels, hands and feet[8-11], but although hip arthritis is a key cause of functional impairment, no detailed scoring system for this region has been validated for inflammatory spondyloarthritides. Based on consensus in the international OMERACT MRI in Arthritis Working Group, it was decided to further develop and

[1] Abbreviations: axSpA, axial spondyloarthritis; HIMRISS, Hip Inflammation MRI Scoring System; ICC, intraclass correlation coefficient; kappa, Cohen's kappa, quadratically weighted; MR, magnetic resonance; MRI, Magnetic resonance imaging; MRI-WIPE, OMERACT, Outcome Measures in Rheumatology; OA, osteoarthritis; OMERACT MRI Whole-body score for Inflammation in peripheral joints and Entheses in inflammatory arthritis; PsA, psoriatic arthritis; RETIC, real-time iterative calibration; SpA, spondyloarthritis; SRM, standardized response mean; STIR, short-tau inversion recovery; T1w, T1-weighted; TNF, tumor necrosis factor; Whole-body MRI, WB-MRI.

validate MRI-WIPE by investigating methods for evaluation of WB-MRI for individual regions, i.e. a modular approach.

The Hip Inflammation Magnetic resonance imaging Scoring System (HIMRISS), a semiquantitative method, was developed and validated in osteoarthritis (OA) showing good reliability for status and change in bone marrow lesions [12].

Consequently, the aim of the current study was to investigate the two WB-MRI methods (MRI-WIPE and HIMRISS) for evaluation of osteitis (bone marrow oedema), synovitis and soft tissue inflammation in the hip/pelvis region in patients with SpA including PsA, and to assess interreader agreement, responsiveness and correlation between the scoring systems.

2. Materials and Method

2.1 Materials

The OMERACT MRI-WIPE scoring system was developed and preliminarily validated by The OMERACT MRI in Arthritis Working Group [6, 7]. In 2019 the group decided that a next step should be to investigate the hip/pelvis region with the MRI-WIPE and an alternative system to validate the scoring systems in accordance with the OMERACT Filter (2.1) Instrument Selection Algorithm (OFISA)[13]. HIMRISS, which has been developed and validated for hip joint OA[12, 14, 15], was chosen. In 2020, ten rheumatologists and two radiologists from 7 countries participated in 6 web-conferences and 4 web-based multi-reader exercises. Instructional written presentations for MRI-WIPE in the hip/pelvis region (WIPE-hip/pelvis) and an online real-time iterative calibration (RETIC) module for HIMRISS were available[16].

Anonymized coronal whole-body MRIs for the hip/pelvis (i.e images obtained as part of a WB-MRI-examination) were uploaded to a web-based interface hosted securely by CARE Arthritis, Edmonton, Canada. Images were displayed with semitransparent overlays for assessment using HIMRISS and data entry schematics for WIPE-hip/pelvis. Images were scored according to the semiquantitative system OMERACT MRI-WIPE [6] and the more detailed semiquantitative HIMRISS [12, 14-16] (Appendix). Images were evaluated independently in unknown order by readers with varying expertise in MRI and in the scoring systems.

In exercise 1, coronal T1-weighted (T1w) and short-tau inversion recovery (STIR) hip/pelvis images from 3 cases with axSpA were assessed by 12 readers (10/2 rheumatologists/radiologists) to train inexperienced readers and identify pitfalls.

In exercise 2, coronal T1w and STIR hip/pelvis images from 7 cases with axSpA were assessed by 9 readers (8/1 rheumatologists/radiologist) to subsequently discuss difficulties and discrepancies to further improve consensus.

In exercise 3, coronal T1w and STIR hip/pelvis images of 11 cases with axSpA or PsA and 2 timepoints (in 9 of 11 patients before and after tumor necrosis factor (TNF) inhibitor) were assessed by 9 readers (7/2 rheumatologists/radiologists). Interreader agreement was analysed for all readers and for the 3 readers with the overall highest agreement. The latter was done to evaluate the reliability among more calibrated and experienced readers. Subsequently, a selection of reference images for WIPE-hip/pelvis were discussed online to enhance understanding and reliability for the next exercise.

In exercise 4, coronal T1w and STIR hip/pelvis images from 20 cases with axSpA, PsA or peripheral SpA, 10 with 2 timepoints (before and after TNF inhibitor therapy) and 10 cases with 1 timepoint were assessed by 9 readers (7/2 rheumatologists/radiologists). Inflammation in hip/pelvis

region at baseline was not necessarily present. Interreader agreement was analysed for all readers and for the 3 readers with the overall highest agreement identified in exercise 3.

2.2 Statistics

For exercises 3-4 agreement at patient level was evaluated using intraclass correlation coefficient (ICC; two-way mixed model, single-measure, absolute agreement definition)[17, 18]. Correlations between methods were assessed with Spearman's rank correlation coefficient (rho), changes from baseline to follow-up with the Wilcoxon signed-rank test and responsiveness with the standardized response mean (SRM)[19]. Agreement at lesion level for MRI-WIPE was evaluated using Cohen's kappa (quadratically weighted)[20]. Statistical analyses were performed in SPSS version 25.0 or R version 3.6.1. $P<0.05$ was considered statistically significant.

3. Results

Twelve rheumatologists and radiologists from 7 countries participated in web-meetings and exercises. The first two exercises were used to calibrate readers. Six readers completed the RETIC modules for HIMRISS before exercise 3. Two of 3 readers with overall highest interreader agreement in exercise 3 did not complete calibration modules prior to exercise 3 but were experienced readers (1/1 radiologist/rheumatologists) and developers of one of the scoring systems.

In exercises 3-4, variations in reliability for status and change in sum scores between reader pairs were seen and overall agreement (ICC and kappa) improved when data from the readers with the overall highest interreader agreement in exercise 3 was analyzed. In exercise 3, agreement for status in osteitis was good for WIPE-hip/pelvis with ICC 0.63 (WIPE-osteitis) and very good for HIMRISS with ICC 0.88 (HIMRISS-osteitis). Interreader agreement for change in osteitis was not done due to minimal change over time in this parameter. Interreader agreement for status and

change in synovitis/effusion was good for WIPE-hip/pelvis and very good for HIMRISS with ICC 0.60/0.60 (WIPE-synovitis/effusion) and 0.89/0.78 (HIMRISS-synovitis/effusion) (Table 1).

In exercise 4, agreement varied from poor to very good. For osteitis ICCs for all patients were 0.58 (WIPE) and 0.10 (HIMRISS). For synovitis/effusion, ICCs for all patients' status/change were 0.38/0.52 (WIPE), and 0.73/0.85 (HIMRISS). In the subgroup with two timepoints ICCs for WIPE osteitis and synovitis status were 0.65 and 0.72 (Table 1).

WIPE-hip/pelvis and HIMRISS correlated significantly regarding osteitis status and for status and change in synovitis/effusion (Table 2). In exercise 4 Wilcoxon signed-rank test showed a significant change in osteitis between baseline and follow-up using WIPE-hip/pelvis and SRM was large (1.23), while it was lower for WIPE-synovitis/effusion as well as for HIMRISS (Table 2).

4. Discussion

In this OMERACT study a modular approach to whole-body MRI was applied. Inflammation in the hip/pelvis region was evaluated in patients with SpA using the two different scoring methods MRI-WIPE for the hip/pelvis region and HIMRISS. The study showed variable, but mostly good reliability for status in osteitis and for status and change in synovitis/effusion for the two methods.

This is the first OMERACT validation of HIMRISS in patients with SpA. Furthermore, this is the first study where the OMERACT MRI-WIPE is used to assess individual regions on whole-body MR images. The interreader agreement was very variable between reader pairs, in accordance with varying reader experience, training and calibration. Better agreement between experienced readers indicates that the methods will be reliable among experienced and well-calibrated readers, and that improved pre-reading calibration is required before future reading exercises.

It should be noted that WIPE-hip/pelvis and HIMRISS do not measure the same. In WIPE-hip/pelvis soft tissue and bone marrow inflammation at various locations in the hip/pelvis region are assessed individually, including entheseal regions such as the greater trochanter and ischial tuberosity. HIMRISS provides detailed assessment of osteitis and synovitis/effusion in the hip joint itself and does not include assessment of entheseal regions. Thus, the scoring systems cannot be directly compared and can be considered complementary.

Our study included a relatively small number of cases and osteitis and/or enthesitis were not necessarily present in the hip/pelvis region. The observed range of scores for osteitis was overall very small compared to the maximum possible score and only minimal change was seen. Therefore, we chose not to assess interreader agreement for change in this parameter. It would have been ideal to have an image dataset with more osteitis, synovitis/effusion and change over time. However, our WB-MRI image dataset was limited and did not allow this. Furthermore, experience of readers varied (some readers had no previous experience in scoring MRIs of this region) and not all completed the calibration modules. This was not considered obligatory, since the study was preliminary. This should be taken into consideration when interpreting the results.

In conclusion, two complementary semiquantative MRI scoring systems MRI-WIPE for the hip/pelvis region and HIMRISS, allow assessment of inflammation in the hip/pelvis region in SpA. The methods showed mostly good, but varying from poor to very good agreement between reader pairs. Before future reads, obligatory completion of prespecified calibration should be included. Furthermore, an atlas with reference images for WIPE-hip/pelvis should be available for future exercises.

The hip/pelvis region is an important part of whole-body MRI assessment in spondyloarthritis. WIPE-hip/pelvis and HIMRISS are promising outcome tools, which need further validation before general use in randomized controlled trials in patients with spondyloarthritis can be recommended.

5. Disclosures

RGWL has received consulting fees from CARE Arthritis, Parexel and Pfizer. WPM is Chief Medical Officer CARE Arthritis Limited and has acted as a paid consultant/participated in advisory boards for AbbVie, Boehringer Ingelheim, Celgene, Eli Lilly, Galapagos, Janssen, Novartis, Pfizer and UCB, received research and/or educational grants from AbbVie, Novartis, Pfizer and UCB, and received speaker fees from AbbVie, Janssen, Novartis, Pfizer and UCB. SJP has been an advisory board member for AbbVie and Novartis, received research support from AbbVie, MSD, and Novartis and speaker fees from MSD, Pfizer, AbbVie, Novartis and UCB. PB participated in advisory boards and received speaker fees from Janssen, Abbvie, UCB, Celgene, BMS, Novartis, Pfizer, Gilead, Eli-Lilly. HMO has received research grants from Janssen and Novartis, and honoraria/speaker fess from AbbVie, Celgene, Janssen, Lilly, Novartis, Pfizer, Takeda and UCB. PC has received research grants from UCB, MSD and Pfizer, speaker/consultant fees for Pfizer, MSD, Novartis, BMS, AbbVie, UCB, Eli Lilly, Gilead and Celgene. PGC has received speaker or consultancy fees from AbbVie, AstraZeneca, BMS, Eli Lilly, EMD Serono, Flexion Therapeutics, Galapagos, Gilead, Novartis, Pfizer and Stryker. MØ has received research support, consultancy fees and/or speaker fees from Abbvie, BMS, Boehringer-Ingelheim, Celgene, Eli-Lilly, Hospira, Janssen, Merck, Novartis, Pfizer, Regeneron, Roche, Sandoz, Sanofi and UCB. MW, IE, MS, SK, AM, VF, FG, JP, GDM, AEFP and JLJ have no declarations for this work.

6. Acknowledgement

We thank CARE Arthritis Limited (carearthritis.com) for help with setting up the web-based scoring interface, scoring exercises, and the web-based meetings. We acknowledge the contribution of SIG (Special Interest Group) participants at the virtual OMERACT meeting October 29, 2020.

HMO, GDM and PGC are supported in part by the National Institute for Health Research (NIHR) Leeds Biomedical Research Centre, United Kingdom. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

7. Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Sudoł-Szopińska I, Matuszewska G, Kwiatkowska B, et al. Diagnostic imaging of psoriatic arthritis. Part I: etiopathogenesis, classifications and radiographic features. J Ultrason. 2016;16(64):65-77.

2. Sudoł-Szopińska I, Pracoń G. Diagnostic imaging of psoriatic arthritis. Part II: magnetic resonance imaging and ultrasonography. J Ultrason. 2016;16(65):163-74.

3. Poggenborg RP, Pedersen SJ, Eshed I, et al. Head-to-toe whole-body MRI in psoriatic arthritis, axial spondyloarthritis and healthy subjects: first steps towards global inflammation and damage scores of peripheral and axial joints. Rheumatology (Oxford). 2015;54(6):1039-49.

4. Poggenborg RP, Eshed I, Ostergaard M, et al. Enthesitis in patients with psoriatic arthritis, axial spondyloarthritis and healthy subjects assessed by 'head-to-toe' whole-body MRI and clinical examination. Ann Rheum Dis. 2015;74(5):823-9.

5. Krabbe S, Østergaard M, Eshed I, et al. Whole-body Magnetic Resonance Imaging in Axial Spondyloarthritis: Reduction of Sacroiliac, Spinal, and Entheseal Inflammation in a Placebo-controlled Trial of Adalimumab. J Rheumatol. 2018;45(5):621-9.

6. Krabbe S, Eshed I, Gandjbakhch F, et al. Development and Validation of an OMERACT MRI Whole-Body Score for Inflammation in Peripheral Joints and Entheses in Inflammatory Arthritis (MRI-WIPE). J Rheumatol. 2019;46(9):1215-21.

7. Østergaard M, Eshed I, Althoff CE, et al. Whole-body Magnetic Resonance Imaging in Inflammatory Arthritis: Systematic Literature Review and First Steps Toward Standardization and an OMERACT Scoring System. J Rheumatol. 2017;44(11):1699-705.

8. Boers M, Kirwan JR, Tugwell P, Beaton D, Bingham CO III, Conaghan PG, et al. The OMERACT Handbook 2018 [cited 2021 January 7]. Available from: https://www.omeract.org/handbook.

9. Ostergaard M, McQueen F, Wiell C, et al. The OMERACT psoriatic arthritis magnetic resonance imaging scoring system (PsAMRIS): definitions of key pathologies, suggested MRI sequences, and preliminary scoring system for PsA Hands. J Rheumatol. 2009;36(8):1816-24.

10. Mathew AJ, Krabbe S, Kirubakaran R, et al. Utility of Magnetic Resonance Imaging in Diagnosis and Monitoring Enthesitis in Patients with Spondyloarthritis: An OMERACT Systematic Literature Review. J Rheumatol. 2019;46(9):1207-14.

11. Mathew AJ, Krabbe S, Eshed I, et al. The OMERACT MRI in Enthesitis Initiative: Definitions of Key Pathologies, Suggested MRI Sequences, and a Novel Heel Enthesitis Scoring System. J Rheumatol. 2019;46(9):1232-8.

12. Jaremko JL, Lambert RGW, Pedersen SJ, et al. OMERACT Hip Inflammation Magnetic Resonance Imaging Scoring System (HIMRISS) Assessment in Longitudinal Study. J Rheumatol. 2019;46(9):1239-42.

13. D'Agostino MA, Beaton DE, Maxwell LJ, Cembalo SM, Hoens AM, Hofstetter C et al. Improving domain definition and outcome instrument selection: Lessons learned for OMERACT from imaging. Semin Arthritis Rheum. 2021 (In Press).

14. Maksymowych WP, Cibere J, Loeuille D, et al. Preliminary validation of 2 magnetic resonance image scoring systems for osteoarthritis of the hip according to the OMERACT filter. J Rheumatol. 2014;41(2):370-8.

15. Maksymowych WP, Pitts M, Budak MJ, et al. Development and Preliminary Validation of a Digital Overlay-based Learning Module for Semiquantitative Evaluation of Magnetic Resonance Imaging Lesions in Osteoarthritis of the Hip. J Rheumatol. 2016;43(1):232-8.

16. [cited 2021 January 10]. Available from: https://www.carearthritis.com/mriportal/himriss/index/.

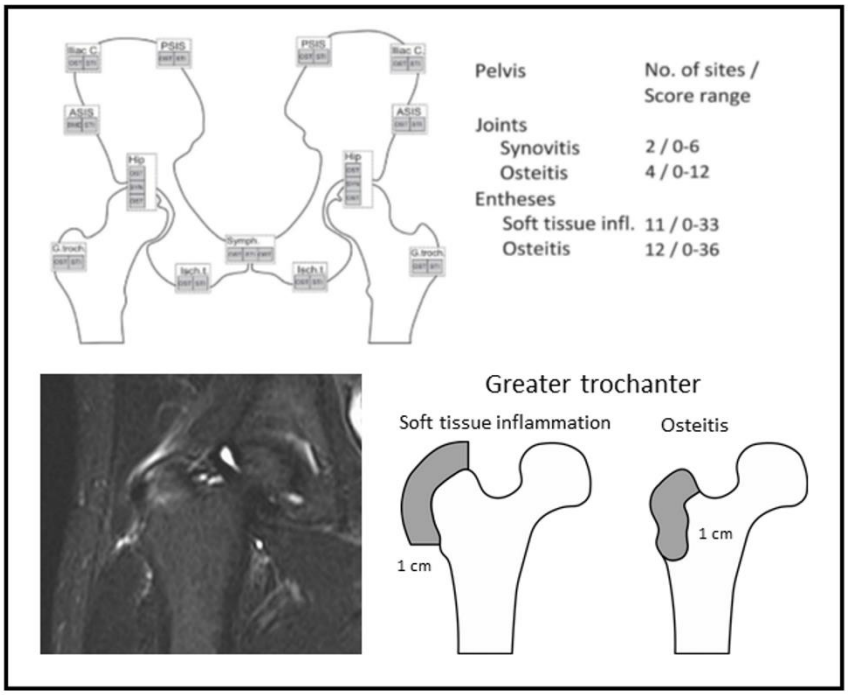
17. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155-63.

18. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol. 2012;8(1):23-34.

19. Husted JA, Cook RJ, Farewell VT, et al. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol. 2000;53(5):459-68.

20. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82.

Figure A.1.



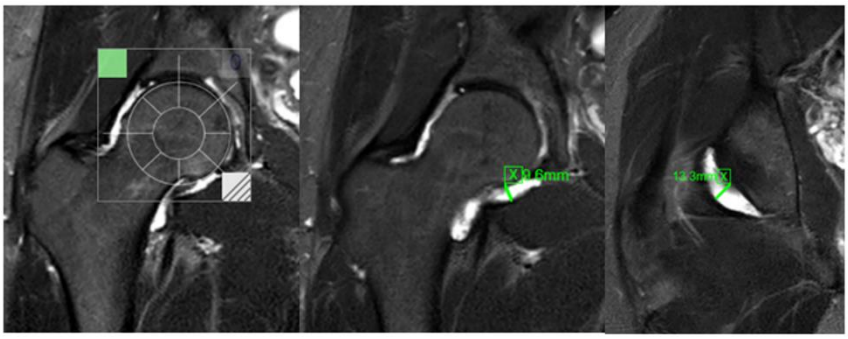
MRI-WIPE schematic and scoring ranges for the hip and pelvis insertion of region (upper row), from Krabbe et al.[6] and a schematic drawing of the principle of scoring osteitis and soft tissue inflammation using MRI-WIPE (lower row), illustrated by the greater trochanter (coronal STIR whole-body MR image of hip region shown to the left).

The top row illustrates that osteitis of the hip/pelvis is assessed separately for acetabulum and femur and osteitis of the pubic symphysis is assessed separately for left and right pubic bone. Using MRI-WIPE osteitis is assessed in the bone from the articular surface/entheseal insertion to a depth of 1 cm on all available images (as shown in schematic of the greater trochanter assessment). The osteitis grading scale is 0-3 based on the proportion of bone with oedema, compared to the

“assessed bone volume”, judged on all available images: 0: normal; 1: mild (1-33% of bone oedematous); 2: moderate (34-66% of bone oedematous); 3: severe (67-100% of bone oedematous). Soft tissue inflammation is assessed inside the ligament/tendon and it’s the immediate surroundings to 1 cm from the entheseal insertion: 0: normal; 1: mild; 2: moderate; 3: severe – by thirds of the maximum potential volume of inflammatory tissue. Synovitis is assessed in the entire synovial compartment on all available images: 0: normal; 1: mild; 2: moderate; 3: severe – by thirds by thirds of the maximum potential volume of enhancing tissue in the synovial compartment[6].

PSIS, posterior superior iliac spine; Iliac C.: iliac crest; ASIS: anterior superior iliac spine; G. troch: greater trochanter; Isch.t.: ischial tuberosity; Symph.: pupic symphysis; OST, osteitis; STI, soft tissue inflammation; SYN, synovitis; MRI-WIPE, OMERACT MRI Whole-body score for Inflammation in peripheral joints and Entheses in inflammatory arthritis; MR, magnetic resonance; STIR, short-tau inversion recovery.

Figure A.2.



Coronal STIR whole-body MR image of the hip with the web-based semi-transparent HIMRISS overlay positioned for osteitis scoring in femoral head and acetabulum (right hip) and 2 examples of synovitis/effusion measuring according to HIMRISS. Osteitis is scored on consecutive sagittal slices through the hip joint. The reader marks the first slice and the last slice where femur bone is visible, and the overlay is moved by the reader to fit the central slices (the slice where the femoral head appears the largest). The overlay separates subarticular bone in the femoral head and acetabulum into approximately 1x1 cm regions. On each slice, the reader clicks each area with osteitis and sum scores of these regions are automatically calculated and adjusted for the scoring range of each region (total scoring range 0-100). Hip synovitis/effusion is measured in each coronal

image as the longest diameter perpendicular to the longest axis of synovitis/effusion collection (score 0: 0-1.9 mm; score 1: 2-3.9 mm; score 2: ≥4 mm, scoring range 0-30)[12, 14-16].

STIR, short-tau inversion recovery; MR, magnetic resonance; HIMRISS, Hip Inflammation MRI Scoring System.

Tables

Table 1. MRI-WIPE in the hip/pelvis and HIMRISS interreader reliability for exercises 3 and 4

			MRI-WIPE hip/pelvis						HIMRISS			
			Osteitis			Synovitis/effusion			Osteitis		Synovitis/effusion	
Variables	**No. patients (cases)**	**Type of score**	**Mean score**	**ICC**	**Kappa**	**Mean score**	**ICC**	**Kappa**	**Mean score**	**ICC**	**Mean score**	**ICC**
Exercise 3 9 readers	11	Status	2.3 (0-10)	0.69 (0.23-0.93)	0.50 (0.15-0.93)	1.4 (0-4)	0.58 (-0.06-0.96)	0.62 (-0.06-0.96)	8.2 (1-60)	0.84 (0.56-0.99)	12.8 (3-25)	0.52 (0.00-.91)
	11	Change	-0.2 (-1-1)	NA	NA	-0.2 (-3-1)	0.50 (0.10-0.87)	0.30 (-0.25-0.90)	-0.35 (-3-1)	NA	-1.8 (-17-10)	0.50 (-0.05-0.89)
Exercise 3 3 readers	11	Status	1.8 (0-10)	0.63 (0.46-0.93)	0.47 (0.27-0.62)	1.7 (0-5)	0.60 (0.34-0.80)	0.78 (0.68-0.85)	6.6 (0-65)	0.88 (0.77-0.94)	12.8 (2-28)	0.89 (0.87-0.91)
	11	Change	-0.12 (-1-1)	NA	NA	-0.12 (-3-2)	0.60 (0.48-0.83)	0.48 (0.24-0.90)	-0.7 (-7-0)	NA	-1.6 (-21-8)	0.78 (0.70-0.87)
Exercise 4 9 readers	10 (case 1-10)	Status	1.2 (0-4)	0.21 (-0.39-0.91)	0.21 (-0.03-0.66)	1.1 (0-2)	0.19 (-0.31-0.69)	0.17 (-0.20-0.61)	1.8 (0-6)	0.07 (-0.17-0.83)	16.4 (9-23)	0.31 (0.00-0.89)
	10 (case 11-20)	Status	1.6 (0-6)	0.51 (-0.08-0.99)	0.55 (0-0.92)	1 (0-3)	0.40 (-0.17-0.88)	0.52 (0.02-0.90)	3.5 (1-8)	0,08 (-0.21-0.95)	11.2 (5-24)	0.49 (0.00-0.94)
	10 (case 11-20)	Change	-0.4 (-2-0)	NA	NA	-0.39 (-2-0)	0.22 (-0.68-0.83)	0.31 (-0.09-0.71)	-2.2 (-7-2)	NA	-5.2 (-18-0)	0.57 (0.02-0.92)
	20 (case 1-20)	Status	1.4 (0-6)	0.41 (-0.35-0.92)	0.44 (0-0.76)	1.0 (0-3)	0.27 (-0.07-0.75)	0.36 (0.01-0.83)	2.7 (0-9)	0.09 (-0.17-0.85)	13.8 (5-25)	0.45 (0.01-0.90)
Exercise 4 3 readers	10 (case 1-10)	Status	0.8 (0-4)	0.29 (0.01-0.78)	0.25 (0.10-0.36)	1.3 (0-2)	-0.02 (-0.29-0.12)	0.01 (-0.20-0.13)	0.4 (0-2)	-0.04 (-0.04-0.04)	15.8 (5-26)	0.73 (0.59-0.89)
	10 (case 11-20)	Status	1.8 (0-9)	0.65 (0.52-0.76)	0.63 (0.54-0.75)	1.2 (0-4)	0.72 (0.62-0.81)	0.71 (0.62-0.80)	1.7 (0-5)	0.06 (-0.17-0.35)	9.2 (2-26)	0.68 (0.53-0.88)
	10 (case 11-20)	Change	-0.6 (-2-0)	NA	NA	-0.5 (-3-1)	0.52 (0.49-0.55)	0.48 (0.41-0.59)	-0.2 (-2-1)	NA	-2.8 (-19-6)	0.85 (0.82-0.88)
	20 (case 1-20)	Status	1.3 (0-9)	0.58 (0.43-0.69)	0.55 (0.46-0.66)	1.2 (0-4)	0.38 (0.31-0.44)	0.42 (0.35-0.49)	1.0 (0-5)	0.10 (-0.09-0.33)	12.5 (2-26)	0.73 (0.69-0.77)

Sum scores are mean (range) of the patient scores (each patient´s score is the average of the scores assigned to that patient). ICC and Kappa values are mean (range). NA: not done, due to minimal findings/change over time in this parameter. MRI-WIPE hip range for osteitis is 0-48 and for synovitis/effusion 0-6[6]. HIMRISS osteitis total range is 0-100 and range for synovitis/effusion is 30[12, 14, 16]. ICC is 2-way model, single measure, by absolute agreement. ICC values ≤0.49 were considered as poor, 0.50–0.79 as good, ≥0.80 as very good. Scorings at lesion level were assessed using Cohen's kappa, quadratically weighted. Kappa 0–0.20 was considered as no agreement, 0.21–0.39 as slight, 0.40–0.59 as weak, 0.60–0.79 as moderate, 0.80–0.90 as strong and >0.90 as almost perfect agreement[20]. Readers: ⁺IE, MW, MØ*, PB, SJP, WPM* (all exercises), ⁺RGL*, VF (exercise 1, 3, 4), MS (exercise 1, 2, 4), AM (exercise 1-3), SK (exercise 1, 2), FG (exercise 1).⁺Musculoskeletal radiologist. *the readers with overall highest agreement in Exercise 3 (MØ, RGL, WPM).

HIMRISS, Hip Inflammation MRI Scoring System; ICC, intraclass correlation coefficient; Kappa: Cohen's Kappa, quadratic weighted; MRI-WIPE, OMERACT MRI Whole-body score for Inflammation in Peripheral joints and Entheses in inflammatory arthritis.

Table 2. Sensitivity to change and correlation between methods in exercises 3 and 4[1]

Exercise 3	Baseline	Follow-up	Change	p-value	SRM
MRI-WIPE hip/pelvis					
Osteitis	1.8 (3.13)	1.7 (3.21)	-0.1 (1.19)	0.279	0.10
Synovitis/effusion	1.7 (1.66)	1.6 (1.74)	-0.1 (0.34)	0.891	0.35
HIMRISS					
Osteitis	6.6 (19.32)	5.9 (17.35)	-0.7 (1.99)	0.109	0.35
Synovitis/effusion	12.8 (8.84)	11.2 (8.80)	-1.6 (7.23)	0.562	0.23
Correlation MRI-WIPE hip/pelvis vs. HIMRISS					
Osteitis	0.77** (0.006)	0.63* (0.04)	0.32 (0.337)	-	-
Synovitis/effusion	0.89***(<0.001)	0.77** (0.006)	0.63* (0.039)	-	-
Exercise 4					
MRI-WIPE hip/pelvis					
Osteitis	1.8 (2.62)	1.2 (2.51)	-0.6* (0.51)	0.011	1.23
Synovitis/effusion	1.2 (1.46)	0.7 (0.96)	-0.5 (1.22)	0.203	0.41
HIMRISS					
Osteitis	1.6 (1.68)	1.4 (1.73)	-0.2 (0.72)	0.465	0.28
Synovitis/effusion	9.2 (7.33)	6.4 (2.08)	-2.8 (7.24)	0.415	0.39
Correlation MRI-WIPE hip/pelvis vs. HIMRISS					
Osteitis	0.72* (0.019)	0.94***(<0.001)	0.53 (0.115)	-	-
Synovitis/effusion	0.83**(0.003)	0.16 (0.651)	0.73*(0.017)	-	-

[1]Values are shown for the 3 readers with overall highest interreader agreement in exercise 3 (WPM, RL, MØ). Data are shown as mean (SD) and correlation coefficient (p-value). Comparison of scores at timepoints are done with Wilcoxon signed-rank test. Spearman Rank Correlation analysis is done for baseline and change for MRI-WIPE versus HIMRISS. Standardized response mean (SRM) is calculated as mean change score divided by standard deviation (SD) of the change score and interpreted as follows: no: <0.20; small: ≥0.20 and <0.50; moderate: ≥0.50 and <0.80; large ≥0.80[19]. *p<0.05, **p<0.01, ***p<0.001.

HIMRISS, Hip Inflammation MRI Scoring System; MRI-WIPE, OMERACT MRI Whole-body score for Inflammation in Peripheral joints and Entheses in inflammatory arthritis; SRM, standardized response mean.