



















# The OMERACT MRI in Enthesitis Initiative: Definitions of Key Pathologies, Suggested MRI Sequences, and a Novel Heel Enthesitis Scoring System

Ashish J. Mathew , Simon Krabbe , Iris Eshed , Frédérique Gandjbakhch , Paul Bird , Susanne J. Pedersen , Maria S. Stoeniou , Violaine Foltz , Daniel Glinatsi , Robert G. Lambert , Kay Geert A. Hermann , Walter P. Maksymowych , Ida K. Haugen , Jacob L. Jaremko , René P. Poggenborg, Joel Paschke, Jean-Denis Laredo , Philippe Carron , Philip G. Conaghan , and Mikkel Østergaard 

**ABSTRACT. Objective.** To develop and validate an enthesitis magnetic resonance imaging (MRI) scoring system for spondyloarthritis/psoriatic arthritis, using the heel as model.

**Methods.** Consensus definitions of key pathologies and 3 heel enthesitis multireader scoring exercises were done, separated by discussion, training, and calibration.

**Results.** Definitions for bone and soft tissue pathologies were agreed. In the final exercise, median pairwise single-measures intraclass correlation coefficients (ICC; patient-level) for enthesal inflammation status/change scores were 0.83/0.82 for all readers. For radiologists and selected rheumatologists, ICC were 0.91/0.84 and quadratic-weighted  $\kappa$  (lesion-level) 0.57–0.91/0.45–0.81.

**Conclusion.** The proposed definitions and Heel Enthesitis Scoring System (HEMRIS) are reliable among trained readers and promising for clinical trials. (J Rheumatol First Release April 1 2019; doi:10.3899/jrheum.181093)

## Key Indexing Terms:

ENTHESOPATHY    MAGNETIC RESONANCE IMAGING    SPONDYLOARTHROPATHY  
PSORIATIC ARTHRITIS    OMERACT

From the Department of Clinical Immunology and Rheumatology, Christian Medical College, Vellore, India; Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen; Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet, Glostrup, Denmark; Department of Diagnostic Imaging, Sheba Medical Center, Affiliated to the Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel; Hôpitaux Universitaires Pitié Salpêtrière, Paris 6 University, GRC-UPMC 08, Pierre Louis Institute of Epidemiology and Public Health; Service de Radiologie, Hôpital Lariboisière, AP-HP and Université Paris-Diderot, Paris, France; Division of Medicine, University of New South Wales, Sydney, Australia; Cliniques Universitaires Saint-Luc, Institut de Recherche Expérimentale et Clinique (IREC), Université catholique de Louvain, Brussels; Department of Rheumatology, Ghent University Hospital, Ghent, Belgium; Department of Radiology and Diagnostic Imaging, and CaRE (Canadian Research Education) Arthritis and Department of Medicine, University of Alberta, Edmonton, Alberta, Canada; Department of Radiology, Arthritis Imaging Research Group, University Hospital Charité, Berlin, Germany; Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway; Faculty of Medicine and Health, University of Leeds, Leeds, UK.

PGC is supported in part by the UK National Institute for Health Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health.

A.J. Mathew, MBBS, DNB, DM, Associate Professor, Department of Clinical Immunology and Rheumatology, Christian Medical College, and PhD Fellow, Department of Clinical Medicine, Faculty of Health and Medical Sciences, COPECARE, Center for Rheumatology and Spine Diseases, Rigshospitalet Glostrup, University of Copenhagen; S. Krabbe, MD, PhD Fellow, Department of Clinical Medicine, Faculty of Health and Medical Sciences, COPECARE, Center for Rheumatology and Spine Diseases, Rigshospitalet Glostrup, University of Copenhagen; I. Eshed, MD, Professor, Department of Diagnostic Imaging, Sheba Medical Center, Affiliated to the Sackler School of Medicine, Tel Aviv University; F. Gandjbakhch, MD, Practicing Rheumatologist, Hôpitaux Universitaires

Pitié Salpêtrière, Paris 6 University, GRC-UPMC 08, Pierre Louis Institute of Epidemiology and Public Health; P. Bird, B Med (Hons), FRACP, PhD, Grad Dip MRI, Associate Professor, University of New South Wales; S.J. Pedersen, MD, PhD, Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet; M.S. Stoeniou, MD, PhD, Cliniques Universitaires Saint-Luc, Institut de Recherche Expérimentale et Clinique (IREC), Université catholique de Louvain; V. Foltz, MD, Practising Rheumatologist, Hôpitaux Universitaires Pitié Salpêtrière; D. Glinatsi, MD, PhD, Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet; R.G. Lambert, MB BCh, FRCP, Professor, Department of Radiology and Diagnostic Imaging, University of Alberta; K.G. Hermann, MD, PhD, Senior Consultant, Department of Radiology, Arthritis Imaging Research Group, University Hospital Charité; W.P. Maksymowych MD, FRCP(C), Professor, Division of Rheumatology, Faculty of Medicine and Dentistry, University of Alberta; I.K. Haugen, MD, PhD, Department of Rheumatology, Diakonhjemmet Hospital; J.L. Jaremko, MD, PhD, FRCP, Associate Professor, Department of Radiology, University of Alberta; R.P. Poggenborg, MD, PhD, Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet; J. Paschke, BSc, CaRE (Canadian Research Education) Arthritis; J.D. Laredo, MD, Professor, Service de Radiologie, Hôpital Lariboisière, AP-HP and Université Paris-Diderot; P. Carron, MD, PhD, Department of Rheumatology, Ghent University Hospital; P.G. Conaghan, MB, BS, PhD, FRACP, FRCP, Professor of Musculoskeletal Medicine, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals National Health Service Trust; M. Østergaard, MD, PhD, DMSc, Professor, Department of Clinical Medicine, Faculty of Health and Medical Sciences, COPECARE, Center for Rheumatology and Spine Diseases, Rigshospitalet Glostrup, University of Copenhagen.

Address correspondence to Dr. M. Østergaard, COPECARE – Center for Rheumatology and Spine Diseases, Rigshospitalet, Nordre Ringvej 57, DK-2600 Glostrup, Denmark. E-mail: mo@dadlnet.dk

Accepted for publication January 9, 2019.

Enthesitis — inflammation at insertion sites of ligaments, fasciae, tendons, and joint capsules to bone — is a central feature of spondyloarthritis (SpA), including psoriatic arthritis (PsA). Sensitive and objective assessment of enthesitis is important in SpA clinical trials. Conventional clinical methods have limited reliability, validity, and sensitivity<sup>1,2,3</sup>. Magnetic resonance imaging (MRI) is a sensitive method for detecting enthesitis in peripheral SpA and the only method allowing detection of perientheseal osteitis<sup>4,5,6</sup>. MRI studies have demonstrated decreased entheseal inflammation after anti-tumor necrosis factor (TNF) therapy, but no validated MRI scoring systems exist for evaluating enthesitis in clinical trials<sup>7</sup>. Our aim was to create consensus-based MRI definitions of key enthesitis pathologies and through multireader exercises to develop and validate an MRI score for assessing enthesitis in patients with SpA, focusing on the heel region.

## MATERIALS AND METHODS

The Outcome Measures in Rheumatology (OMERACT) MRI in Arthritis Working Group initially performed a systematic literature review (SLR) of studies with MRI being used for assessment of enthesitis<sup>8</sup>. Based on this SLR,

MRI sequences for optimal visualization of enthesitis were identified, and MRI definitions of key enthesitis pathologies were decided by consensus among group members through meetings/e-mails. The heel region (insertions of Achilles tendon and plantar fascia) was chosen for initial testing because of its frequent involvement. Three multireader exercises, with consensus discussion and calibration in-between, were then performed. A graphical data entry schematic (Figure 1) was created, and subsequently a Web-based interface that simultaneously displayed DICOM (Digital Imaging and Communications in Medicine) images and the data entry schematic (Figure 2). In exercise 1, performed to identify challenges and pitfalls, sagittal T1-weighted (T1W) and sagittal and axial T2W fat-suppressed (T2wFS) MR images of 10 ankles [4 inflammatory enthesitis (peripheral SpA), 4 mechanical enthesitis, and 2 normal controls] were scored by 15 readers from 10 countries, with varying expertise in ankle MRI, for enthesitis at Achilles tendon and plantar fascia insertions. This was followed by a Web-based calibration exercise leading to minor score sheet modifications. In exercise 2, 16 ankle MRI [8 inflammatory enthesitis (peripheral SpA), 3 mechanical enthesitis, and 5 normal controls; MRI sequences as above] were scored by 16 readers. In exercise 3, ankle MRI (sagittal T2wFS only) of 21 patients with SpA from a clinical trial, obtained before and after anti-TNF therapy, were scored for inflammatory pathologies by 10 readers, blinded to chronological order. For assessing the reliability scores among the more experienced readers, agreement was analyzed separately between the participating radiologists and the 3 rheumatologists with best overall intraclass correlation coefficient (ICC) for inflammatory pathologies in exercise 2.

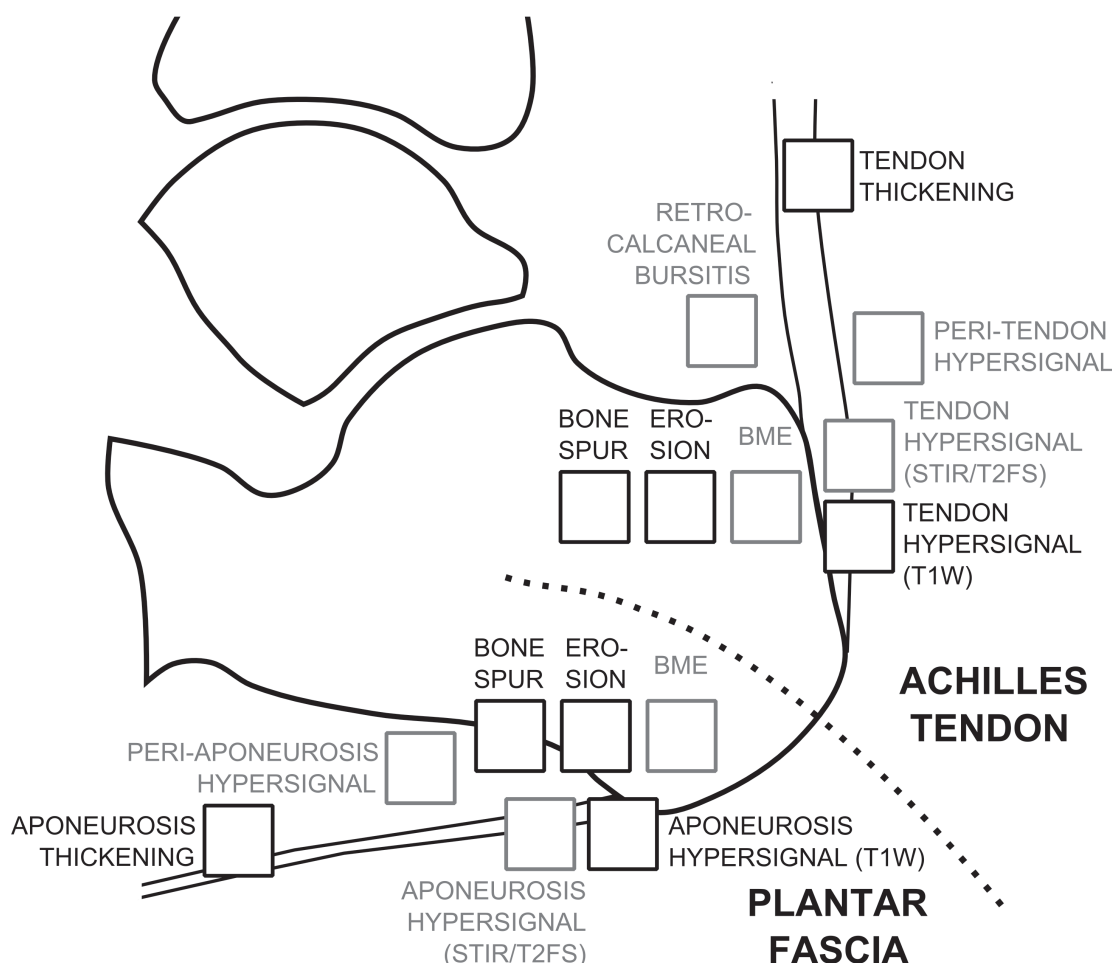


Figure 1. Line drawing of the scoring sheet used in the scoring exercises. BME: bone marrow edema.

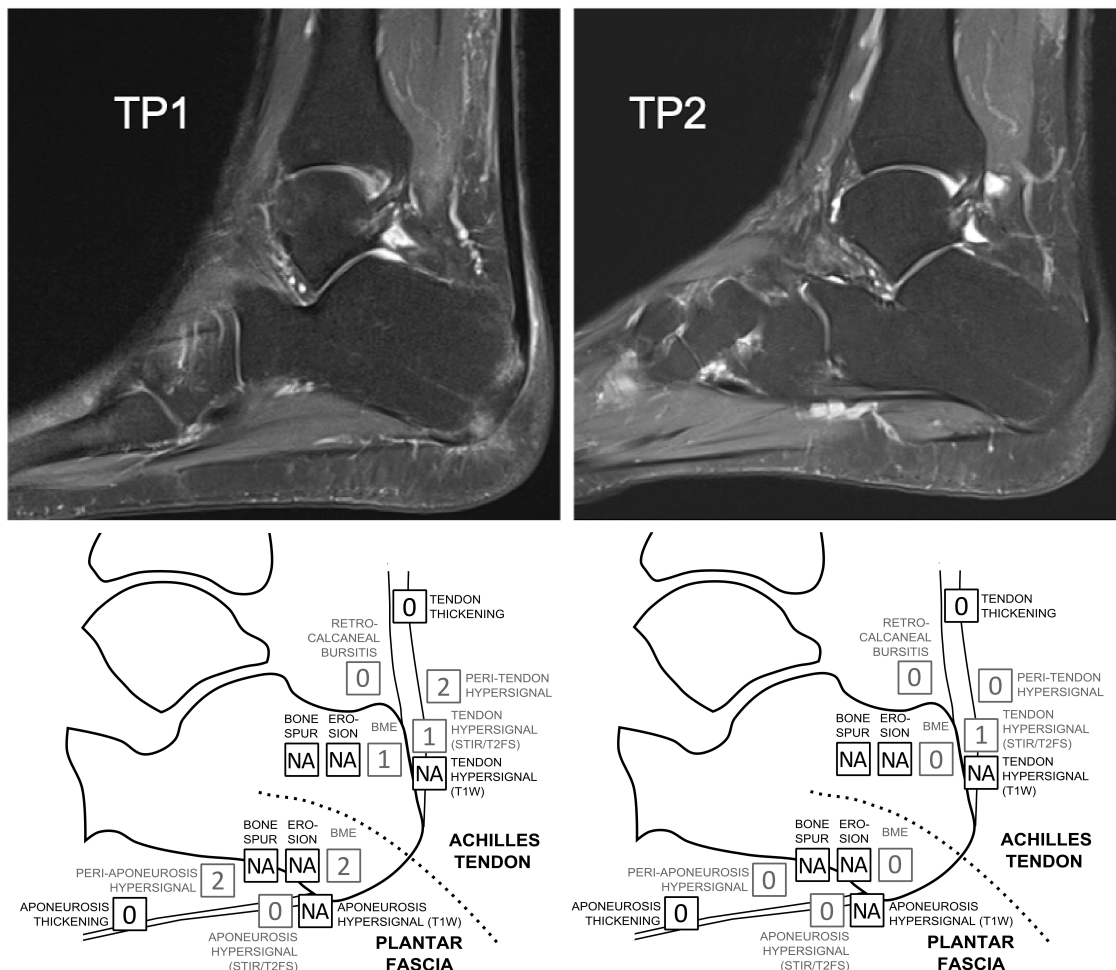


Figure 2. Web-based reader platform used for scoring exercises. T2W fat-suppressed images from the same patient before (TP1) and after (TP2) antitumor necrosis factor therapy. Below the images, data entry forms are shown, which are used to enter the scores of the individual pathologies.

**Statistical analysis.** Exercise 1 was mainly used for qualitative training and understanding principles and pitfalls, while for exercises 2–3 reliability statistics [pairwise single-measures and average-measures ICC by absolute agreement for sum scores (patient level) and squared weights Cohen’s  $\kappa$  for individual component scores (lesion level)] were calculated. In exercise 3, the standardized response mean (SRM) was calculated.

## RESULTS

**Definitions of key pathologies.** Key enthesal pathologies were selected and their definitions agreed upon by consensus within the OMERACT MRI in Inflammatory Arthritis Working Group (Table 1), based on knowledge from an SLR<sup>8</sup>, and published OMERACT MRI definitions for comparable conditions<sup>9,10,11</sup>. The selected pathologies were intratendon hypersignal (enthesal tendonitis), peritendon hypersignal (enthesal peritendonitis), bone marrow edema (BME; enthesal osteitis), bursitis, tendon thickening, enthesophyte, enthesal bone erosion, and intratendon hypersignal on T1W sequence.

**MRI sequences and planes.** For evaluating inflammatory pathologies, it was agreed to include a fluid-sensitive

sequence [short-tau inversion recovery (STIR) or T2wFS], and/or a fat-suppressed T1W sequence following intravenous gadolinium (Gd) injection (Figure 3). A T1W sequence prior to contrast injection (T1-pre-Gd) was considered helpful in determining the exact localization of inflammatory pathologies because of its high anatomical resolution and is essential for assessment of structural pathologies.

**Scoring system.** It was decided to score all assessed pathologies on a semiquantitative scale of 0–3 (none/mild/moderate/severe), following the principles from the RAMRIS (rheumatoid arthritis magnetic resonance imaging) and PsAMRIS (psoriatic arthritis magnetic resonance imaging) systems<sup>9,10,11</sup>, and to create a total enthesal inflammation score by summation of scores of all inflammatory variables (intratendon hypersignal on T2w/STIR sequences, peritendon hypersignal, BME, and bursitis). Similarly, a total enthesal structural damage score by summation of structural scores (enthesophyte, bone erosion, tendon thickening) was developed. Intratendon hypersignal on T1W sequences was

Table 1. Magnetic resonance imaging definitions of key enthesal pathologies.

Pathology	Definition
1. Intratendon hypersignal (enthesal tendonitis)	Signal characteristics consistent with increased water content/inflammation* within the tendon/ ligament/aponeurosis close to its insertion
2. Peritendon hypersignal (enthesal peritendonitis)	Signal characteristics consistent with increased water content/inflammation* in the soft tissues surrounding the tendon/ ligament/aponeurosis, close to its insertion
3. Bone marrow edema (enthesal osteitis)	Bone lesion with ill-defined margins and signal characteristics consistent with increased water content/inflammation,* close to the tendon/aponeurosis insertion
4. Bursitis <sup>†</sup>	Signal characteristics consistent with increased water content/inflammation* in an above normal-sized bursa
5. Tendon/aponeurosis thickening	Abnormal thickening of the tendon/aponeurosis close to its insertion
6. Enthesophyte	Abnormal bone formation at the insertion of tendon/ ligament/aponeurosis insertion into the bone
7. Bone erosion (enthesal bone erosion)	A sharply marginated bone lesion, with typical signal characteristics** and a visible cortical break, located close to the tendon/ ligament/aponeurosis insertion
8. Intra-tendon hypersignal on T1W	Increased signal in T1-weighted (T1W) sequence within the tendon/ ligament/aponeurosis close to its insertion

<sup>†</sup>This lesion should only be assessed in enthesal regions in which a relevantly located bursa is present. \*High signal intensity on short-tau inversion recovery/T2wFS images and/or above normal post-gadolinium enhancement on T1W images. \*\*On T1W images without contrast injection: loss of normal low signal intensity of cortical bone and loss of normal high signal intensity of marrow fat. T2wFS: T2w fat-suppressed [images].

not included in sum scores. In exercises described in the present paper, scoring of entheses of the heel region was chosen, i.e., at calcaneal insertions of the Achilles tendon and plantar fascia, respectively.

**Exercise 1.** Exercises 1 and 2 included single-point images of the heel region, which were scored for the selected pre-defined pathologies. Exercise 1 was used for initial learning, calibration, and identification of pitfalls. Mean pairwise inter-reader single-measure ICC for inflammatory and structural variables, done without calibration, were 0.40 and 0.41, respectively.

**Exercise 2.** In exercise 2, agreement between reader pairs varied from poor to very good for various lesion types and their sum scores (Table 2). When limiting the analyses to 3 participating musculoskeletal radiologists and 3 rheumatologists with best ICC for inflammatory pathologies in exercise 2, reliability improved to moderate to very good. For this subset of readers, median single-measure ICC for total inflammation scores was 0.85, while for total structural damage scores was 0.68. Median  $\kappa$  for different inflammatory pathologies varied from 0.60 to 0.89, and for individual structural pathologies from 0.41 to 0.78. Average-measure ICC based on 2 readers among the preselected 6 readers (median 0.92 for total inflammatory score, 0.81 for total damage scores) were better than the single-measure ICC.

**Exercise 3.** This exercise included 2 timepoint images, in which inflammatory pathologies were scored. Mean pairwise interreader ICC and lesion-wise  $\kappa$  agreement demonstrated moderate to good reliability when all readers were considered (Table 3). The subset of readers (3 rheumatologists with best agreement for inflammatory measures in exercise 2 and the participating radiologist in exercise 3) demonstrated good to very good reliability, both for baseline scores and for change in scores (Table 2). The median baseline single-measures ICC

for total inflammation was 0.91, while it was 0.84 for change in score. Median average-measure ICC based on 2 readers [status: 0.95 (range 0.95–0.97), change: 0.92 (0.89–0.96)] were higher than single-measure ICC. Using 3 readers demonstrated numerically higher average-measure ICC [status: median 0.97 (0.97–0.97), change 0.94 (0.94–0.95)].

The Heel Enthesitis Scoring System (HEMRIS) showed moderate responsiveness, with SRM of 0.70 (95% CI 0.38–1.05) for all readers in exercise 3.

## DISCUSSION

To the best of our knowledge, our study is the first international consensus effort toward development of a comprehensive MRI scoring system, combined with MRI definitions and reader rules, for enthesitis in patients with SpA. The work was informed by an SLR<sup>8</sup>, which clarified knowledge gaps and the need for development of a validated MRI enthesitis scoring system to be used as outcome measure in clinical trials. Enthesitis, often located at the heels, is a typical feature of SpA and is easily accessible for MRI<sup>12</sup>. Further, enthesitis in SpA may show changes both in inflammation (such as BME and perienthesal inflammation) and damage (such as erosion and new bone formation)<sup>13,14</sup>. Thus, both inflammatory and structural MRI findings were considered relevant to include in the scoring system. A series of multireader scoring exercises focused on the heel region, using an intuitive Web-based data entry and image display platform. The preliminary OMERACT-HEMRIS showed good interreader agreement for status scores and for change over time in inflammatory measures. Considering that baseline heel enthesitis was not mandatory in exercise 3, the moderate SRM (0.70) supports that responsiveness of the HEMRIS score would likely be good in trials with baseline enthesitis as an inclusion criterion. Thus, HEMRIS appears promising for further validation and future use in randomized controlled trials.



**MRI Sequences:** STIR/T2wFS or, alternatively, T1W post-Gd; T1W without contrast (not mandatory if only inflammation is assessed)

**Imaging Planes:** Achilles tendon: Sagittal and preferably coronal; Plantar aponeurosis: Sagittal and preferably axial

**Area to Score:** At the heel region, the entheses are evaluated within 1 cm from the tendon/aponeurosis insertion.

**Scoring Procedure:**

Enteseal Soft Tissue Inflammation

- If T1W post-Gd images are available, enteseal soft tissues are assessed on these and the intra-tendon/peritendon/bursal hypersignal is defined as above-normal post-Gd enhancement on T1W images
- If only STIR/T2wFS images are available, enteseal soft tissues are assessed on these and the intra-tendon/peritendon/bursal hypersignal is defined as high signal intensity on STIR/T2wFS images
- Grading scale is 0–3 based on thirds of the maximum potential volume of enhancing soft tissue: Score 0 = normal; 1 = mild; 2 = moderate; and 3 = severe

Enteseal Osteitis

- If STIR/T2wFS images are available, enteseal osteitis is assessed on these defined as a lesion within the enteseal bone marrow with ill-defined margins and high signal intensity on STIR/T2wFS images (“bone marrow edema”)
- If only T1W post-Gd images are available, enteseal osteitis is assessed on these, and defined as a lesion within the enteseal bone marrow, with ill-defined margins, which shows above-normal enhancement on T1W post-Gd images (“bone marrow postcontrast enhancement”)
- Grading scale is 0–3 based on the proportion of bone with edema, compared to the “assessed bone volume,” judged on all available images: 0 = no edema; 1 = 1–33% of the bone is edematous; 2 = 34–66% of the bone is edematous; 3 = 67–100% of the bone is edematous.

Enteseal Structural Damage Variables

- Enteseal structural damage variables are scored based on T1W pre-Gd images

Figure 3. Reader rules for heel enthesitis. STIR: short-tau inversion recovery; T2wFS: T2W fat-suppressed; T1W: T1-weighted; Gd: gadolinium.

The strengths of this initiative include taking an SLR as starting point to clarify unmet need, the involvement of experienced MRI researchers in the development of consensus-based definitions and scoring systems, and the participation of multiple readers with both radiological and rheumatological backgrounds in interactive Web-based exercises with standardized image display and scoring module. Limitations include varying experience and backgrounds of readers in the exercises, which needs to be taken into consideration when interpreting the results. This was addressed by subanalysis of scores of a subset of experienced readers, who had showed high scoring proficiency in previous exercises. Longitudinal studies incorporating T1W images are needed for assessment of the sensitivity to change of structural variables. Future developments should also

include an MRI enthesitis reference image atlas, and image sets for training and calibration. The definitions and scoring principle may be applicable to other entheses. Thus, validation of the definitions and scoring system in other anatomical regions are also suggested.

The heel enthesitis MRI score appears to be particularly reliable if the mean score of 2 readers (compared to 1) is used in the final study analysis; the average-measure ICC for 2 readers were markedly higher (0.92–0.95 for inflammation total status/change score in last exercise) than single-measure ICC. This will be relevant in real-life clinical trials where 2 independent readers generally score images.

Increasing the number of novel therapeutic options in SpA and PsA increases the potential utility of an objective and reproducible enthesitis outcome measure. The proposed

**Table 2.** Exercise 2: Single-measure interreader intraclass correlation coefficient (ICC; sum scores), quadratic weighted  $\kappa$  (individual component scores, per lesion), and mean scores of all readers.

Variables	Reproducibility, smICC and $\kappa$				Range of Scores	Reader Scores	
	All Readers		Subset of Readers*			All Readers	Subset of Readers*
	Median	Mean (Range)	Median	Mean (Range)		**Mean (Range)	Mean (Range)
Inflammatory pathologies							
Interreader smICC (patient level)							
Total inflammation scores	0.58	0.56 (0.11–0.90)	0.85	0.83 (0.76–0.90)	0–21	3.94 (0.67–8.27)	4.83 (0.5–10.67)
Interreader quadratic weighted $\kappa$ (lesions level)							
Achilles tendon							
Peritendon hypersignal	0.41	0.45 (0.14–1.00)	0.66	0.64 (0.28–0.89)	0–3	0.45 (0–2.00)	0.64 (0–2.50)
Intratendon hypersignal	0.50	0.47 (0.04–0.91)	0.68	0.71 (0.53–0.90)	0–3	0.62 (0.07–2.07)	0.73 (0–2.17)
Retrocalcaneal bursitis	0.47	0.45 (–0.06 to 0.86)	0.60	0.62 (0.47– 0.71)	0–3	0.43 (0–1.67)	0.50 (0–2)
Bone marrow edema	0.83	0.78 (0.26–1.00)	0.89	0.90 (0.83–1.00)	0–3	0.44 (0–2.27)	0.52 (0–2.50)
Plantar fascia							
Periaponeurosis hypersignal	0.67	0.63 (0.12–0.91)	0.83	0.83 (0.74– 0.91)	0–3	0.82 (0–2.53)	1.02 (0–3.00)
Intraaponeurosis hypersignal	0.45	0.40 (0–0.92)	0.70	0.69 (0.54– 0.92)	0–3	0.51 (0–1.60)	0.69 (0–2.33)
Bone marrow edema	0.84	0.77 (0.11–0.98)	0.86	0.86 (0.73–0.94)	0–3	0.66 (0–2.47)	0.74 (0–2.67)
Structural pathologies							
Interreader smICC (patient level)							
Total structural damage score	0.27	0.35 (–0.04 to 0.85)	0.68	0.66 (0.37– 0.85)	0–18	1.54 (0.2–4.4)	2.33 (0.33–7.00)
Interreader quadratic weighted $\kappa$ (lesion level)							
Achilles tendon							
Tendon thickness	0.52	0.48 (0–0.92)	0.76	0.72 (0.4–0.92)	0–3	0.54 (0–2.27)	0.78 (0–3.00)
Bone erosion	0.54	0.45 (0–1.00)	0.78	0.78 (0.52–1.00)	0–3	0.14 (0–1.4)	0.19 (0–1.83)
Bone spur	0.00	0.26 (–0.08 to 1.0)	0.41	0.37 (0–0.87)	0–3	0.13 (0–0.87)	0.22 (0–1.33)
Intratendon hypersignal on T1W <sup>@</sup>	0.30	0.33 (–0.09 to 0.88)	0.64	0.63 (0.36–0.96)	0–3	0.46 (0.07–1.47)	0.58 (0–2.00)
Plantar fascia							
Tendon thickness	0.31	0.35 (–0.23 to 0.97)	0.86	0.72 (0.26–0.97)	0–3	0.50 (0–1.53)	0.75 (0–2.5)
Bone erosion	0.00	0.02 (–0.17 to 0.64)	0.00	0.03 (–0.05 to 0.14)	0–3	0.06 (0–0.27)	0.11 (0–0.5)
Bone spur	0.00	0.12 (–0.18 to 0.76)	0.42	0.4 (–0.18 to 0.76)	0–3	0.17 (0–0.53)	0.28 (0–1.17)
Intraaponeurosis hypersignal on T1W <sup>@</sup>	0.21	0.25 (–0.19 to 0.84)	0.40	0.42 (0.05– 0.83)	0–3	0.13 (0–1.27)	0.49 (0–2.00)

\* Three participating radiologists and 3 rheumatologists with best individual ICC with other readers for inflammatory pathologies in exercise 2. \*\*Each patient's score was calculated as the mean of all readers. The presented mean and ranges are means/ranges of these values (range of readers' mean scores). Readers: AJM, DG, FG, IH, IE<sup>†</sup>, KGH<sup>†</sup>, MS, MØ<sup>‡</sup>, PB, RGL<sup>†</sup>, SK<sup>‡</sup>, SJP, VF<sup>‡</sup>, WM (exercises 1 and 2); JJ<sup>†</sup> (only exercise 1); RPP (only exercise 2). <sup>†</sup>Musculoskeletal radiologists. <sup>‡</sup>Three rheumatologists with best individual ICC for inflammatory pathologies in exercise 2. <sup>@</sup>Not included in total structural damage score; it may occur both on inflammatory and structural backgrounds. smICC: single-measures intraclass correlation coefficient by 2-way random effects, absolute agreement for sum scores (patient level); T1W: T1-weighted.

OMERACT MRI heel enthesitis scoring system (HEMRIS) is a promising tool for further refinement and validation through the OMERACT filter and for future use in clinical trials<sup>15,16</sup>.

## ACKNOWLEDGMENT

We thank the Canadian Research and Education (CaRE) Arthritis organization ([www.carearthritis.com](http://www.carearthritis.com)) for help with setting up online meetings and exercises and development of the Web-based scoring interface.

## REFERENCES

- Schett G, Lories RJ, D'Agostino MA, Elewaut D, Kirkham B, Soriano ER, et al. Enthesitis: from pathophysiology to treatment. *Nat Rev Rheumatol* 2017;13:731–41.
- Polachek A, Li S, Chandran V, Gladman DD. Clinical enthesitis in a prospective longitudinal psoriatic arthritis cohort: incidence, prevalence, characteristics, and outcome. *Arthritis Care Res* 2017;69:1685–91.
- Mease PJ. Measures of psoriatic arthritis: tender and swollen joint assessment, Psoriasis Area and Severity Index (PASI), Nail Psoriasis Severity Index (NAPSI), Modified Nail Psoriasis Severity Index (mNAPSI), Mander/Newcastle Enthesitis Index (MEI), Leeds Enthesitis Index (LEI), Spondyloarthritis Research Consortium of Canada (SPARCC), Maastricht Ankylosing Spondylitis Enthesis Score (MASES), Leeds Dactylitis Index (LDI), Patient Global for Psoriatic Arthritis, Dermatology Life Quality Index (DLQI), Psoriatic Arthritis Quality of Life (PsAQOL), Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-F), Psoriatic Arthritis Response Criteria (PsARC), Psoriatic Arthritis Joint Activity Index (PsAJAI), Disease Activity in Psoriatic Arthritis (DAPSA), and Composite Psoriatic Disease Activity Index (CPDAI). *Arthritis Care Res* 2011;63 Suppl 11:S64–85.
- Eshed I, Bollow M, McGonagle DG, Tan AL, Althoff CE, Asbach P, et al. MRI of enthesitis of the appendicular skeleton in spondyloarthritis. *Ann Rheum Dis* 2007;66:1553–9.
- Mathew AJ, Coates LC, Danda D, Conaghan PG. Psoriatic arthritis: lessons from imaging studies and implications for therapy. *Exp Rev Clin Immunol* 2017;13:133–42.
- Benjamin M, Moriggl B, Brenner E, Emery P, McGonagle D, Redman S. The “enthesitis organ” concept: why enthesopathies may not present as focal insertional disorders. *Arthritis Rheum* 2004;50:3306–13.

Table 3. Exercise 3: Baseline and change single-measure interreader intraclass correlation coefficient (ICC; sum scores), quadratic weighted  $\kappa$  (individual component scores, per lesion), and mean scores of all readers.

Variables	Reproducibility, smICC and $\kappa$				Range of Scores	Reader Scores	
	All Readers		Subset of Readers*			All Readers	Subset of Readers*
	Median	Mean (Range)	Median	Mean (Range)		**Mean (Range)	Mean (Range)
Baseline scores							
Interreader smICC (total score)	0.83	0.81 (0.57–0.95)	0.91	0.91 (0.90–0.94)	0–21	3.55 (0.1–13.6)	4.04 (0.25–14.5)
Achilles tendon ( $\kappa$ — lesion)							
Peritendon hypersignal	0.64	0.62 (0.29–0.87)	0.79	0.78 (0.68–0.87)	0–3	0.47 (0–2.00)	0.41 (0–2.00)
Intratendon hypersignal	0.55	0.51 (0.12–0.89)	0.79	0.81 (0.77–0.89)	0–3	0.34 (0–1.50)	0.33 (0–1.50)
Retrocalcaneal bursitis	0.55	0.49 (–0.12 to 0.93)	0.57	0.60 (0.42–0.78)	0–3	0.34 (0–2.1)	0.40 (0–2.5)
Bone marrow edema	0.86	0.86 (0.75–0.97)	0.89	0.88 (0.84–0.92)	0–3	0.38 (0–3.00)	0.39 (0–3.00)
Plantar fascia ( $\kappa$ — lesion)							
Periaponeurosis hypersignal	0.65	0.60 (0.06–0.89)	0.80	0.80 (0.66–0.89)	0–3	0.85 (0–2.80)	1.01 (0–3.00)
Intraaponeurosis hypersignal	0.55	0.46 (–0.0 to 0.93)	0.84	0.84 (0.79–0.93)	0–3	0.56 (0–2.00)	0.80 (0–3.00)
Bone marrow edema	0.89	0.89 (0.76–0.98)	0.87	0.88 (0.81–0.96)	0–3	0.61 (0–2.90)	0.65 (0–3.00)
Change scores							
Interreader smICC (total score)	0.82	0.80 (0.57–0.92)	0.84	0.85 (0.79–0.85)	0–21	1.54 (0.1–4.9)	1.99 (0–6.25)
Achilles tendon ( $\kappa$ — lesion)							
Peritendon hypersignal	0.49	0.47 (0.21–0.75)	0.50	0.53 (0.42–0.75)	0–3	0.20 (0–0.8)	0.17 (0–1.00)
Intratendon hypersignal	0.41	0.41 (0.09–0.63)	0.54	0.52 (0.35–0.63)	0–3	0.15 (0–1.00)	0.17 (0–1.25)
Retrocalcaneal bursitis	0.24	0.26 (–0.2 to 1.00)	0.45	0.42 (0.15–0.67)	0–3	0.21 (0–0.8)	0.29 (0–1.00)
Bone marrow edema	0.52	0.53 (0.30–0.82)	0.47	0.54 (0.45–0.76)	0–3	0.14 (0–1.1)	0.20 (0–1.25)
Plantar fascia ( $\kappa$ — lesion)							
Periaponeurosis hypersignal	0.66	0.61 (0.17–0.87)	0.77	0.78 (0.70–0.85)	0–3	0.33 (0–1.30)	0.42 (0–1.75)
Intraaponeurosis hypersignal	0.53	0.42 (–0.08 to 0.80)	0.62	0.64 (0.57–0.77)	0–3	0.24 (0–0.90)	0.40 (0–1.25)
Bone marrow edema	0.78	0.77 (0.57–0.94)	0.81	0.79 (0.69–0.88)	0–3	0.28 (0–1.30)	0.58 (0–1.25)

\*One participating radiologist and 3 rheumatologists with best individual ICC with other readers for inflammatory pathologies in exercise 2. \*\*Each patient's score was calculated as the mean of all readers. The presented mean and ranges are means/ranges of these values (range of readers' mean scores). Readers: AJM, DG, FG, IE<sup>†</sup>, MS, MØ<sup>‡</sup>, PB, SK<sup>‡</sup>, SJP, VF<sup>‡</sup>. <sup>†</sup>Musculoskeletal radiologist. <sup>‡</sup>Three rheumatologists with best individual ICC for inflammatory pathologies in exercise 2. smICC: single-measures intraclass correlation coefficient by 2-way random effects, absolute agreement for sum scores (patient level).

7. Dougados M1, Combe B, Braun J, Landewé R, Sibilia J, Cantagrel A, et al. A randomised, multicentre, double-blind, placebo-controlled trial of etanercept in adults with refractory heel enthesitis in spondyloarthritis: the HEEL trial. *Ann Rheum Dis* 2010;69:1430–35.
8. Mathew AJ, Krabbe S, Kirubakaran R, Barr AJ, Conaghan PG, Bird P, et al. Utility of magnetic resonance imaging (MRI) in diagnosing and monitoring enthesitis in patients with spondyloarthritis: a systematic literature review [accepted]. *J Rheumatol* 2019;46:xx–xx.
9. Østergaard M, Peterfy C, Conaghan P, McQueen F, Bird P, Ejbjerg B, et al. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system. *J Rheumatol* 2003;30:1385–6.
10. Østergaard M, McQueen F, Wiell C, Bird P, Bøyesen P, Ejbjerg B, 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;36:1816–24.
11. Østergaard M, Peterfy CG, Bird P, Gandjbakhch F, Glinatsi D, Eshed I, et al. The OMERACT rheumatoid arthritis magnetic resonance imaging (MRI) scoring system: updated recommendations by the OMERACT MRI in arthritis working group. *J Rheumatol* 2017;44:1706–12.
12. Dougados M, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. *Arthritis Rheum* 1991;34:1218–27.
13. Feydy A, Lavie-Brion MC, Gossec L, Lavie F, Guerini H, Nguyen C, et al. Comparative study of MRI and power Doppler ultrasonography of the heel in patients with spondyloarthritis with and without heel pain and in controls. *Ann Rheum Dis* 2012; 71:498–503.
14. Kehl AS, Corr M, Weisman MH. Review: enthesitis: new insights into pathogenesis, diagnostic modalities, and treatment. *Arthritis Rheumatol* 2016;68:312–22.
15. Boers M, Kirwan JR, Tugwell P, Beaton D, Bingham CO III, Conaghan PG, et al. The OMERACT Handbook. [Internet. Accessed September 1, 2018.] Available from: <https://omeract.org/resources>
16. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *J Clin Epidemiol* 2014;67:745–53.