

Development and Reliability of a Preliminary Foot Osteoarthritis Magnetic Resonance Imaging Score

Jill Halstead, Carmen Martín-Hervás, Elizabeth M.A. Hensor, Dennis McGonagle, Anne-Maree Keenan, Anthony C. Redmond, and Philip G. Conaghan

ABSTRACT. Objective. Foot osteoarthritis (OA) is very common but underinvestigated musculoskeletal condition and there is little consensus as to common magnetic resonance imaging (MRI) features. The aim of this study was to develop a preliminary foot OA MRI score (FOAMRIS) and evaluate its reliability.

Methods. This preliminary semiquantitative score included the hindfoot, midfoot, and metatarsophalangeal joints. Joints were scored for joint space narrowing (JSN; 0–3), osteophytes (0–3), joint effusion/synovitis, and bone cysts (present/absent). Erosions and bone marrow lesions (BML) were scored (0–3) and BML were evaluated adjacent to entheses and at sub-tendon sites (present/absent). Additionally, tenosynovitis (0–3) and midfoot ligament pathology (present/absent) were scored. Reliability was evaluated in 15 people with foot pain and MRI-detected OA using 3.0T MRI multi-sequence protocols, and assessed using ICC as an overall score and per anatomical site.

Results. Intrareader agreement (ICC) was generally good to excellent across the foot in joint features (JSN 0.90, osteophytes 0.90, effusion/synovitis 0.46, cysts 0.87), bone features (BML 0.83, erosion 0.66, BML entheses 0.66, BML sub-tendon 0.60) and soft tissue features (tenosynovitis 0.83, ligaments 0.77). Interreader agreement was lower for joint features (JSN 0.43, osteophytes 0.27, effusion/synovitis 0.02, cysts 0.48), bone features (BML 0.68, erosion 0.00, BML entheses 0.34, BML sub-tendon 0.13), and soft tissue features (tenosynovitis 0.35, ligaments 0.33).

Conclusion. This preliminary FOAMRIS demonstrated good intrareader reliability and fair interreader reliability when assessing the total feature scores. Further development is required in cohorts with a range of pathologies and to assess the psychometric measurement properties. (First Release June 1 2017; *J Rheumatol* 2017;44:1257–64; doi:10.3899/jrheum.160617)

Key Indexing Terms:

OSTEOARTHRITIS

FOOT

RELIABILITY

MAGNETIC RESONANCE IMAGING

SEMIQUANTITATIVE SCORE

Osteoarthritis (OA) of the foot is a common cause of pain and disability^{1,2,3}. Radiographic studies suggest OA is much more common in the foot than previously suspected^{3,4,5}. The

prevalence was reported to be between 60.7% and 94.6% for the foot joints in those aged 62–94 years². Magnetic resonance imaging (MRI) has been used in describing and

From the Leeds Institute of Rheumatic and Musculoskeletal Medicine, and the School of Healthcare, University of Leeds; UK National Institute for Health Research (NIHR) Biomedical Research Centre, Leeds Teaching Hospitals Trust, Leeds; Arthritis Research UK Experimental Osteoarthritis Treatment Centre, Leeds; Arthritis Research UK Centre for Sports, Exercise and Osteoarthritis, Nottingham/Leeds; Salford Royal Hospital UK National Health Service (NHS) Foundation Trust, Manchester, UK; Department of Musculoskeletal Radiology, La Paz University Hospital, Autonomous University of Madrid; Biomedical Research Networking Centre on Bioengineering, Biomaterials and Nanomedicine, Madrid, Spain.

Authors A.C. Redmond, P.G. Conaghan, A.M. Keenan, and E.M. Hensor are funded in part by the NIHR Leeds Biomedical Research Centre. The work was directly supported by an Arthritis Research UK grant (no. 18256) and the Leeds Experimental Osteoarthritis Treatment Centre, supported by Arthritis Research UK grant (no. 20083) and the Arthritis Research UK Sports, Exercise and Osteoarthritis Centre grant (no. 20194).

J. Halstead, PhD, Visiting Research Fellow, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and Principal Podiatrist, Salford Royal Hospital NHS Foundation Trust; C. Martín-Hervás, PhD, MD, Consultant Radiologist, Department of Musculoskeletal Radiology, La Paz University Hospital, and Associate Professor of Radiology, School of Medicine, Autonomous University of Madrid, and Biomedical Research Networking Centre on Bioengineering, Biomaterials and Nanomedicine; E.M. Hensor, PhD, Biostatistician, Leeds

Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals Trust; D. McGonagle, PhD, Professor of Investigative Rheumatology, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals Trust; A.M. Keenan, PhD, Professor of Allied Health Research, School of Healthcare, University of Leeds, and NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals Trust, and Arthritis Research UK Experimental Osteoarthritis Treatment Centre; A.C. Redmond, PhD, Professor of Clinical Biomechanics, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals Trust, and Arthritis Research UK Experimental Osteoarthritis Treatment Centre, and Arthritis Research UK Centre for Sports, Exercise and Osteoarthritis; P.G. Conaghan, PhD, Professor of Musculoskeletal Medicine, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals Trust, and Arthritis Research UK Experimental Osteoarthritis Treatment Centre, and Arthritis Research UK Centre for Sports, Exercise and Osteoarthritis. P.G. Conaghan and A.C. Redmond contributed equally to this study.

Address correspondence to Professor A.C. Redmond, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, 2nd Floor, Chapel Allerton Hospital, Leeds, LS7 4SA, UK. E-mail: a.redmond@leeds.ac.uk

Accepted for publication April 11, 2017.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2017. All rights reserved.

defining knee OA pathology; however, its use in foot OA is limited, possibly because of the complexity of foot anatomy and image acquisition. Further, while semiquantitative scores have been developed for the knee, hip, and hand^{6,7,8,9,10,11,12,13}, none exist for OA of the foot. The aim of our study was to develop a foot OA MRI score (FOAMRIS) for assessing pathological features of OA and soft tissue features that may be commonly associated with foot pain.

MATERIALS AND METHODS

Development of the FOAMRIS. Following a review of MRI scoring systems^{8,13,14,15}, a consensus process was undertaken involving 2 musculoskeletal radiologists, 2 rheumatologists, and 3 podiatrists. A preliminary scoring system was developed to identify and grade typical pathological features of OA in the joints, bones, and soft tissue features associated with foot pain.

The new system included 16 joints: first to fifth metatarsophalangeal (MTP) joints and tarsometatarsal joints, navicular-medial-cuneiform, navicular-intermediate-cuneiform, navicular-lateral-cuneiform, talonavicular, calcaneal-cuboid, and subtalar. Twelve bones were included: first to fifth metatarsals (divided into the distal, central and proximal regions), lateral cuneiform, intermediate cuneiform, medial cuneiform, navicular, cuboid, calcaneus, and talus. The interphalangeal joints and toes were not included in this assessment score because these are often not in the field of view in a foot and ankle MRI coil.

Tendons and ligaments of the foot were included, in 8 sites of tenosynovitis: tibialis anterior, extensor hallucis longus, extensor digitorum longus, peroneus brevis, peroneus longus, tibialis posterior, flexor hallucis longus, and flexor digitorum longus. The Lisfranc ligament complex and intertarsal ligaments were included, although not every ligament in the Lisfranc (midfoot) region was individually scored because of the large degree of anatomical variation¹⁶. These sites were included because of the association of soft tissue disorders in OA¹⁷, which has been shown for Lisfranc injuries and tendon damage^{18,19}.

Five sub-tendon sites of the foot (bone regions adjacent to overlying tendons) were also included: lateral calcaneus under long peroneal tendon, lateral cuboid under long peroneal tendon, medial calcaneus under posterior tibial tendon, medial navicular under posterior tibial tendon, and medial cuneiform under anterior tibial tendon. These sub-tendon sites, where tendons wrap around the bones, have been described as “functional entheses” and are sites associated with pain in mechanical foot disorders^{20,21,22,23}. On MRI, these regions can be associated with abnormal signal in the tendon and at the adjacent bone of the ankle²³, and it is unclear whether this may be the case in the foot.

Enthesopathy has been shown to be somewhat associated with OA in the hands^{13,24,25,26}. It is as yet unclear whether there may be an association in the foot, given the weight-bearing design of the structures; therefore enthesopathy was scored at 9 sites in the foot: the tibialis anterior tendon at the plantar distal medial base of the first metatarsal bone and plantar distal medial cuneiform bone; peroneus longus tendon at the plantar base of the first metatarsal bone and plantar distal base medial cuneiform bone; tibialis posterior tendon at the plantar insertion at the base of the second, third, or fourth metatarsal bones, the plantar proximal medial cuneiform bone, the plantar medial of lateral cuneiform bone, and plantar medial navicular bone; and finally the peroneus brevis tendon at the dorsal lateral base of the fifth metatarsal.

A set of MRI features was determined, and semiquantitative scores for each feature were then developed. The term “bone marrow lesion” (BML) was adopted in this system, rather than bone marrow edema, because bone signal in OA may not be attributed solely to fluid²⁷. During the consensus process, it became apparent that the development of a cartilage score posed

challenges because of the small cross-sectional surfaces and complexity of the anatomy. Therefore, a pragmatic approach was taken and a joint space narrowing (JSN) definition was agreed. To provide a score that could be applied in the absence of contrast agent, we did not include multiple severity categories for scoring or differentiate between synovitis and effusion [previously adopted in rheumatoid arthritis (RA) of the foot and OA of the hand^{12,28}], but pragmatically scored for the presence or absence of joint effusion/synovitis. The final definitions of each MRI feature, anatomical locations, and semiquantitative scores are summarized in Table 1.

Image acquisition. Fifteen participants were recruited as part of a larger study. In accordance with the Declaration of Helsinki, ethical approval was provided (Leeds West Ethics Committee 09/H1305/10). Participants were included if they reported foot pain on weight-bearing and the musculoskeletal radiologist judged there to be MRI features of OA, which were based on knee MRI and foot radiographic criteria in at least 1 foot joint^{4,29}. Inclusion was based, therefore, on the presence of osteophytes judged to be at least moderate in size (\geq grade 2) or, where the osteophytes were graded “small,” this was accompanied by JSN (partial to full thickness, grade \geq 2) and subchondral BML with cysts.

Participants were scanned using a Siemens Magnetom Verio (3T) large-bore MRI scanner (Siemens Medical Solutions). All scans were acquired using an 8-channel foot and ankle coil, with the foot placed perpendicular to the ankle and magnetic field (β_0) and centered over the navicular bone. The following protocol was used: T2-weighted fat-saturated sequence variables were TR: 3000–3600 ms, TE: 69, flip angle: 155–160°, echo train length 8, 2-mm slices, and 0.4-mm inter-slice gap, matrix 256 × 256, and field of view (FOV) 150 × 150 mm in 3 planes. Short-tau inversion recovery sequence (STIR) variables were TR: 4500 ms, TE: 31, NEX 2, TI 200, flip angle 150°, echo train length 11, 3-mm slices and 0.6-mm inter-slice gap, matrix 320 × 256, and FOV 150 × 150 mm in 3 planes. T1-weighted high-resolution spin echo sequence variables were TR: 700 ms, TE: 10, FS 3, flip angle: 90°, 1.2-mm slices and 1.32-mm inter-slice gap, matrix 512 × 512, and FOV 150 × 150 mm in the sagittal plane. Gradient recalled echo sequence variables were TR: 450, TE: 2.5, flip angle 30°, echo train length 1, 3-mm slices, 0.6 mm interslice gap, Matrix 336 × 448, and FOV 250 × 250 mm in the sagittal plane.

FOAMRIS reliability. Anonymized scans were analyzed using OsiriX 64-bit Version 5.6 (OsiriX Foundation). All images were scored using the standardized score sheet (Supplementary Data 1, available with the online version of this article) and the FOAMRIS system (Table 1, Figure 1, Figure 2, and Figure 3). Intrareader reliability was undertaken by an experienced musculoskeletal radiologist who read the same images twice in a random order more than 1 week apart. An interreader reliability exercise was undertaken by a second reader. Both readers undertook a consensus exercise together using 5 separate foot images prior to second reader scoring.

Features were scored for each joint, bone, and soft tissue site, with all sites grouped. Reliability scores were evaluated using descriptive statistics; percentage of exact agreement (PEA) and Chamberlain percent positive agreement (PPA), which is the proportion of the total number of ratings made in a given category during the 2 readings (either intra- or interreader pairs) that were in agreement. Additionally, ICC were calculated using generalizability theory; the Brennan method was used to account for negative variance components³⁰. The individual joint or bone was considered the facet of differentiation. Joint or bones were considered to be nested within patients. Patient, occasion (for intrareader reliability), and reader (for interreader reliability) were considered random facets of generalization. Occasionally a negative ICC was obtained; when this occurred, we reported that the result was negative (indicating poor agreement), but did not report the actual value. ICC could not be calculated when all joints or bones scored 0.

The reliability results were evaluated according to the Cicchetti criteria as < 0.40 poor, 0.40–0.59 fair, 0.60–0.74 good, and 0.75–1.00 excellent³¹. Analysis was undertaken using Stata 13.1 (StataCorp) and G_STRING IV (a wrapper for urGENOVA, University of Iowa).

Table 1. Definitions of each MRI feature and the related semiquantitative scores.

MRI Feature and Anatomical Location	Definition	Score
JSN: All joints of the hindfoot, tarsus, midfoot, and metatarsophalangeal joints	Increased signal in T2-weighted sequences (fat-suppressed or inversion recovery sequences) and/or loss of joint space as a partial or complete loss on T1-weighted images and/or gradient echo sequence. Visible in 2 planes.	JSN was scored as 0–3: 0 = normal thickness and signal, 1 = increased signal, 2 = partial thickness focal loss, 3 = full thickness loss of joint space ($\geq 75\%$ of the region).
Osteophytes: All joints of the hindfoot, tarsus, midfoot, and metatarsophalangeal joints Effusion/synovitis: All joints of the hindfoot, tarsus, midfoot, and metatarsophalangeal joints	Abnormal bone formation in the periarticular region on T1-weighted images. The presence of increased intraarticular fluid, demonstrated as high signal intensity on T2-weighted sequences (fat-suppressed or inversion recovery sequences). Visible in 2 planes: coronal and sagittal.	Osteophytes were scored as 0–3: 0 = none, 1 = mild, 2 = moderate, 3 = large. Effusion/synovitis was scored as 0–1: 0 = absent, 1 = present
Subchondral cyst: All joints of the hindfoot, tarsus, midfoot, and metatarsophalangeal joints	A sharply marginated subchondral bone lesion that showed increased signal intensity on T2-weighted images (fat-suppressed or inversion recovery sequences). Visible in 2 planes without a cortical break.	Cysts were scored as 0–1: 0 = absent, 1 = present.
BML: All bones of the hindfoot, tarsus, midfoot, and metatarsals	An area of poorly delineated signal within the trabecular bone that shows decreased signal intensity on T1-weighted images and increased signal intensity on T2-weighted images (fat-suppressed or inversion recovery sequences). Visible in at least in 2 planes.	BML was scored as 0–3 according to the proportion of bone with abnormal signal: 0 = none, 1 = 1–33%, 2 = 34–66%, 3 = 67–100%, except in the long bones of the metatarsals, where the BML was scored in 3 regions per bone: (1) At the proximal joint, the base (up to the epiphysis and metaphysis) was included in the tarsometatarsal joint and scored 0–3. (2) At the central region, metatarsal shaft (diaphysis) was divided into proximal, central, and distal in one-third increments (33%) on a bone level and scored 0–3. (3) At the distal region, the head (to the epiphysis and metaphysis) was included in the metatarsophalangeal joint and scored 0–3.
Bone erosion: All bones of the hindfoot, tarsus, midfoot, and metatarsals	A bone defect in the cortical and juxtacortical region, with sharp margins visible on T1-weighted images and with a loss of normal low signal intensity of cortical bone and loss of normal high signal intensity of marrow fat. Visible in 2 planes with a cortical break seen in at least 1 plane.	Erosions were scored from 0–3 according to the volume of the erosion as a proportion of the joint margin: 0 = no erosion; 1 = 1–33%, 2 = 34–66%, 3 = 67–100%
Enthesopathy: Locations at the insertion of the tendons: posterior tibial, anterior tibial, flexor digitorum, flexor hallucis, extensor digitorum, extensor hallucis, peroneus longus, and peroneus brevis. Location at sites at the attachments of the Lisfranc and intertarsal ligament complex	A BML pattern where altered signal intensity within the bone was adjacent to insertions of anatomically defined ligaments and/or tendons of the foot. Visible in at least 2 planes.	Enthesopathy scored as 0–1: 0 = absent, 1 = present.
Sub-tendon BML (functional enthesopathy): All bones of the hindfoot, tarsus, midfoot, and metatarsals adjacent to the course of a tendon at the medial (posterior tibial and anterior tibial tendons), lateral (peroneus longus and peroneus brevis), plantar (flexor digitorum and flexor hallucis longus tendons), and dorsal (extensor digitorum and extensor hallucis longus tendons) regions	A BML pattern where increased signal intensity within the bone was adjacent to the course of a tendon and away from an articular surface. Shown as hyperintensity on T2-weighted sequences and decreased signal intensity on T1-weighted images. Visible in at least 2 planes.	Sub-tendon BML was scored 0–1: 0 = absent, 1 = present.
Tenosynovitis tendon locations: Posterior tibial, anterior tibial, flexor digitorum, flexor hallucis, extensor digitorum, extensor hallucis, peroneus longus, and peroneus brevis	Decreased signal intensity on T1-weighted images and increased signal on T2-weighted (fat-suppressed or inversion recovery sequences) in a region of the tendon with an enclosing tendon sheath. Visible in at least 2 planes.	Tenosynovitis was scored 0–3: 0 = normal, 1 = < 2 mm peritendinous effusion, 2 = > 2 and < 5 mm peritendinous effusion and/or thickening and high intratendinous signal intensity on T2-weighted sequences, 3 = > 5 mm peritendinous effusion and/or higher thickening and high intratendinous signal intensity.
Ligament abnormality: ligament locations: Lisfranc and intertarsal ligament complex	Thickening and high signal intensity seen on T2-weighted images (fat-suppressed or inversion recovery sequences) with or without disruption. Visible in at least 2 planes: axial and coronal.	Ligament abnormality was scored 0–1: 0 = absent, 1 = present.

MRI: magnetic resonance imaging; JSN: joint space narrowing; BML: bone marrow lesion.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2017. All rights reserved.

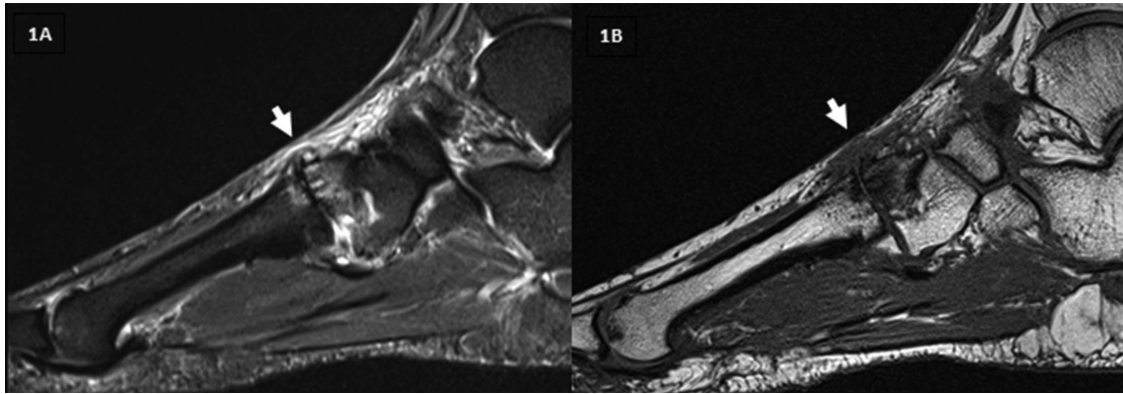


Figure 1. (A) and (B) show T1W and fat-saturated T2W sagittal images. Arrow shows the second tarsometatarsal joint, scored as a grade 2 (partial thickness or focal loss) for JSN, grade 2 (moderate) osteophyte, the presence of effusion/synovitis, the presence of cysts, grade 2 (34%–66%) BML at the intermediate cuneiform, and grade 1 (1%–33%) in the base of the second metatarsal. The second metatarsophalangeal joint was scored as a grade 2 for JSN, grade 2 osteophytes, in addition to grade 1 BML at the head of the second metatarsal. JSN: joint space narrowing; BML: bone marrow lesion.

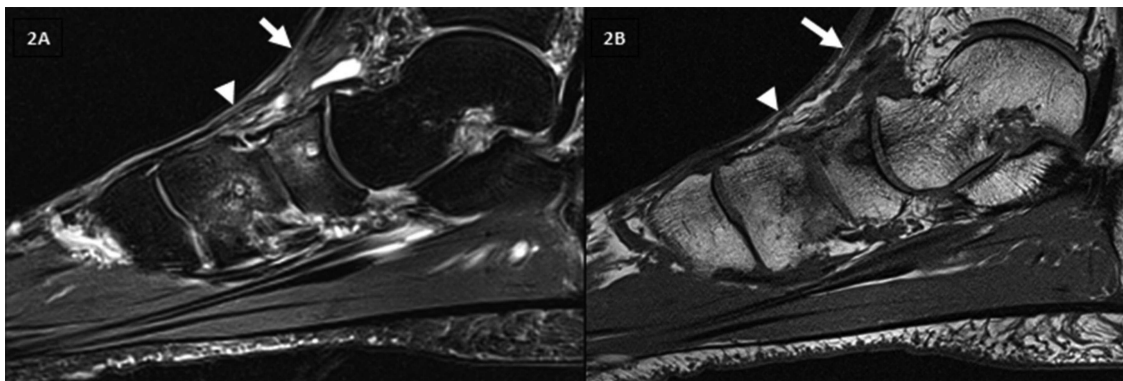


Figure 2. (A) and (B) show T1W and fat-saturated T2W sagittal images. White arrow shows the talonavicular joint, scored as a grade 0 (normal joint space and signal) JSN and grade 2 (moderate) osteophyte. The presence of effusion/synovitis, the presence of cysts, and navicular bone were scored as grade 2 (34%–66%) BML. White arrowhead shows the navicular-medial cuneiform, scored as grade 1 (increased signal in the joint space) JSN and grade 1 (mild) osteophytes. The presence of effusion/synovitis, the presence of cysts, and medial cuneiform bone were scored as grade 2 (34%–66%) BML. JSN: joint space narrowing; BML: bone marrow lesion.

RESULTS

The musculoskeletal radiologist read 61 sequential MRI, of which 35 were classified as having foot OA and deemed eligible for the study. Fifteen participants' scans were chosen at random for the reliability study. The participants were aged between 41 and 66 years [median 51 yrs, interquartile range (IQR) 46–60], included 10 women, and had a median body mass index (BMI) of 31.5 (IQR 26.3–34.5, range 23.5–40.1). OA was present in a single talonavicular joint in 5 participants, in 1–2 joints in the tarsi in 6 participants, and in 2 joints (MTP and tarsal joints) in 5 participants. An experienced radiologist performed the full FOAMRIS in 30 min per foot and reported the presence of the following conditions: JSN in 12 participants (total 31 sites), osteophytes in all participants (total 77 sites), effusion/synovitis in all participants (total 182 sites), cysts in 13 participants (total 28 sites), BML in all participants (total 74 sites), erosion in 5 participants

(total 10 sites), enthesopathy in 7 participants (total 9 sites), tenosynovitis in the entire group (total 47 sites), and ligament abnormalities in 6 participants (total 7 sites).

The intrareader reliability was summarized per imaging pathology (amalgamating anatomical locations; Table 2 and Table 3), and the range across the anatomical locations (Supplementary Tables 1–7, available with the online version of this article). It should be noted that ICC represent a ratio of between-object variability to total variability and can therefore be low if there is little variation in scores between different joints/bones, which was an issue when assessing agreement in specific sites.

Combining all joints, the results showed excellent agreement for the presence of JSN (ICC total = 0.90, range across joints = 0.65–1) and osteophytes (ICC total = 0.90, range across joints = 0.00–1), although there was a low proportion of severe scores in this sample and for some

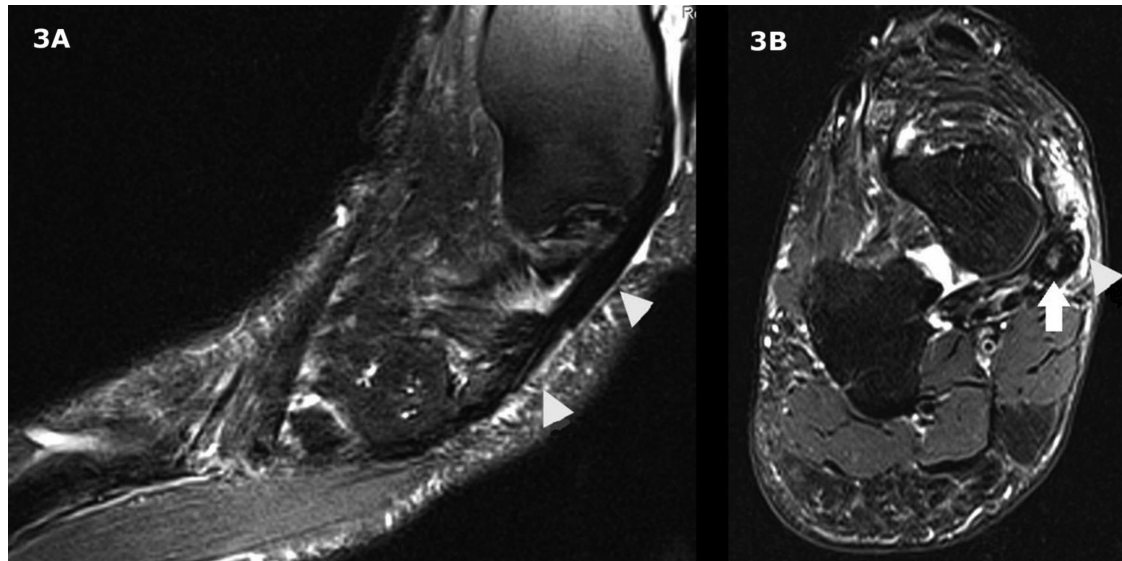


Figure 3. (A) and (B) show fat-saturated T2W images in the sagittal and axial plane, respectively. Tenosynovitis of tibialis posterior defined as high signal on a T2-weighted sequence in a region of the tendon (white arrow), and the enclosing tendon sheath (white arrowhead) in 2 planes was shown. This was scored as grade 2 (> 2 mm and < 5 mm peritendinous effusion and/or thickening and high intratendinous signal intensity).

Table 2. Intrarater repeatability results of scoring all joints and bones for JSN, osteophytes, effusion/synovitis, cysts, bone erosions, and marrow lesions. Values are % (n/N) unless otherwise stated.

Statistic	JSN 0–3	Osteophytes 0–3	Effusion/synovitis 0–1	Cysts 0–1	Erosion 0–3	BML 0–3
PEA	96 (238/248)	93 (231/248)	81 (202/248)	98 (242/248)	97 (242/250)	91 (225/248)
PEA range*	80–100	80–100	50–100	71–100	86–100	73–100
% ± 1 category	100 (248/248)	100 (248/248)	n/a	n/a	100 (250)	> 99 (247/248)
PPA score = 0	98 (422/429)	96 (322/335)	62 (74/120)	99 (432/438)	98 (470/478)	95 (338/354)
PPA score = 1	85 (44/52)	86 (100/117)	88 (330/376)	90 (52/58)	60 (12/20)	79 (88/111)
PPA score = 2	77 (10/13)	91 (40/44)	n/a	n/a	100 (2/2)	79 (22/28)
PPA score = 3	0 (0/2)	—	n/a	n/a	—	67 (2/3)
ICC	0.90	0.90	0.46	0.87	0.66	0.83
ICC range**	0.65–1	0.00–1	0.00–1	0.00–1	0.00–1	0.49–1

* PEA per location (Supplementary Tables, available with the online version of this article). ** ICC per location (Supplementary Tables). JSN: joint space narrowing; BML: bone marrow lesion; PEA: percent exact agreement; PPA: percent positive agreement; n/a: not applicable.

Table 3. Intrarater repeatability results for scoring sites of enthesopathy, sub-tendon BML, tenosynovitis, and ligament abnormalities. Values are % (n/N) unless otherwise stated.

Statistic	Enthesopathy 0–1	Sub-tendon BML 0–1	Tenosynovitis 0–3	Ligament 0–1
PEA	96 (130/135)	91 (68/75)	88 (106/120)	90 (27/30)
PEA range*	80–100	73–100	73–100	87–93
PEA ± 1 category	n/a	n/a	100 (120/120)	n/a
PPA score = 0	98 (248/253)	94 (120/127)	92 (134/146)	100 (30/30)
PPA score = 1	71 (12/17)	70 (16/23)	83 (68/82)	93 (42/45)
PPA score = 2	n/a	n/a	83 (10/12)	n/a
PPA score = 3	n/a	n/a	—	n/a
ICC	0.66	0.60	0.83	0.77
ICC range**	0.44–1	0.00–1	0.43–1	0.65–0.74

* PEA per location (Supplementary Tables, available with the online version of this article). ** ICC per location (Supplementary Tables). BML: bone marrow lesion; PEA: percent exact agreement; PPA: percent positive agreement; n/a: not applicable.

individual sites, the ICC was poor. There were very few JSN grade 3 scores and no scores for osteophytes grade 3 (the majority were grades 1–2), therefore the reliability in this category remains to be determined; however, for grades 0–2 the category-specific agreement was generally substantial (range 60%–100%). The presence of effusion/synovitis was the least reliably scored (ICC total 0.46, range across joints = negative to 1). Lower reliability in scoring effusion/synovitis was due to poor agreement over the absence of effusion/synovitis at the MTP joints. The repeatability for the scoring of presence of cysts was excellent when all joints were combined, although ICC was low for some individual sites (ICC total = 0.87, range across joints = 0.00–1).

The intrareader reliability for combined sites was excellent for BML (ICC total = 0.83, range across bones = 0.49–1) and erosions (ICC total = 0.66, range across bones = 0.00–1). As was observed for the joints, in the bony features there was a relatively low prevalence of more severe scores. Scores for severity of BML suggest similar repeatability for the range of scores 1 to 3, although only 3 bones across the sample scored a grade 3. While the agreement results for erosions were not equal across the severity scale, the results showed a lower level of agreement for a score of 1; however, only 20 erosion scores were assigned grade 1, 2 erosion scores assigned grade 2, and none were assigned grade 3. The reliability in this category still remains to be determined.

The intrareader reliability of bone-related and soft tissue result, and the patterns of BML associated with tendon enthesopathy (ICC total = 0.66, range across the locations 0.44–1) and at the sub-tendon BML regions (ICC total = 0.60, range across the locations 0.00–1) were similar, with excellent agreement scores when all sites were combined. Reliability of scores for tenosynovitis was also excellent (ICC total = 0.83, range = 0.43–1). The repeatability of scoring tenosynovitis was stable across scores ranging from 0 to 2. Score category 3 was not assigned during either of the repeated reads in our study; therefore, the repeatability in this category remains to be determined. The agreement scores for all ligament abnormality were excellent (ICC total = 0.77, range across the 2 sites = 0.65–0.74), with greater scores for the Lisfranc ligament.

The interreader reliability scores are summarized in Supplementary Tables 7–12 (available with the online version of this article), and as might be expected, the intrareader scoring showed greater reliability than interreader. The results demonstrated good agreement for the presence of JSN (ICC total = 0.43, range across joints = negative to 1) and poor agreement for osteophytes (ICC total = 0.27, range across joints = 0.00–1). The interreader reliability scores for the presence of effusion/synovitis were poor across the joints of the foot (ICC total = 0.02, range across joints = negative to 0.13). The repeatability for the scoring of presence of cysts was fair (ICC total = 0.48, range across joints = negative to 1).

The interreader reliability was excellent for sites of BML

(ICC total = 0.68, range across bones = 0.00–1), but was poor for erosion scores in bones with erosions present (ICC total = 0.00, values for all bones 0.00 where calculable). There were several sites for which both scorers agreed on the absence of any erosions, but ICC could not be calculated if there were no scores above 0. The interreader reliability of bone-related and soft tissue scores of BML associated with tendon enthesopathy was poor (ICC total = 0.34, range across the locations 0.00–1), but scores were less reliable at the sub-tendon BML regions (ICC total = 0.13, range across the locations negative to 0.65). Interreader reliability scores for tenosynovitis were poor (ICC total = 0.35, range = 0.00–0.61). The interreader reliability scores for all ligament abnormality were also poor (ICC total = 0.33, range across the 2 sites = 0.00–0.18), with higher scores for the Lisfranc ligaments.

DISCUSSION

To our knowledge, there are no MRI scoring systems for OA foot pathology, although a previous study has defined some MRI features in foot OA³². This new scoring system was deliberately inclusive of not only “traditional” OA features, but also included features that may inform studies investigating the broader construct of foot pain.

In our study, intrareader reliability of the total MRI features was shown to be generally excellent when assessed at a whole foot level, while interreader reliability was more variable. The best intra- and interreader reliability was seen for joint-specific features (JSN, osteophytes, and cysts), and compared well to scores such as those evaluating small joints in hand OA¹². The presence of joint effusion/synovitis showed worse intra- and interreader reliability and was lower because of poor agreement, particularly at the MTP joints, which may or not be considered a normal finding. The reliability scores may have been affected by the size of the joint because joint effusion/synovitis scores have been shown previously to be more variable in small joints of the hands¹². In a later reliability study of joint effusion/synovitis in the hand, the agreement improved once an atlas was developed¹³. In addition, administration of a contrast agent may have aided precision in estimating the volume of joint fluid, particularly in differentiating fluid from synovial hypertrophy. Further studies with contrast administration may be needed to refine the scoring and better characterize OA-related pathology.

Bony features demonstrated excellent intrareader agreement across the foot as a whole. Descriptively, the erosion scores were highly reliable across nearly all sites; however, this may have been influenced by the low number of lesions present. ICC values (where calculable) were variable, which may reflect both limitations in agreement over the presence of erosion and limitations in the amount of “true” variation between bones. The BML scores were also variable, and lower agreement was shown in the cuboid and the proximal metatarsals. Where patterns of BML were associated with the

tendon enthesis, intra- and interreader reliability was good, but at the sub-tendon region, reliability was lower. No reliability studies of these MRI features have been previously reported, and there is likely to be difficulty in scoring these regions where planar anatomy is subject to partial volume artifact; in these regions, an atlas would be beneficial.

The intra- and interreader reliability of scoring of soft tissue features was similar across ligament abnormalities and tenosynovitis. Similar levels of agreement have been reported for scores of hand tenosynovitis in RA³³ and hand OA¹². A limited number of ligaments of the midfoot were included in this score, which have been well described^{16,34}. Other foot ligaments were not included because of potential issues with poor visualization and requirement for specialist views and sequences, e.g., calcaneocuboid and calcaneonavicular ligaments³⁵.

The results of our study should be considered in light of the following limitations. The sample for our preliminary study included a group with relatively mild structural OA, and more severe damage was limited to few joint regions. In addition, the definition of OA on MRI as applied in our study, while based on consensus approaches developed for other joints, requires further work and validation and this has implications for the results presented. Definitions of the individual features are difficult because of the variation in presentation in the various anatomical sites and the technical aspects of acquiring MRI. For example, we did not use contrast-enhanced imaging in our study and so have not differentiated between synovitis and effusion. A detailed definition of osteophyte grading was not provided in this score, given the widely varying presentation of periarticular bone change in sites such as the first MTP joint versus the small joints of the hindfoot or midfoot. Future work is required to refine the FOAMRIS approach and analyze validity in larger and more diverse samples. In addition, 8 participants were obese (≥ 30 BMI), which may influence the frequency of the tendon and ligament pathology because greater occurrence has been shown in obese people at the ankle³⁶.

The foot poses unique challenges when using MRI because of the complexity of the anatomy and inherent variability in the shape and size. This manifests as problems with coil positioning, homogeneous fat saturation, imaging wrap, and magic-angle effect³⁷. In our study, a foot and ankle coil was used, which was beneficial for maintaining a consistent position within the magnet; however, this can be limited with larger and longer feet. Using a larger coil may allow for imaging of the entire foot, although the positioning might be difficult because of flexibility and foot type. In future studies, it may be appropriate to reposition the target for the hindfoot, midfoot, and forefoot, although this will increase acquisition times and may not be desirable.

The issue of how many planes and sequences to acquire is a complex one. In our study, both T2-weighted water-sensitive and STIR sequences in 3 planes were included to account

for possible failure of the fat saturation. T1-weighted sequences included high-resolution spin echo and gradient recalled echo in a single plane, which may have affected the scoring of erosions and osteophytosis. In practice, where acquisition time is of primary importance, a T2 fat-saturated sequence may suffice.

A minimum of 2 planes for each T1-weighted sequence could improve scoring; however, defining the optimum plane for each foot joint requires further work and a 3-D sequence may provide a compromise. Gradient recalled echo sequences are sensitive in delineating subchondral cysts and were helpful in the verification in our study. These sequences, however, are insensitive to diffuse marrow abnormalities because of trabecular magnetic susceptibility and will not show the full extent of these lesions, so in our study, spin echo sequences were also used for better BML detection⁸. Further consensus regarding sequence choice is recommended.

Across most scores, interreader reliability scores were lower than intrareader. We have identified training (the second reader was less experienced) and case definitions as likely contributors to these findings. Improved description of certain scoring features, accompanied by an atlas, would be a natural next step because this process has improved interreader reliability in other MRI scores¹³.

Finally, it is recognized that ICC can be affected by the degree of “true” variability in the sample, which in this relatively mild group was limited for some features. Further validation in more diverse samples should give a more accurate assessment of inter- and intrareader reliability.

We have proposed a set of definitions and scoring criteria for a semiquantitative MRI investigation of multiple foot pathology: FOAMRIS. This preliminary scoring system generally showed acceptable reliability for a broad range of pathologies except for effusion/synovitis, and for some features at anatomical sites where visualization may be particularly influenced by acquisition plane. Iterative development is now needed, and will include application in other cohorts, expert consensus on acquisition protocol, use of contrast, and the development of an atlas to aid scoring.

ACKNOWLEDGMENT

We acknowledge the contribution of Dr. Eiji Fukuba, musculoskeletal radiologist, who was involved in the consensus exercise in the development of the magnetic resonance imaging (MRI) scores. We also acknowledge the expertise of Dr. Richard Hodgson, Rob Evans, and Dr. Carole Burnett of the NIHR Leeds Musculoskeletal Biomedical Research Unit in the acquisition of the MRI scans used in the project.

ONLINE SUPPLEMENT

Supplementary material accompanies the online version of this article.

REFERENCES

1. Roddy E, Zhang W, Doherty M. Prevalence and associations of hallux valgus in a primary care population. *Arthritis Rheum* 2008;59:857-62.
2. Menz HB, Munteanu SE, Landorf KB, Zammit GV, Cicuttini FM.

- Radiographic evaluation of foot osteoarthritis: sensitivity of radiographic variables and relationship to symptoms. *Osteoarthritis Cartilage* 2009;17:298-303.
3. Roddy E, Thomas MJ, Marshall M, Rathod T, Myers H, Menz HB, et al. The population prevalence of symptomatic radiographic foot osteoarthritis in community-dwelling older adults: cross-sectional findings from the Clinical Assessment Study of the Foot. *Ann Rheum Dis* 2015;74:156-63.
 4. Menz HB, Munteanu SE, Landorf KB, Zammit GV, Cicuttini FM. Radiographic classification of osteoarthritis in commonly affected joints of the foot. *Osteoarthritis Cartilage* 2007;15:1333-8.
 5. van Saase JL, van Romunde LK, Cats A, Vandenbroucke JP, Valkenburg HA. Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations. *Ann Rheum Dis* 1989;48:271-80.
 6. Peterfy CG, Guermazi A, Zaim S, Tirman PF, Miaux Y, White D, et al. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. *Osteoarthritis Cartilage* 2004;12:177-90.
 7. Hunter DJ, Lo GH, Gale D, Grainger AJ, Guermazi A, Conaghan PG. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score). *Ann Rheum Dis* 2008;67:206-11.
 8. Hunter DJ, Guermazi A, Lo GH, Grainger AJ, Conaghan PG, Boudreau RM, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). *Osteoarthritis Cartilage* 2011;19:990-1002.
 9. Kornaat PR, Ceulemans RY, Kroon HM, Riyazi N, Kloppenburg M, Carter WO, et al. MRI assessment of knee osteoarthritis: Knee Osteoarthritis Scoring System (KOSS)—inter-observer and intra-observer reproducibility of a compartment-based scoring system. *Skeletal Radiol* 2005;34:95-102.
 10. Roemer FW, Hunter DJ, Winterstein A, Li L, Kim YJ, Cibere J, et al. Hip Osteoarthritis MRI Scoring System (HOAMS): reliability and associations with radiographic and clinical findings. *Osteoarthritis Cartilage* 2011;19:946-62.
 11. Lee S, Nardo L, Kumar D, Wyatt CR, Souza RB, Lynch J, et al. Scoring hip osteoarthritis with MRI (SHOMRI): a whole joint osteoarthritis evaluation system. *J Magn Reson Imaging* 2015;41:1549-57.
 12. Haugen IK, Lillegraven S, Slatkowsky-Christensen B, Haavardsholm EA, Sesseng S, Kvien TK, et al. Hand osteoarthritis and MRI: development and first validation step of the proposed Oslo Hand Osteoarthritis MRI score. *Ann Rheum Dis* 2011;70:1033-8.
 13. Haugen IK, Østergaard M, Eshed I, McQueen FM, Bird P, Gandjbakhch F, et al. Iterative development and reliability of the OMERACT hand osteoarthritis MRI scoring system. *J Rheumatol* 2014;41:386-91.
 14. 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.
 15. Østergaard M, Edmonds J, McQueen F, Peterfy C, Lassere M, Ejbjerg B, et al. An introduction to the EULAR-OMERACT rheumatoid arthritis MRI reference image atlas. *Ann Rheum Dis* 2005;64 Suppl:i3-i7.
 16. Castro M, Melão L, Canella C, Weber M, Negrão P, Trudell D, et al. Lisfranc joint ligamentous complex: MRI with anatomic correlation in cadavers. *AJR Am J Roentgenol* 2010;195:W447-55.
 17. Øiestad BE, Engebretsen L, Storheim K, Risberg MA. Knee osteoarthritis after anterior cruciate ligament injury: a systematic review. *Am J Sports Med* 2009;37:1434-43.
 18. Hardcastle PH, Reschauer R, Kutscha-Lissberg E, Schoffmann W. Injuries to the tarsometatarsal joint. Incidence, classification and treatment. *J Bone Joint Surg Bri* 1982;64:349-56.
 19. Bluman EM, Title CI, Myerson MS. Posterior tibial tendon rupture: a refined classification system. *Foot Ankle Clin* 2007;12:233-409, v.
 20. Benjamin M, McGonagle D. The anatomical basis for disease localisation in seronegative spondyloarthropathy at entheses and related sites. *J Anat* 2001;199:503-26.
 21. O'Donnell P, Saifuddin A. Cuboid oedema due to peroneus longus tendinopathy: a report of four cases. *Skeletal Radiol* 2005;34:381-8.
 22. Lo LD, Schweitzer ME, Fan JK, Wapner KL, Hecht PJ. MR imaging findings of entrapment of the flexor hallucis longus tendon. *AJR Am J Roentgenol* 2001;176:1145-8.
 23. Morrison WB, Carrino JA, Schweitzer ME, Sanders TG, Raiken DP, Johnson CE. Subtendinous bone marrow edema patterns on MR images of the ankle: association with symptoms and tendinopathy. *AJR Am J Roentgenol* 2001;176:1149-54.
 24. McGonagle D, Tan AL, Carey J, Benjamin M. The anatomical basis for a novel classification of osteoarthritis and allied disorders. *J Anat* 2010;216:279-91.
 25. Tan AL, Grainger AJ, Tanner SF, Shelley DM, Pease C, Emery P, et al. High-resolution magnetic resonance imaging for the assessment of hand osteoarthritis. *Arthritis Rheum* 2005;52:2355-65.
 26. Kortekaas MC, Kwok WY, Reijnen M, Wolterbeek R, Bøyesen P, van der Heijde D, et al. Magnetic resonance imaging in hand osteoarthritis: intraobserver reliability and criterion validity for clinical and structural characteristics. *J Rheumatol* 2015;42:1224-30.
 27. Zanetti M, Bruder E, Romero J, Hodler J. Bone marrow edema pattern in osteoarthritic knees: correlation between MR imaging and histologic findings. *Radiology* 2000;215:835-40.
 28. Baan H, Bezooijen R, Avenarius JK, Dubbeldam R, Drossaers-Bakker WK, van de Laar MA. Magnetic resonance imaging of the rheumatic foot according to the RAMRIS system is reliable. *J Rheumatol* 2011;38:1003-8.
 29. Hunter DJ, Arden N, Conaghan PG, Eckstein F, Gold G, Grainger A, et al; OARSI OA Imaging Working Group. Definition of osteoarthritis on MRI: results of a Delphi exercise. *Osteoarthritis Cartilage* 2011;19:963-9.
 30. Brennan RL. Generalizability theory. New York: Springer-Verlag; 2001.
 31. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instrument in psychology. *Psychol Assess* 1994;6:284-90.
 32. Halstead J, Bergin D, Keenan AM, Madden J, McGonagle D. Ligament and bone pathologic abnormalities more frequent in neuropathic joint disease in comparison with degenerative arthritis of the foot and ankle: implications for understanding rapidly progressive joint degeneration. *Arthritis Rheum* 2010;62:2353-8.
 33. Haavardsholm EA, Østergaard M, Ejbjerg BJ, Kvan NP, Kvien TK. Introduction of a novel magnetic resonance imaging tenosynovitis score for rheumatoid arthritis: reliability in a multireader longitudinal study. *Ann Rheum Dis* 2007;66:1216-20.
 34. Raikin SM, Elias I, Dheer S, Besser MP, Morrison WB, Zoga AC. Prediction of midfoot instability in the subtle Lisfranc injury. Comparison of magnetic resonance imaging with intraoperative findings. *J Bone Joint Surg AM* 2009;91:892-9.
 35. Melão L, Canella C, Weber M, Negrão P, Trudell D, Resnick D. Ligaments of the transverse tarsal joint complex: MRI-anatomic correlation in cadavers. *AJR Am J of Roentgenol* 2009;193:662-71.
 36. Galli MM, Protzman NM, Mandelker EM, Malhotra A, Schwartz E, Brigid SA. Comparing tendinous and ligamentous ankle pathology in atraumatic overweight and nonoverweight patients: a comprehensive MRI review. *Foot Ankle Spec*;7:449-56.
 37. Rosenberg ZS, Bencardino J, Mellado JM. Normal variants and pitfalls in magnetic resonance imaging of the ankle and foot. *Top Magn Reson Imaging* 1998;9:262-72.