

Development and Reliability of the OMERACT Thumb Base Osteoarthritis Magnetic Resonance Imaging Scoring System

Féline P.B. Kroon, Philip G. Conaghan, Violaine Foltz, Frédérique Gandjbakhch, Charles Peterfy, Iris Eshed, Harry K. Genant, Mikkel Østergaard, Margreet Kloppenburg, and Ida K. Haugen

ABSTRACT. Objective. To develop the Outcome Measures in Rheumatology (OMERACT) thumb base osteoarthritis (OA) magnetic resonance imaging (MRI) scoring system (TOMS) for the assessment of inflammatory and structural abnormalities in this hand OA subset, and test its cross-sectional reliability.

Methods. Included features and their scaling were agreed upon by members of the OMERACT MRI Task Force using the Hand OA MRI scoring system as a template. A reliability exercise was performed in which 3 readers participated, using a preliminary atlas with examples to facilitate reading. Each reader independently scored a set of 20 MRI (coronal and axial T1- and T2-weighted fat-suppressed images, of which 5 included T1-weighted fat-suppressed post-Gadolinium images). Intra- and inter-reader reliability were assessed using ICC, percentage exact agreement (PEA), and percentage close agreement (PCA).

Results. The TOMS assessed the first carpometacarpal (CMC-1) and scaphotrapezotrapezoid (STT) joints for synovitis, subchondral bone defects (including erosions, cysts, and bone attrition), osteophytes, cartilage, and bone marrow lesions on a 0–3 scale (normal to severe). Subluxation was evaluated only in the CMC-1 joint (absent/present). Reliability of scoring for both joints was comparable. Interreader ICC were good for all features (0.77–0.99 and 0.74–0.96 for CMC-1 and STT joints, respectively). Intrareader reliability analyses gave similar results. PCA was $\geq 65\%$ for all features. PEA was low to moderate, with better performance for subchondral bone defects, subluxation, and bone marrow lesions.

Conclusion. A thumb base OA MRI scoring system has been developed. The OMERACT TOMS demonstrated good intrareader and interreader reliability. Longitudinal studies are warranted to investigate reliability of change scores and responsiveness. (J Rheumatol First Release March 1 2017; doi:10.3899/jrheum.161099)

Key Indexing Terms:

OMERACT
MAGNETIC RESONANCE IMAGING

HAND
MAGNETIC RESONANCE IMAGING

THUMB BASE

OSTEOARTHRITIS
OUTCOMES RESEARCH

From the Departments of Rheumatology and Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands; Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds; National Institute for Health Research, Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK; Department of Rheumatology, Pitié Salpêtrière Hospital, APHP, Université Pierre et Marie Curie, Paris, France; Spine Sciences Inc., Boca Raton, Florida, USA; Department of Diagnostic Imaging, Sheba Medical Center, Tel Aviv University, Tel Aviv, Israel; Departments of Radiology and Medicine, University of California San Francisco, San Francisco, California, USA; Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Glostrup Hospital; Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway.

F.P. Kroon, MD, Department of Rheumatology, Leiden University Medical Center; P.G. Conaghan, MB, BS, PhD, FRACP, FRCP, Professor of Musculoskeletal Medicine, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and National Institute for Health Research, Leeds Musculoskeletal Biomedical Research Unit; V. Foltz, MD, Practicing Rheumatologist, Department of Rheumatology,

Pitié Salpêtrière Hospital, APHP, Université Pierre et Marie Curie; F. Gandjbakhch, MD, Practicing Rheumatologist; Department of Rheumatology, Pitié Salpêtrière Hospital, APHP, Université Pierre et Marie Curie; C. Peterfy, MD, PhD, FRCP, Chief Executive Officer, Spine Sciences Inc.; I. Eshed, MD, Associate Professor of Radiology, Department of Diagnostic Imaging, Sheba Medical Center, Tel Aviv University; H.K. Genant, MD, FACR, FRCP, Professor Emeritus of Radiology, Medicine and Orthopedics, Departments of Radiology and Medicine, University of California San Francisco; M. Østergaard, MD, PhD, DMSc, Professor of Rheumatology, Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Glostrup Hospital, and Department of Clinical Medicine, University of Copenhagen; M. Kloppenburg, MD, PhD, Professor of Rheumatology, Departments of Rheumatology and Clinical Epidemiology, Leiden University Medical Center; I.K. Haugen, MD, PhD, Postdoctoral Researcher; Department of Rheumatology, Diakonhjemmet Hospital.

Address correspondence to Dr. F.P. Kroon, Department of Rheumatology, Leiden University Medical Center, Post Box 9600, 2300 RC Leiden, the Netherlands. E-mail: f.kroon.reum@lumc.nl

Accepted for publication January 13, 2017.

Hand osteoarthritis (OA) affects the interphalangeal (IP) joints and the thumb base, including the first carpometacarpal (CMC-1) and scaphotrapezotrapezoid (STT) joints¹. Thumb base OA may consist of a separate hand OA subset, with distinct risk factors¹. However, much is unknown about the pathophysiology and disease course of hand OA subsets. New imaging modalities including magnetic resonance imaging (MRI) with visualization of all affected joint compartments may lead to increased insights into this disease.

Previously, the Hand OA MRI scoring system (HOAMRIS) for IP OA was developed with good cross-sectional and moderate longitudinal reliability^{2,3}. However, although the thumb base is commonly affected in patients with hand OA⁴, no MRI scoring systems assessing these joints exist to date. MRI studies of the thumb base of patients with hand OA can contribute to the understanding of this disease subset, including its differences from and similarities with IP OA.

The aim was to develop the Outcome Measures in Rheumatology (OMERACT) thumb base OA MRI scoring system (TOMS) for the assessment of inflammatory and structural abnormalities in thumb base OA, and to test its cross-sectional reliability using the OMERACT methodology^{5,6}.

MATERIALS AND METHODS

Development of the OMERACT TOMS. Using HOAMRIS as a template, members of the OMERACT MRI Task Force iteratively discussed in several Web-based meetings the joints and features (including definitions and scaling) to be included, as well as a list of preferred sequences and planes, and agreed by consensus.

Table 1 provides an overview of the proposed MRI features. Each feature was evaluated on 0–3 scales in the CMC-1 and STT joints, except subluxation, which was scored absent/present in the CMC-1 joint only. The proximal and distal joint parts were scored separately for subchondral bone defects, osteophytes, and bone marrow lesions (BML). For CMC-1, the proximal part of the first metacarpal bone (from the articular surface to a 1-cm depth) and distal half of the trapezium were evaluated (range 0–6). For STT, the proximal half of the trapezium and trapezoid and the distal half of the scaphoid were scored (range 0–9). Increments of 0.5 were introduced for synovitis, subchondral bone defects, and BML to increase potential responsiveness of the score.

Reliability exercise. A reliability exercise was conducted by 2 rheumatologists (VF, FG) and 1 radiologist (CP) with extensive experience in assessing hand/wrist MRI. Two readers (VF, FG) repeated the exercise after 1 month, after recoding and rearranging the MRI in a different order. A preliminary atlas with examples of most grades of each feature was developed prior to the exercise, approved by the members of the task force and distributed among readers to facilitate scoring. Each reader scored 20 MRI: 15 MRI were acquired on a 1.5T extremity MRI unit (ONI, GE) in patients with hand OA from the Hand Osteoarthritis in Secondary Care (HOSTAS) study at Leiden University Medical Center (Leiden, the Netherlands), and 5 MRI were acquired on a 3.0T MRI unit (Philips Ingenia) in patients with hand OA from Sheba Medical Center (Tel Aviv, Israel). MRI were selected by a nonreader to include a wide range of severity of pathology in the thumb base (based on Kellgren-Lawrence arthritis grading scale scores). MRI from HOSTAS included coronal and axial T1-weighted (T1w) fast spin echo (FSE), and T2w FSE images with fat-saturation (fs; Supplementary File 1 is available with the online version of this article). MRI from the Sheba Medical Center additionally included coronal and axial T1w-fs

post-gadolinium (Gd) images. A general wrist acquisition was used. Data collection in both centers was approved by the local ethics committee. All HOSTAS participants signed informed consent; written consent was waived for the use of MRI from the Sheba Medical Center.

Statistical analysis. Each MRI feature was analyzed separately for the CMC-1 and STT joints. Separate scores for the distal and proximal joint parts were combined into a single sum score per joint where appropriate. Median and interquartile range were calculated for each feature based on the mean value of the 3 readers. Reliability was assessed by calculating ICC, percentage exact agreement (PEA), and percentage close agreement (PCA). Single and average measure ICC (mixed-effect models, absolute agreement) were calculated to assess intrareader and interreader reliability, respectively. ICC values ≤ 0.20 were considered poor, > 0.20 to < 0.40 fair, ≥ 0.40 to < 0.60 moderate, ≥ 0.60 to < 0.80 good, and ≥ 0.80 very good reliability⁷. PEA was defined as a difference of 0 between minimum and maximum scores across readers, and PCA as a difference of ≤ 1 between minimum and maximum scores.

RESULTS

Supplementary Table 1 (available with the online version of this article) shows characteristics of the 15 HOSTAS patients. Most MRI features were present in the majority of patients (Table 2). STT joint scores were overall lower compared with CMC-1, despite higher possible score range for certain features. Time required to perform TOMS was comparable to that required to score 2 joints with HOAMRIS.

All features demonstrated good to very good interreader ICC values (Table 3). PCA was $\geq 65\%$ for all features. PEA was low to moderate, with better performance for subchondral bone defects, subluxation, and BML. Similar results were found for intrareader reliability (Supplementary Table 2 is available with the online version of this article). Reliability of the CMC-1 and STT joints were generally comparable.

When analyzing the reliability of subchondral bone defects, osteophytes, and BML for the distal and proximal joint parts separately, we generally saw ICC comparable to the aggregated scores. However, for subchondral bone defects in the trapezoid and osteophytes in the trapezoid and the proximal side of the trapezium, ICC were moderate (data not shown).

Readers gave slightly higher scores when assessing synovitis on post-Gd images as compared with the T2w-fs images (data not shown), whereas reliability was comparable (CMC-1: ICC 0.75, 95% CI 0.05–0.97 vs ICC 0.83, 95% CI 0.59–0.94; and STT: ICC 0.68, 95% CI –0.37 to 0.96 vs ICC 0.78, 95% CI 0.47–0.92 for images with vs without Gd).

DISCUSSION

In our study, the OMERACT MRI Task Force proposed the first thumb base MRI scoring system, TOMS, and evaluated its cross-sectional reliability. The score was feasible and had good to very good reliability for the assessment of structural and inflammatory features in the CMC-1 and STT joints.

The previously published OMERACT HOAMRIS for the IP joints was used as a prototype in the development of the TOMS³. Two major differences between the scoring systems

Table 1. Definitions and scaling of features in the proposed Outcome Measures in Rheumatology thumb base osteoarthritis MRI scoring system.

MRI Feature	Definition	Scaling*	Advised Plane and MRI Sequence
Synovitis [†]	Thickened synovium with enhancement after Gd injection.	0 = normal, 1 = mild (1%–33%), 2 = moderate (34%–66%), 3 = severe (67%–100%). Based on thirds of the presumed maximum thickness of enhancing tissue in the synovial compartment.	Coronal and axial. T1w pre- and post-Gd with fs. In absence of post-Gd images, T2w-fs/STIR/PD-fs can be used.
Subchondral bone defects ^{†#}	Subchondral bone loss, including erosions (sharply marginated bone lesions with cortical break), cysts (sharply marginated bone lesions without cortical break), and bone attrition (diffuse loss of bone contour).	0 = no bone defects, 1 = mild ($\leq 25\%$ of bone volume or joint surface affected), 2 = moderate (26%–50% of bone volume or joint surface affected), 3 = severe ($> 50\%$ of bone volume or joint surface affected).	Coronal and axial. T1w and T2w-fs/STIR/PD-fs.
Osteophytes [#]	Abnormal bone protuberance at joint margins or surfaces.	0 = no osteophytes, 1 = mild (1–2 small osteophytes), 2 = moderate [≥ 3 small osteophytes and/or ≥ 1 moderate osteophyte(s)], 3 = severe [≥ 1 large osteophyte(s)].	Coronal (and sagittal if available). T1w.
Cartilage assessment	Loss of cartilage, or loss of cartilage space based on the inter-bone distance. If assessment of cartilage and cartilage space are in conflict, direct visualization of the cartilage should be prioritized.	0 = no loss of cartilage or cartilage space, 1 mild (cartilage loss without complete denuding, or cartilage space loss without bone-to-bone contact), 2 = moderate (cartilage loss with denuding $\leq 50\%$ of joint surface or focal complete cartilage space loss with bone-to-bone contact $\leq 50\%$ of the articulating area), 3 = severe (cartilage loss with denuding $> 50\%$ of joint surface or complete cartilage space loss over $> 50\%$ of the articulating area).	Coronal. T1w-fs-3D-GE, otherwise use T1w-fs, T2w-fs, or PD-fs.
Subluxation [^]	Subluxation of the CMC-1 joint in the frontal plane.	0 = MC-1 subluxed 0%–25% of the MC width, 1 = MC-1 subluxed $\geq 26\%$ of the MC width.	Coronal. T1w.
Bone marrow lesions ^{†#}	Lesions within the trabecular bone with signal characteristic consistent with increased water content* and with ill-defined margins.	0 = no bone marrow lesions, 1 = mild (1%–33%), 2 = moderate (34%–66%), 3 = severe (67%–100%). Based on thirds of assessed bone volume.	Coronal and axial. T2w-fs/STIR/PD-fs.

[†] In longitudinal studies, 0.5 increments may be used for synovitis, subchondral bone defects, and bone marrow lesions. [#] Proximal and distal parts of a joint are scored separately for subchondral bone defects, osteophytes, and bone marrow lesions. [^] Only the CMC-1 joint is evaluated for this feature. * High signal intensity on STIR/T2w-fs images. MRI: magnetic resonance imaging; Gd: gadolinium; CMC-1: first carpometacarpal joint; MC-1: first metacarpal; w: weighted; fs: fat saturation; STIR: short-tau inversion recovery; PD: proton density; GE: gradient echo.

Table 2. Median (IQR) scores of each MRI feature and n (%) with each feature present for the CMC-1 and STT joints (n = 20). Separate scores for the distal and proximal part of the joint were combined into a single sum score per joint. The n (%) is of patients with each feature present according to at least 1 of 3 readers.

MRI Feature [Range for CMC-1/STT]	CMC-1 Joint		STT Joint	
	Median (IQR)	n (%)	Median (IQR)	n (%)
Synovitis [0–3/0–3]	1.4 (1.0–2.3)	20 (100)	1.0 (0.4–1.7)	18 (90)
Subchondral bone defects [0–6/0–9]	1.4 (1.0–2.8)	18 (90)	1.0 (0.3–2.0)	17 (85)
Osteophytes [0–6/0–9]	2.2 (1.2–4.0)	19 (95)	0.3 (0.0–0.9)	13 (65)
Cartilage assessment [0–3/0–3]	1.5 (0.4–2.3)	16 (80)	1.0 (0.4–1.3)	16 (80)
Subluxation [absent or present]	—	12 (60)	—	—
Bone marrow lesions [0–6/0–9]	1.7 (0.0–3.8)	13 (65)	1.4 (0.1–2.9)	15 (75)

IQR: interquartile range; MRI: magnetic resonance imaging; CMC-1: first carpometacarpal; STT: scaphotrapezio-trapezoid.

can be noted. First, erosive damage and cysts were combined into 1 score (subchondral bone defects) because it was judged that the distinction could not be made reliably in the thumb base joints. Second, because of larger joint size, it was reasoned that direct cartilage assessment is feasible in the

thumb base when using appropriate MRI sequences, and should be prioritized over indirect cartilage assessment. Further, it was decided to score distal and proximal joint parts separately for some features, similar to the first Oslo MRI scoring system for IP OA⁸. Because only 2 joints are

Table 3. Interreader reliability of MRI features for the CMC-1 and STT joints (3 readers). Separate scores for the distal and proximal part of the joint were combined into a single sum score per joint to calculate ICC.

Features	CMC-1 Joint			STT Joint		
	AvmICC (95% CI)	PEA, n/N (%)	PCA, n/N (%)	AvmICC (95% CI)	PEA, n/N (%)	PCA, n/N (%)
Synovitis	0.81 (0.60–0.92)	3/20 (15)	15/20 (75)	0.75 (0.48–0.90)	7/20 (35)	18/20 (90)
Subchondral bone defects	0.88 (0.73–0.95)	23/40 (58)	36/40 (90)	0.81 (0.60–0.92)	42/60 (70)	58/60 (97)
Osteophytes	0.83 (0.56–0.93)	10/40 (25)	31/40 (78)	0.74 (0.44–0.89)	41/60 (68)	56/60 (93)
Cartilage assessment	0.79 (0.48–0.92)	6/20 (30)	13/20 (65)	0.83 (0.64–0.93)	8/20 (40)	16/20 (80)
Subluxation	0.77 (0.52–0.91)	13/20 (65)	—	—	—	—
Bone marrow lesions	0.99 (0.98–1.00)	32/40 (80)	40/40 (100)	0.96 (0.91–0.98)	43/60 (72)	57/60 (95)

MRI: magnetic resonance imaging; CMC-1: first carpometacarpal; STT: scaphotrapeziotrapezoid; AvmICC: average measure ICC; PEA: percent exact agreement; PCA: percent close agreement.

evaluated, this addition provides more detailed information without decreasing feasibility. In future studies of pharmacological and nonpharmacological interventions, the HOAMRIS and TOMS can be used as complementary scoring systems, because both assess similar features. Combined assessment with MRI of the fingers and thumb base of patients with hand OA in future trials can provide information about hypothesized differences in the pathophysiology of these OA subtypes¹.

Assessment of the scaphotrapezoid articulation was also included in the scoring system. Previous cadaver studies have shown frequent degenerative changes of the scaphotrapezoid joint, although its relative contribution to STT joint OA complaints is unclear, partly because of poor visualization with traditional radiography^{9,10}.

All included MRI were performed using a standard wrist acquisition technique. Although dedicated thumb base acquisitions do exist, these are not widely used in clinical practice. It is unclear whether the use of a dedicated thumb base acquisition would yield different results, and this should be evaluated in future studies.

Only 5 MRI included post-Gd imaging. No previous studies have compared the reliability and validity of MRI-defined synovitis with and without contrast in patients with hand OA. In knee OA, synovitis is commonly assessed without contrast, although contrast-enhanced MRI appears to be a more reliable and valid measure of synovial inflammation, with the ability to differentiate inflamed synovium from effusion^{11,12}. Østergaard, *et al* found that omitting contrast from MRI examination of synovitis in the metacarpophalangeal and wrist joints in patients with rheumatoid arthritis decreased reliability¹³. In our sample, reliability was good using both contrast and noncontrast images. This warrants more detailed analysis, preferably comparing synovitis scores between different sequences within the same patient in a larger sample.

Before TOMS can be recommended as a core instrument according to the OMERACT filter⁶, assessment of the reliability of change scores and its responsiveness in longitudinal studies is needed. Future studies will reveal whether reli-

bility of TOMS is similar when used by other trained readers compared with expert readers who developed the scoring system, which for the HOAMRIS was shown to be either better or worse^{14,15}. Further, readers used a preliminary atlas during the exercises, which has likely increased agreement across readers, as was previously shown for the HOAMRIS³. A comprehensive atlas including all grades of all features in both joints would facilitate scoring and increase reliability of the TOMS. Validity of the scoring system should be investigated in future studies by assessing correlations with signs and symptoms, and other imaging modalities including traditional radiography and ultrasound.

ACKNOWLEDGMENT

We are indebted to I. Eshed (Department of Diagnostic Imaging of the Sheba Medical Center, Tel Aviv, Israel) and the Department of Radiology of the Leiden University Medical Center (Leiden, the Netherlands) for providing the magnetic resonance images for the scoring exercise.

ONLINE SUPPLEMENT

Supplementary material accompanies the online version of this article.

REFERENCES

1. Kloppenburg M, Kwok WY. Hand osteoarthritis—a heterogeneous disorder. *Nat Rev Rheumatol* 2012;8:22–31.
2. Haugen IK, Eshed I, Gandjbakhch F, Foltz V, Østergaard M, Bøyesen P, et al. The longitudinal reliability and responsiveness of the OMERACT Hand Osteoarthritis Magnetic Resonance Imaging Scoring System (HOAMRIS). *J Rheumatol* 2015;42:2486–91.
3. 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.
4. Dahaghin S, Bierma-Zeinstra SM, Ginai AZ, Pols HA, Hazes JM, Koes BW. Prevalence and pattern of radiographic hand osteoarthritis and association with pain and disability (the Rotterdam study). *Ann Rheum Dis* 2005;64:682–7.
5. Boers M, Kirwan JR, Gossec L, Conaghan PG, D'Agostino MA, Bingham CO 3rd, et al. How to choose core outcome measurement sets for clinical trials: OMERACT 11 approves filter 2.0. *J Rheumatol* 2014;41:1025–30.
6. Boers M, Kirwan JR, Tugwell P, Beaton D, Bingham CO 3rd, Conaghan PG, et al. The OMERACT handbook. [Internet. Accessed January 18, 2017.] Available from: www.omeract.org/pdf/OMERACT_Handbook.pdf

7. Müller R, Büttner P. A critical discussion of intraclass correlation coefficients. *Stat Med* 1994;13:2465-76.
8. 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.
9. Bhatia A, Pisoh T, Touam C, Oberlin C. Incidence and distribution of scaphotrapezotrapezoidal arthritis in 73 fresh cadaveric wrists. *Ann Chir Main Memb Super* 1996;15:220-5.
10. Moritomo H, Viegas SF, Nakamura K, Dasilva MF, Patterson RM. The scaphotrapezio-trapezoidal joint. Part 1: An anatomic and radiographic study. *J Hand Surg Am* 2000;25:899-910.
11. Guermazi A, Roemer FW, Crema MD, Englund M, Hayashi D. Imaging of non-osteochondral tissues in osteoarthritis. *Osteoarthritis Cartilage* 2014;22:1590-605.
12. Hayashi D, Roemer FW, Katur A, Felson DT, Yang SO, Alomran F, et al. Imaging of synovitis in osteoarthritis: current status and outlook. *Semin Arthritis Rheum* 2011;41:116-30.
13. Østergaard M, Conaghan PG, O'Connor P, Szkudlarek M, Klarlund M, Emery P, et al. Reducing invasiveness, duration, and cost of magnetic resonance imaging in rheumatoid arthritis by omitting intravenous contrast injection — Does it change the assessment of inflammatory and destructive joint changes by the OMERACT RAMRIS? *J Rheumatol* 2009;36:1806-10.
14. Ramonda R, Favero M, Vio S, Lacognata C, Frallonardo P, Belluzzi E, et al. A recently developed MRI scoring system for hand osteoarthritis: its application in a clinical setting. *Clin Rheumatol* 2016;35:2079-86.
15. Kortekaas MC, Kwok WY, Reijnierse 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.