Article:
Haugen, IK, Østergaard, M, Eshed, I et al. (10 more authors) (2014) Iterative development and reliability of the OMERACT hand osteoarthritis MRI scoring system. The Journal of Rheumatology, 41 (2). pp. 386-391. ISSN 0315-162X

https://doi.org/10.3899/jrheum.131086

Development and reliability of the OMERACT thumb base osteoarthritis MRI scoring system.

Féline PB Kroon[1], Philip G Conaghan[2], Violaine Foltz[3], Frédérique Gandjbakhch[4], Charles Peterfy[5], Iris Eshed[6], Harry K Genant[7], Mikkel Østergaard[8], Margreet Kloppenburg[9], Ida K Haugen[10]

[1]MD, Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands; f.kroon.reum@lumc.nl
[2]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, Leeds, United Kingdom; P.Conaghan@leeds.ac.uk
[3]MD, Practicing Rheumatologist; Department of Rheumatology, Pitié Salpêtriere Hospital, APHP, Université Pierre et Marie Curie, Paris, France; violaine.foltz@aphp.fr
[4]MD, Practicing Rheumatologist; Department of Rheumatology, Pitié Salpêtriere Hospital, APHP, Université Pierre et Marie Curie, Paris, France; frederique.gandjbakhch@aphp.fr
[5]MD, PhD, FRCP, Chief Executive Officer; Spire Sciences Inc., Boca Raton, Florida, USA; charles.peterfy@spiresciences.com
[6]MD, Associate Professor of Radiology; Department of Diagnostic Imaging, Sheba Medical Center, Tel Aviv University, Tel Aviv, Israel; iriseshed@gmail.com
[7]MD, FACR, FRCR, Professor Emeritus of Radiology, Medicine and Orthopedics; Departments of Radiology and Medicine, University of California San Francisco, San Francisco, California, USA; Harry.Genant@ucsf.edu
[8]MD, PhD, DMSc, Professor of Rheumatology; Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Glostrup Hospital, Copenhagen, Denmark and Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; mo@dadlnet.dk
[9]MD, PhD, Professor of Rheumatology; Departments of Rheumatology and Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; g.kloppenburg@lumc.nl
[10]MD, PhD, Postdoctoral Researcher; Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway; ida.k.haugen@gmail.com

Correspondence to FPB Kroon, Department of Rheumatology, Leiden University Medical Center, Post Box 9600, 2300 RC Leiden, The Netherlands. E-mail: f.kroon.reum@lumc.nl .

Key Indexing Terms. OMERACT, hand, thumb base, osteoarthritis, magnetic resonance imaging, outcomes research

Funding. none

Running title. TOMS: Development and Reliability

ABSTRACT

Objective. To develop the OMERACT thumb base osteoarthritis MRI scoring system (TOMS) for assessment of inflammatory and structural abnormalities in this hand osteoarthritis 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 Osteoarthritis MRI scoring system (HOAMRIS) 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 MRIs (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 intra-class correlation coefficients (ICC) and percentage exact and close agreement (PEA, PCA).

Results. The TOMS assessed the first carpometacarpal (CMC-1) and scaphotrapeziotrapezoid (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 only evaluated in CMC-1 joint (absent/present). Reliability of scoring for both joints was comparable. Inter-reader ICCs were good for all features (0.77-0.99 and 0.74-0.96 for CMC-1 and STT-joints respectively). Intra-reader reliability analyses gave similar results. PCA was ≥65% for all features. PEA was low to moderate, with better performance for subchondral bone defects, subluxation and bone marrow lesions.

Conclusions. A thumb base osteoarthritis MRI scoring system has been developed. The OMERACT TOMS demonstrated good intra- and inter-reader reliability. Longitudinal studies are warranted to investigate reliability of change scores and responsiveness.

What is new:

- The OMERACT MRI Task Force proposed the first thumb base MRI scoring system (TOMS) to assess inflammatory and structural abnormalities in thumb base OA;
- The OMERACT TOMS demonstrated good intra- and inter-reader reliability in a cross-sectional reliability exercise.

INTRODUCTION

Hand osteoarthritis (OA) affects the interphalangeal (IP) joints and the thumb base, including the first carpometacarpal (CMC-1) and scaphotrapeziotrapezoid (STT) joints(1). Thumb base OA may comprise a separate hand OA subset, with distinct risk factors(1). 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 Osteoarthritis 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 hand OA patients(4), no MRI scoring systems assessing these joints exist to date. MRI studies of the thumb base of hand OA patients can contribute to the understanding of this disease subset, including its differences and similarities with IP OA.

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

METHODS

Development of the OMERACT TOMS

Using HOAMRIS as template, members of the OMERACT MRI Task Force iteratively discussed the joints and features (including definitions and scaling) to be included, as well as a list of preferred sequences and planes, in several Web-based meetings, 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 (BMLs). For CMC-1, the proximal part of the first metacarpal bone (MC-1) (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 BMLs to increase potential responsiveness of the score.

Reliability exercise

A reliability exercise was conducted by two rheumatologists (VF, FG) and one radiologist (CP) with extensive experience in assessing hand/wrist MRIs. Two readers (VF, FG) repeated the exercise after one month after recoding and rearrangement of MRIs 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 MRIs: 15 MRIs were acquired on a 1.5T extremity MRI unit (ONI, GE, Wisconsin, USA) in hand OA patients from the Hand Osteoarthritis in Secondary Care (HOSTAS) study at Leiden University Medical Center (Leiden, Netherlands), and 5 MRIs were acquired on a 3.0T MRI unit (Philips Ingenia) in hand OA patients from Sheba Medical Center (Tel Aviv, Israel). MRIs were selected by a non-reader to include a wide range of severity of pathology in the thumb base (based on Kellgren-Lawrence scores). MRIs from HOSTAS included coronal and axial T1-weighted (T1w) fast spin echo (FSE), and T2w FSE images with fat-saturation (fs) (Supplementary file). MRIs from 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 waved for use of MRIs from Sheba Medical Center.

Statistical analysis

Each MRI feature was analysed 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 (IQR) were calculated for each feature based on the mean value of the three readers. Reliability was assessed by calculating intra-class correlation coefficients (ICCs) and percentage exact and close agreement (PEA/PCA). Single and average measure ICCs (mixed effect models, absolute agreement) were calculated to assess intra- and inter-reader reliability, respectively. ICC values ≤0.20 were considered as poor, 0.20≤ICC<0.40 as fair, 0.40≤ICC<0.60 as moderate, 0.60≤ICC<0.80 as good, and ICC ≥0.80 as very good reliability(7). 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 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 to CMC-

1, despite higher possible score range for certain features. Time required to perform TOMS was comparable to that required to score two joints with HOAMRIS.
All features demonstrated good to very good inter-reader ICC values (Table 3). PCA was ≥65% for all features. PEA was low to moderate, with better performance for subchondral bone defects, subluxation and BMLs. Similar results were found for intra-reader reliability (Supplementary Table 2). Reliability of the CMC-1 and STT-joints were generally comparable.

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

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

DISCUSSION

In this 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 assessment of structural and inflammatory features in the CMC-1 and STT-joint.

The previously published OMERACT HOAMRIS for the IP joints was used as a prototype in the development of the TOMS(3). Two major differences between the scoring systems can be noted. First, erosive damage and cysts were combined into one score (subchondral bone defects), because it was judged that the distinction could not be made reliably in the thumb base joints. Second, due to 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. Furthermore, it was decided to score distal and proximal joint parts separately for some features, similar to the first Oslo MRI scoring system for IP OA(8). Since only two joints are evaluated, this addition provides more detailed information without decreasing feasibility. In future studies of pharmacological and non-pharmacological interventions, HOAMRIS and TOMS can be used as complimentary scoring systems, since both assess similar features. Combined assessment of the fingers and thumb base of hand OA patients with MRI in future trials can provide information about hypothesized differences in the pathophysiology of these OA subtypes(1).

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 MRIs included 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 five MRIs included post-Gd imaging. No previous studies have compared the reliability and validity of MRI-defined synovitis with and without contrast in hand OA patients. 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 rheumatoid arthritis patients decreased reliability(13). In our sample reliability was good using both contrast and non-contrast images. This warrants more detailed exploration, preferably comparing synovitis scores between different sequences within the same patient in a larger sample.

Before TOMS can be recommended as core instrument according to the OMERACT filter(6), assessment of the reliability of change scores and its responsiveness in longitudinal studies is needed. Future studies will reveal whether reliability of TOMS is similar when used by other trained readers compared to expert readers who developed the scoring system, which for HOAMRIS was shown to be either better or worse(14,15). Furthermore, readers used a preliminary atlas during the exercises, which has likely increased agreement across readers, as was previously shown for HOAMRIS(3). 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 MRIs for the scoring exercise.

REFERENCES

1. Kloppenburg M, Kwok W-Y. Hand osteoarthritis--a heterogeneous disorder. Nat Rev Rheumatol. 2012;8:22-31.
2. Haugen IK, Eshed I, Gandjbakhch F, Foltz V, Østergaard M, Boyesen 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, Ostergaard 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 S, Ginai A, Pols H, Hazes J, Koes B. 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 J, Gossec L, Conaghan P, D'Agostino M, Bingham C, 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, Tugwell P, Beaton D, Bingham CO, Conaghan P, D'Agostino M-A, et al. The OMERACT Handbook: Available from: http://www.omeract.org/pdf/OMERACT_Handbook.pdf.; 2016.
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;70:1033-38.
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 S, Nakamura K, Dasilva M, Patterson R. The scaphotrapezio-trapezoidal joint. Part 1: An anatomic and radiographic study. J Hand Surg. 2000;25A: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.

Table 1. Definitions and scaling of features in proposed OMERACT 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 (≤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 ≤50% of joint surface or focal complete cartilage space loss with bone-to-bone contact ≤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 ≥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. [*High signal intensity on STIR/T2w-fs images]	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 joint are scored separately for subchondral bone defects, osteophytes and bone marrow lesions. ^Only the CMC-1 joint is evaluated for this feature. CMC-1, first carpometacarpal joint. fs, fat saturated. Gd, gadolinium. GE, gradient echo. MC-1, first metacarpal. PD, proton density. w, weighted.

Table 2. Median (interquartile range) scores of each MRI feature and number of patients (%) with each feature present for the CMC-1 and STT joint (n=20).

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)

Separate scores for the distal and proximal part of the joint were combined into a single sum score per joint. Number (%) of patients with each feature present in at least one of three readers. CMC-1, first carpometacarpal. IQR, interquartile range. MRI, magnetic resonance imaging. n, number. STT, scaphotrapeziotrapezoid.

Table 3. Inter-reader reliability of MRI features for the CMC-1 and STT joint (3 readers).

	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)

Separate scores for the distal and proximal part of the joint were combined into a single sum score per joint to calculate ICCs. AvmICC, average measure intra-class correlation coefficient. CI, confidence interval. CMC-1, first carpometacarpal. IQR, interquartile range. MRI, magnetic resonance imaging. N, number. PCA, percent close agreement. PEA, percent exact agreement. STT, scaphotrapeziotrapezoid.

ONLINE SUPPLEMENTARY FILE

Additional MRI sequence information HOSTAS

MRIs from the HOSTAS study included T1-weighted (T1w) fast spin echo (FSE) images in coronal and axial planes (TR/TE 575/11.2, slice thickness 2.0 and 3.0 mm, slice gap 0.2 and 0.3 mm), and T2w FSE images with frequency-selective fat-saturation in coronal and axial planes (TR/TE 3000/61.8, slice thickness 2.0 and 3.0 mm, slice gap 0.2 and 0.3 mm).

Supplementary table 1. Demographic and clinical characteristics of HOSTAS patients included in the reliability exercise (n=15).

	Patients in exercise (n=15)*
Women, n (%)	12 (80%)
Age, mean (SD) years	65.3 (9.0)
Body mass index, mean (SD) kg/m^2	29.6 (5.4)
Kellgren-Lawrence grade CMC-1†#, n (%)	
Grade 1	5 (36%)
Grade 2	5 (36%)
Grade 3	3 (21%)
Grade 4	1 (7%)
Grip strength†, mean (SD) kg	23.4 (8.3)
AUSCAN pain, mean (SD) [0-20]	8.9 (3.2)
AUSCAN function, mean (SD) [0-36]	14.3 (5.5)
VAS pain†, mean (SD) [0-100]	36.4 (20.6)
Self-reported joint pain thumb†, n(%)	11 (73%)
Self-reported joint stiffness thumb†, n(%)	7 (47%)
Bony swelling CMC-1 joint†, n (%)	5 (33%)
Tenderness on palpation CMC-1 joint†, n(%)	7 (47%)
Limited range of motion CMC-1†, n(%)	4 (27%)

**Information only available for patients from the HOSTAS cohort. †Only data of the imaged hand are displayed. #Data from n=14 patients. AUSCAN, Australian/Canadian hand osteoarthritis index. CMC-1, first carpometacarpal. kg, kilogram. n, number. SD, standard deviation.*

Supplementary table 2. Intra-reader reliability of MRI features for the CMC-1 and STT joint (2 readers).

	CMC-1 joint			STT joint		
	SmICC	PEA (%)	PCA (%)	SmICC	PEA (%)	PCA (%)
Synovitis	0.53-0.83	65	90-95	0.72-0.89	65-90	100
Subchondral bone defects	0.89-0.89	95	100	0.62-0.70	90-95	100
Osteophytes	0.71-0.73	60-70	100	0.44-0.71	90-100	100
Cartilage assessment	0.61-0.86	75	95-100	0.71-0.84	65-70	100
Subluxation	0.53-0.91	80-95				
Bone marrow lesions	0.98-0.96	100	100	0.87-0.92	95	100

Values of both readers shown separately (lowest-highest), unless values were not different. Separate scores for the distal and proximal part of the joint were combined into a single sum score per joint to calculate ICCs. CI, confidence interval. CMC-1, first carpometacarpal. IQR, interquartile range. MRI, magnetic resonance imaging. N, number. PCA, percent close agreement. PEA, percent exact agreement. SmICC, single measure intra-class correlation coefficient. STT, scaphotrapeziotrapezoid.