Longitudinal Reliability of the OMERACT Thumb Base Osteoarthritis Magnetic Resonance Imaging Scoring System (TOMS)

Féline P.B. Kroon, Sjoerd van Beest, Frédérique Gandjbakhch, Charles G. Peterfy, Su Chen, Philip G. Conaghan, Iris Eshed, Violaine Foltz, Harry K. Genant, Ida K. Haugen, Jeroen K. Medema, Mikkel Østergaard, Lanju Zhang, Marc C. Levesque, and Margreet Kloppenburg

ABSTRACT. ***Objective.*** To assess the longitudinal reliability of the Outcome Measures in Rheumatology (OMERACT) Thumb base Osteoarthritis Magnetic resonance imaging (MRI) Scoring system (TOMS).

Methods. Paired MRI of patients with hand osteoarthritis were scored in 2 exercises (6-mo and 2-yr followup) for synovitis, subchondral bone defects (SBD), osteophytes, cartilage assessment, bone marrow lesions (BML), and subluxation. Interreader reliability of delta scores was assessed.

Results. Little change occurred. Average-measure intraclass correlation coefficients were good-excellent (≥ 0.71), except synovitis (0.55–0.83) and carpometacarpal-1 osteophytes/cartilage assessment (0.47/0.39). Percentage exact/close agreement was 52–92%/68–100%, except BML in 2 years (28%/64–76%). Smallest detectable change was below the scoring increment, except in SBD and BML.

Conclusion. TOMS longitudinal reliability was moderate-good. Limited change hampered assessment.
(First Release December 15 2018; J Rheumatol 2019;46:1228–31; doi:10.3899/jrheum.180949)

Key Indexing Terms:
OMERACT HAND THUMB BASE OSTEOARTHRITIS
MAGNETIC RESONANCE IMAGING OUTCOMES RESEARCH

From the Departments of Rheumatology and Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands; Paris 6, Sorbonne Universités, GRC-08 (EEMOIS); AP-HP, Department of Rheumatology, Pitié Salpêtrière University Hospital, Paris, France; Spire Sciences Inc., Boca Raton, Florida; Data and Statistical Sciences, and Immunology Development, AbbVie Inc., North Chicago, Illinois; Departments of Radiology and Medicine, University of California at San Francisco, San Francisco, California; AbbVie Inc., Cambridge, Massachusetts, USA; Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds and National Institute for Health Research, Leeds Biomedical Research Centre, Leeds, UK; Department of Diagnostic Imaging, Sheba Medical Center, Tel Aviv University, Tel Aviv, Israel; Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway; Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet, Glostrup; Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.

PGC is supported in part by the UK National Institute for Health Research infrastructure at Leeds.

F.P. Kroon, MD, Department of Rheumatology, Leiden University Medical Center; S. van Beest, MD, Department of Rheumatology, Leiden University Medical Center; F. Gandjbakhch, MD, Practicing Rheumatologist, Paris 6, Sorbonne Universités, GRC-08 (EEMOIS), and AP-HP, Department of Rheumatology, Pitié Salpêtrière University Hospital; C.G. Peterfy, MD, PhD, FRCP, Chief Executive Officer, Spire Sciences Inc.; S. Chen, Data and Statistical Sciences, AbbVie Inc.; 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 Biomedical Research Centre; I. Eshed, MD, Associate Professor of Radiology, Department of Diagnostic Imaging, Sheba Medical Center, Tel Aviv University; V. Foltz, MD, Practicing Rheumatologist, Paris 6, Sorbonne Universités, GRC-08 (EEMOIS), and AP-HP, Department of Rheumatology, Pitié Salpêtrière University Hospital; H.K. Genant, MD, FACR, FRCR, Professor Emeritus of Radiology, Medicine and Orthopedics, Departments of Radiology and Medicine, University of California at San Francisco; I.K. Haugen, MD, PhD, Postdoctoral Researcher, Department of Rheumatology, Diakonhjemmet Hospital; J.K. Medema, MS, Senior Project Leader, Immunology Development, AbbVie Inc.; M. Østergaard, MD, PhD, DMSc, Professor of Rheumatology, Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet, Glostrup, Denmark, and Department of Clinical Medicine, University of Copenhagen; L. Zhang, Data and Statistical Sciences, AbbVie Inc.; M.C. Levesque, MD, PhD, Foundational Immunology Senior Director, AbbVie Inc.; M. Kloppenburg, MD, PhD, Professor of Rheumatology, Departments of Rheumatology and Clinical Epidemiology, Leiden University Medical Center.

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 November 7, 2018.

The thumb base, including the first carpometacarpal (CMC-1) and scaphotrapeziotrapezoid (STT) joints, is often involved in hand osteoarthritis (OA). Thumb base OA is associated with particular risk factors and requires distinct therapeutic interventions compared to interphalangeal finger OA[1]. Therefore, outcome measures specifically assessing thumb base OA are needed.

In response, the Outcome Measures in Rheumatology (OMERACT) magnetic resonance imaging (MRI) working group developed a scoring system of MRI findings in the thumb base: Thumb base OA MRI Scoring system (TOMS)[2]. This tool has been shown to exhibit good cross-sectional reliability, but data concerning longitudinal reliability are lacking[2]. By the term *longitudinal reliability*, we mean the ability to reliably score sequential images, taking into account interreader variability. Understanding the reliability of TOMS for measuring change is needed for effectively implementing this tool.

Our study investigated the longitudinal reliability of TOMS in 2 settings: a prospective observational study with longterm followup and a clinical trial with short-term followup[3,4].

MATERIALS AND METHODS

Reliability exercises. Two reliability exercises were performed. An atlas was available to facilitate scoring[5]. Features assessed were synovitis, subchondral bone defects (SBD), osteophytes, cartilage assessment, bone marrow lesions (BML), and subluxation[2]. All features but subluxation were evaluated on 0–3 scales in the CMC-1 and STT joints, with 0.5 increments for synovitis, SBD, and BML. Proximal and distal joint parts were scored separately for SBD, osteophytes, and BML. Subluxation was scored absent/present in the CMC-1 joint. In both exercises, MRI were selected to represent a large range of pathology.

In the first exercise, paired MRI (baseline, 2-yr followup) of 25 patients from the Hand Osteoarthritis in Secondary Care (HOSTAS) prospective cohort study (Leiden University Medical Center[6]) were scored in known time-order by 3 independent readers [1 rheumatologist (FG) and 2 rheumatology fellows (SvB, FK), all experienced in using TOMS]. Coronal and axial T1-weighted (T1W) fast spin echo (FSE), and T2W FSE images with fat-suppression (FS) were obtained on a 1.5T extremity MRI unit (ONI, GE; Supplementary File, available with the online version of this article). No contrast agent was used. Therefore, synovitis was scored on T2W-FS images, as per the original scoring system[2].

The second exercise was conducted by an experienced radiologist (CP) and a rheumatology fellow (FK). Paired MRI (baseline, 6-mos followup) of 24 patients with hand OA from a multicenter randomized double-blind trial comparing lutikizumab to placebo[7] were scored for synovitis and BML. One reader (CP) scored in unknown and the other in known time-order (FK) for logistical reasons. Coronal and axial T1W-FS images with/without gadolinium-based contrast enhancement, and short-tau inversion recovery or T2W-FS images were obtained according to standardized protocol. Because of incomplete coverage, the STT could only be assessed in 16 patients, and the trapezoid bone was not evaluated.

Data collection for both studies was approved by local ethics committees (P09.004, NCT02384538). All participants provided written informed consent.

Statistical analyses. Separate scores of distal and proximal joint compartments were combined into 1 sum score per joint where applicable. Median and interquartile range (baseline status scores) or range (delta scores) was calculated for each feature, based on the average of the readers. Interreader reliability of delta scores was assessed by calculating intraclass correlation coefficients (ICC; average measure, mixed-effect models, absolute agreement), and percentage exact and close agreement (PEA/PCA). ICC ≤ 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 excellent reliability[8]. PEA/PCA were defined as a difference of 0/≤ 1 between minimum and maximum scores across readers. For each feature, the smallest detectable change (SDC) was calculated[9]. We determined how many patients changed beyond measurement error (i.e., change score > SDC), and whether the smallest scoring increment for each feature could be scored reliably (i.e., smallest increment > SDC).

RESULTS

Table 1[10] presents baseline characteristics of patients from both reliability exercises. Thirteen trial participants received placebo and 11 lutikizumab. Baseline scores of MRI features were generally low (Table 2). Highest scores were given for CMC-1 osteophytes. Overall, more MRI abnormalities were seen in the CMC-1 compared to the STT joint.

Baseline scores of synovitis and BML were comparable in the 2 studies. On average, little change was observed after 6 months and 2 years (Table 2). However, individual patients showed change in synovitis and BML, both increasing and decreasing (Supplementary Figure 1, available with the online version of this article). Cartilage and bone features generally showed less improvement and more deterioration over time.

Table 3 presents the longitudinal reliability in both studies. ICC for most features in both thumb base joints were good to excellent. Fair to moderate ICC were found for cartilage assessment and osteophytes in the CMC-1 joint. ICC for synovitis in the different studies and joints varied from moderate to excellent. ICC could not be estimated for some features (STT synovitis in the clinical trial, STT osteophytes, and subluxation).

Since calculation of ICC was influenced by the small amount of change that occurred over time in both studies, PEA and PCA values were also calculated. PEA/PCA of all features in both joints ranged from 52–92% and 68–100%, except for BML in the CMC-1 in the 3-reader exercise (PEA 28%/PCA 64%). PEA values in that exercise were all lower than for the clinical trial.

Table 1. Baseline characteristics of hand osteoarthritis patients in 2 reliability exercises. Values are n (%) unless otherwise specified.

Clinical Characteristics	HOSTAS Cohort, n = 25	Clinical Trial, n = 24
Women	23 (92)	20 (83)
Age, yrs, mean (SD)	60.0 (7.5)	65.9 (6.8)
Fulfilling ACR hand OA criteria	24 (96)	24 (100)
Pain on palpation thumb base	16 (64)	14 (58)
KL grade CMC-1		
Grade 0	10 (40)	7 (29)
Grade 1	5 (20)	0 (0)
Grade 2	5 (20)	8 (33)
Grade 3	3 (12)	5 (21)
Grade 4	2 (8)	4 (17)
Osteophyte STT†	2 (8)	7 (44)*
Joint space narrowing STT†	6 (24)	6 (38)*

*STT data from 16 patients. †According to the Osteoarthritis Research Society International atlas[10]. HOSTAS: Hand Osteoarthritis in Secondary Care; ACR: American College of Rheumatology; OA: osteoarthritis; KL: Kellgren-Lawrence; CMC-1: first carpometacarpal joint; STT: scapho-trapeziotrapezoid joint.

Table 2. Baseline status (median, interquartile range) and change scores (median, range) of each MRI feature for the CMC-1 and STT joints in 2 reliability exercises. Based on average score of all readers, except subluxation†. Separate scores for the distal and proximal part of the joint combined into sum score per joint.

MRI Feature, Range CMC-1/STT	HOSTAS Cohort, n = 25				Clinical Trial, n = 24			
	CMC-1		STT		CMC-1		STT*	
	Baseline	Change, 2 Yrs	Baseline	Change, 2 Yrs	Baseline	Change, 6 Mos	Baseline	Change, 6 Mos
Synovitis, 0–3/0–3	1.3 (0.7–1.7)	0 (–1.7 to 1)	0.7 (0.3–1.3)	0 (–0.7 to 1)	1.5 (1–2)	0 (–1 to 1)	0.5 (0–1)	0 (0–0.5)
Subchondral bone defects, 0–6/0–9	1.7 (0.7–2.3)	0.2 (–0.7 to 2.2)	0.7 (0–1.7)	0 (–0.2 to 2.3)				
Osteophytes, 0–6/0–9	2.3 (1.7–4)	0 (0–0.7)	0.7 (0.3–1)	0 (0–0.3)				
Cartilage assessment, 0–3/0–3	1 (0.7–1.7)	0 (–0.3 to 0.7)	0.7 (0–1.3)	0 (–1.7 to 1)				
Subluxation, absent or present	8 (32%)†	0 (0 –0.3)						
Bone marrow lesions, 0–6/0–9	1.7 (0.7–4.3)	0 (–3.2 to 4.7)	1 (0.3–2.3)	0 (–2.7 to 3)	1.3 (0.5–2.5)	0 (–5 to 4)	0 (0–1.3)	0 (–2.5 to 3)

*STT scored in 16 patients, trapezoid not included. †n (%) with subluxation scored by at least 2 readers. MRI: magnetic resonance imaging; CMC-1: first carpometacarpal joint; STT: scaphotrapeziotrapezoid joint; HOSTAS: Hand Osteoarthritis in Secondary Care.

Table 3. Interreader reliability of change scores of MRI features for the CMC-1 and STT joint in 2 reliability exercises.

MRI Feature [smallest Increment]	CMC-1				STT			
	AvmICC (95% CI)	PEA, n (%)	PCA, n (%)	SDC	AvmICC (95% CI)	PEA, n (%)	PCA, n (%)	SDC
HOSTAS cohort, n = 25								
Synovitis [0.5]	0.83 (0.68–0.92)	14 (56)	25 (100)	0.45	0.56 (0.12–0.79)	15 (60)	24 (96)	0.48
Subchondral bone defects [0.5]	0.72 (0.47–0.87)	13 (52)	17 (68)	0.73	0.71 (0.44–0.86)	16 (64)	22 (88)	0.63
Osteophytes [1]	0.47 (–0.02 to 0.75)	22 (88)	25 (100)	0.22	†	23 (92)	25 (100)	0.18
Cartilage assessment [1]	0.39 (–0.18 to 0.71)	16 (64)	25 (100)	0.39	0.72 (0.47–0.87)	20 (80)	24 (96)	0.43
Subluxation [1]	†	23 (92)		0.18				
Bone marrow lesions [0.5]	0.84 (0.69–0.93)	7 (28)	16 (64)	1.27	0.92 (0.83–0.96)	7 (28)	19 (76)	0.67
Clinical trial, n = 24*								
Synovitis [0.5]	0.55 (–0.07 to 0.80)	17 (71)	23 (100)	0.65	†	14 (88)	16 (100)	0.37
Bone marrow lesions [0.5]	0.89 (0.75–0.95)	17 (71)	22 (92)	0.87	0.90 (0.68–0.97)	13 (87)	14 (93)	0.77

*STT scored in 16 patients, trapezoid not included. †Reliable estimation of ICC not possible owing to low variability. CMC-1: first carpometacarpal joint; STT: scaphotrapeziotrapezoid joint; MRI: magnetic resonance imaging; AvmICC: average measure intraclass correlation coefficient; PCA: percent close agreement; PEA: percent exact agreement; SDC: smallest detectable change; HOSTAS: Hand Osteoarthritis in Secondary Care.

The SDC was calculated for all features and should be considered in light of the range and smallest increment of that feature's score (Table 3). Most SDC were lower than that feature's smallest scoring increment, although the SDC of SBD and BML in particular were higher than the increment of 0.5. In the cohort study, the SDC for BML in the CMC-1 was even higher than 1 (SDC = 1.27), although in the clinical trial the SDC was better (SDC = 0.87). Most participants did not change more than the SDC (Supplementary Table 1, available with the online version of this article). The largest number of participants with a delta score larger than the SDC, either increasing or decreasing, occurred for synovitis and BML. Features related to cartilage and bone generally deteriorated. Of these, SBD showed the most participants with change.

DISCUSSION

In our report, we show the longitudinal reliability of a recently developed OMERACT MRI scoring system to assess inflammatory and structural features in TOMS. Based on ICC, PEA, and PCA values, our investigation showed that reliability of assessment of delta scores using the TOMS was good.

The longitudinal reliability of the similar Hand Osteoarthritis Magnetic Resonance Imaging Scoring System (HOAMRIS) to evaluate interphalangeal joints was previously published[11]. Because the HOAMRIS and TOMS assess similar features, similar reliability is expected. Reliability of change scores in the HOAMRIS exercise (20 patients, 3 readers) for erosive damage and cysts was similar to those for SBD in TOMS. BML were also reliably assessed in both

studies. However, our results for synovitis, osteophytes, and cartilage assessment were better compared to HOAMRIS. Observed differences between the studies may partly be explained by a higher number of assessed joints for the HOAMRIS, leading to lower PEA/PCA values. Interphalangeal joints are also smaller, and the field strength of the magnetic resonance scanner was lower, which made reliable assessment more difficult.

ICC of the previous cross-sectional reliability exercise of the TOMS were generally higher, while PEA/PCA values were lower[2]. These differences were found because assessment of ICC of delta scores in a cohort with little change over time generally results in lower values, because ICC values are not only dependent on measurement error, but also on between-subject variability. Between-subject variability is part of the calculation used to produce ICC values, and low between-subject variability can cause unreasonably low ICC values[12]. Results of the 2 exercises performed in our study were generally comparable, although the difference in blinding for time-order among readers of the clinical trial may have resulted in lower results for agreement between these readers. PEA values in the 3-reader exercise were all lower than for the 2-reader exercise, which can at least partially be attributed to the higher number of readers who have to reach exact agreement in the first case.

Assessment of longitudinal reliability was hampered by the small magnitude of change. Continuous change scores and the number of patients changing more than the SDC were low. Both cohorts reflect the characteristic disease course. In the cohort study, no intervention was given, and inflammatory features were not expected to change. However, over a 2-year period, cartilage and bone damage were expected to increase, which they did, though only mildly. Generally, radiographic progression in the CMC-1 over 2 years is slow[13]. Moreover, we selected participants with and without thumb base OA for this methodological exercise, which may have contributed to the low amount of change that was observed over time.

Most SDC were low and below the feature's smallest scoring increment, showing that a change of 1 increment reflects a measurable change in that feature. Only SBD and BML had an SDC above their defined smallest increment of 0.5, and it could be argued that 0.5 increments are too small to be reliably assessed for these features.

Results from our study provide evidence that the OMERACT TOMS can be used to evaluate thumb base MRI in studies of different settings. Future studies are warranted, in particular positive clinical trials, to evaluate sensitivity to change, as well as validation studies.

ONLINE SUPPLEMENT

Supplementary material accompanies the online version of this article.

ACKNOWLEDGMENT

We are indebted to AbbVie (North Chicago, Illlinois, USA) and the Department of Radiology of the Leiden University Medical Center (Leiden, the Netherlands) for providing the magnetic resonance images for the reliability exercises.

REFERENCES

1. Kloppenburg M, Kwok WY. Hand osteoarthritis—a heterogeneous disorder. Nat Rev Rheumatol 2011;8:22-31.
2. Kroon FP, Conaghan PG, Foltz V, Gandjbakhch F, Peterfy C, Eshed I, et al. Development and reliability of the OMERACT thumb base osteoarthritis magnetic resonance imaging scoring system. J Rheumatol 2017;44:1694-8.
3. Boers M, Kirwan JR, Tugwell P, Beaton D, Bingham CO, Conaghan PG, et al. The OMERACT Handbook. [Internet. Accessed November 12, 2018.] Available from: https://omeract.org/resources
4. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, D'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol 2014;67:745-53.
5. Kroon FP, Peterfy CG, Conaghan PG, Foltz V, Gandjbakhch F, Eshed I, et al. Atlas for the OMERACT thumb base osteoarthritis MRI scoring system (TOMS). RMD Open 2018;4:e000583.
6. Damman W, Liu R, Kroon FP, Reijnierse M, Huizinga TW, Rosendaal FR, et al. Do comorbidities play a role in hand osteoarthritis disease burden? Data from the Hand OSTeoArthritis in Secondary care cohort. J Rheumatol 2017;44:1659-66.
7. Kloppenburg M, Peterfy C, Haugen I, Kroon F, Chen S, Wang L, et al. A phase 2a, placebo-controlled, randomized study of ABT-981, an anti-interleukin-1alpha and -1beta dual variable domain immunoglobulin, to treat erosive hand osteoarthritis (EHOA) [abstract]. Ann Rheum Dis 2017;76:122.
8. Müller R, Büttner P. A critical discussion of intraclass correlation coefficients. Stat Med 1994;13:2465-76.
9. Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis 2005;64:179-82.
10. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007;15:A1-56.
11. 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.
12. Lee KM, Lee J, Chung CY, Ahn S, Sung KH, Kim TW, et al. Pitfalls and important issues in testing reliability using intraclass correlation coefficients in orthopaedic research. Clin Orthop Surg 2012; 4:149-55.
13. Botha-Scheepers S, Riyazi N, Watt I, Rosendaal FR, Slagboom E, Bellamy N, et al. Progression of hand osteoarthritis over 2 years: a clinical and radiological follow-up study. Ann Rheum Dis 2009;68:1260-4.