

Iterative Development and Reliability of the OMERACT Hand Osteoarthritis MRI Scoring System

Ida K. Haugen, Mikkel Østergaard, Iris Eshed, Fiona M. McQueen, Paul Bird, Frederique Gandjbakhch, Violaine Foltz, Harry Genant, Charles Peterfy, Siri Lillegraven, Espen A. Haavardsholm, Pernille Bøyesen, and Philip G. Conaghan

ABSTRACT. Objective. To develop and test the interreader reliability of the OMERACT Hand Osteoarthritis Magnetic Resonance Scoring System (HOAMRIS) for assessment of structural and inflammatory hand OA features in the interphalangeal joints.

Methods. The HOAMRIS was developed through an iterative process. Selection of features and their scaling was agreed upon through consensus by members of the OMERACT Magnetic Resonance Imaging (MRI) Task Force, using the Oslo Hand Osteoarthritis (OA) MRI Score system as a template. Two reliability exercises were performed, in which 6 and 4 readers participated, respectively. After the first exercise, an atlas was developed and used in the second exercise to facilitate reading. In each exercise, readers independently scored 8 MRI scans from the Oslo Hand OA cohort (coronal/axial short-tau inversion recovery and coronal/axial/sagittal T1-weighted fat-suppressed pre-/post-Gadolinium images). Interreader reliability was assessed by intraclass correlation coefficients (ICC), percentage exact and close agreement (PEA/PCA).

Results. The preliminary OMERACT HOAMRIS included assessment of synovitis, erosive damage, cysts, osteophytes, cartilage space loss, malalignment, and bone marrow lesions (BML), of which all were scored on a 0–3 scale for normal, mild, moderate, and severe (increments of 0.5 for synovitis, erosive damage, and BML). In the first exercise, most features showed good to very good ICC values (0.64–0.94), except synovitis (0.34). In the second exercise using the atlas, the ICC values were > 0.74 for all MRI features, and the PEA/PCA values were higher than in the first exercise.

Conclusion. A preliminary HOAMRIS with good to very good interreader reliability was developed. Longitudinal studies are needed to assess its sensitivity to change. (First Release Dec 1 2013; J Rheumatol 2014;41:386–71; doi:10.3899/jrheum.131086)

Key Indexing Terms:

HAND OSTEOARTHRITIS
OUTCOMES RESEARCH

MAGNETIC RESONANCE IMAGING
RELIABILITY

Hand osteoarthritis (OA) is a frequent disease in the general population¹ that can lead to considerable pain and

functional limitation². Currently, there is limited knowledge about the disease course in hand OA. Magnetic resonance

From the Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway; Department of Rheumatology, Copenhagen University Hospitals at Glostrup, Copenhagen, Denmark; Department of Diagnostic Imaging, Sheba Medical Center, Tel Aviv University, Tel Aviv, Israel; Department of Molecular Medicine and Pathology, Faculty of Medicine and Health Sciences, University of Auckland, Auckland, New Zealand; Department of Rheumatology, St. George Hospital, Sydney, Australia; St. George Clinical School, University of New South Wales, Sydney, Australia; Department of Rheumatology, Pitié Salpêtrière Hospital, APHP, Université Pierre et Marie Curie, Paris, France; Departments of Radiology and Medicine, University of California San Francisco, San Francisco, CA, USA; Synarc Inc., Newark, California, USA; Spire Sciences Inc., Boca Raton, Florida, USA; Leeds Institute of Rheumatic and Musculoskeletal Medicine USA; Section of Musculoskeletal Disease, University of Leeds and UK National Institute for Health Research (NIHR) Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK. The Oslo hand OA cohort is supported by grants from the South-Eastern Norway Regional Health Authority.

I.K. Haugen, MD, PhD, Postdoctoral Researcher, Department of Rheumatology, Diakonhjemmet Hospital; M. Østergaard, MD, PhD, DMSc, Professor of Rheumatology, Department of Rheumatology, Copenhagen University Hospitals at Glostrup; I. Eshed, MD, Consultant Radiologist, Sheba Medical Center, Tel Aviv University; F.M. McQueen,

MBCB, MD, FRACP, Professor of Rheumatology, Department of Molecular Medicine and Pathology, Faculty of Medicine and Health Sciences, University of Auckland; P. Bird, BMed (Hons), FRACP, PhD, Practising Rheumatologist, Department of Rheumatology, St. George Hospital; and Senior Lecturer, University of New South Wales; F. Gandjbakhch, MD, Practising Rheumatologist, Department of Rheumatology, Pitié Salpêtrière Hospital, APHP, Université Pierre et Marie Curie; V. Foltz, MD, Practising Rheumatologist, Department of Rheumatology, Pitié Salpêtrière Hospital, APHP, Université Pierre et Marie Curie; H.K. Genant, MD, FACR, FRCR, Professor Emeritus of Radiology, Medicine and Orthopaedics, University of California San Francisco, and Synarc Inc.; C.G. Peterfy, MD, PhD, FRCPC, Spire Sciences Inc.; S. Lillegraven, MD, MPH, Research Fellow, Department of Rheumatology, Diakonhjemmet Hospital; E.A. Haavardsholm, MD, PhD, Postdoctoral Researcher, Department of Rheumatology, Diakonhjemmet Hospital; P. Bøyesen, MD, PhD, Postdoctoral Researcher, Department of Rheumatology, Diakonhjemmet Hospital; P.G. Conaghan, MB, BS, PhD, FRACP, FRCPC, Professor of Musculoskeletal Medicine, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds and NIHR Leeds Musculoskeletal Biomedical Research Unit.

Address correspondence to Dr. Haugen, Department of Rheumatology, Diakonhjemmet Hospital, Post Box 23 Vinderen, 0319 Oslo, Norway. E-mail: haugen_ida@hotmail.com

imaging (MRI) has the advantage of whole-joint assessment of all the affected joint tissues including cartilage, bone, and soft tissues. Further, MRI is able to demonstrate bone marrow lesions (BML), which may be a potential target for therapeutic interventions. MRI is therefore a valuable tool to increase the understanding of OA processes, and in future clinical trials may serve as an important outcome measure.

A group from Oslo recently proposed the first MRI scoring system for assessment of hand OA features in the distal interphalangeal (DIP) and proximal interphalangeal (PIP) joints³. This scoring system included evaluation of osteophytes, joint space narrowing, erosions, cysts, BML, malalignment, collateral ligament pathology, synovitis, and flexor tenosynovitis. Despite good to very good reliability, associations between certain MRI features and pain, as well as high sensitivity in detection of OA structural features^{3,4,5}, the authors also noted limitations of the proposed system. First, it was time-consuming because of the inclusion of many features and the separate scoring of the proximal and distal parts of the joint. Further, features such as collateral ligament pathology and flexor tenosynovitis were uncommon, had lower reliability, and were not associated with pain.

MRI outcome measures for both rheumatoid arthritis and psoriatic arthritis were developed by the OMERACT MRI Task Force^{6,7} and validated using the OMERACT filter⁸. The aim of our study was to develop a preliminary OMERACT MRI Scoring System for Hand OA (OMERACT HOAMRIS) using OMERACT methodology.

METHODS

Development of the preliminary OMERACT HOAMRIS. Members of the OMERACT MRI Task Force met for a full day in May, 2011. Using the Oslo MRI scoring system as a template, meeting participants made modifications, by consensus, to the selection, definitions, and grading of its pathological features. We also proposed preferred acquisition plane(s) for assessment of the various features and when MRI gadolinium contrast use was preferable.

Iterative reliability exercises. The first interreader reliability exercise was conducted using 6 readers (IKH, IE, FM, PB, FG, VF). Both rheumatologists (n = 4), rheumatology research fellow (n = 1) and radiologist (n = 1) were represented, and all had previous experience in reading MRI scans in different rheumatic diseases affecting the hands. No training was performed prior to the exercise. The readings were performed on different working stations (11–27 in) using different image analysis systems. Each reader scored MRI scans acquired with a 1.0 T extremity MRI scanner (ONI, GE Healthcare) of the DIP and PIP joints in the dominant hand of 8 patients from the Oslo hand OA cohort. The patient images had been selected for a wide range of radiographic hand OA structural severity based on Kellgren-Lawrence scores. The sequences included short-tau inversion recovery (STIR) images in coronal and axial planes (TE 16.3 and 21 ms, TR 2850 and 3150 ms, slice thickness 2–3 mm, gap between slices 0.2 and 1 mm) and T1-weighted gradient-echo fat-suppressed pre- and post-gadolinium images in coronal, axial, and sagittal planes (TE 5 ms, TR 20 ms, slice thickness 1 mm, gap between slices 0 mm).

After the first exercise, a Web-based meeting was arranged to discuss the initial results. To facilitate the reading and to improve interreader reliability, an atlas of the scoring system was developed. The atlas was

distributed and approved by all readers prior to a second exercise. For standardizing purposes, the axial plane (pre- and post-gadolinium) was chosen as the preferred plane for assessment of synovitis in the second exercise.

In the second reliability exercise, 4 of the 6 readers from the first exercise (IKH, IE, FG, VF) participated, and 8 new patient images from the Oslo hand OA cohort were selected.

The data collection in the Oslo hand OA cohort was approved by the regional ethics committee and the data inspectorate. All patients signed informed consent.

Statistical analysis. Interreader reliability was assessed by calculation of percentage exact agreement (PEA), percentage close agreement (PCA), and average measure intraclass correlation coefficients (AvmICC) using mixed effect models (absolute agreement). All features were scored on 0–3 scales with increments of 0.5 for synovitis, erosions, and BML. Hence, PEA was defined as a difference of 0 or 0.5 between minimum and maximum scores in a single joint among all readers, whereas PCA was defined as a difference < 1 between minimum and maximum scores in a single joint among the readers. For calculation of ICC we used the total scores for all 8 joints. ICC values < 0.20 were considered as poor reliability, 0.20 < ICC < 0.40 as fair, 0.40 < ICC < 0.60 as moderate, 0.60 < ICC < 0.80 as good, and 0.80 < ICC < 1.00 as very good reliability (i.e., same cutoffs as recommended for kappa)⁹. We also calculated the median and interquartile range (IQR) for each of the MRI features based on the reader mean values.

RESULTS

Preliminary OMERACT HOAMRIS. Detailed definitions of MRI features in the HOAMRIS, and their grading, are presented in Table 1. Features such as flexor tenosynovitis and collateral ligament pathology were excluded prior to the reliability exercises. All MRI features were scored on a 0–3 scale for normal, mild, moderate, and severe. The distal and proximal part of the DIP and PIP joints were combined, instead of grading separately as in the Oslo scoring system. Increments of 0.5 were provided for synovitis, erosive damage, and BML. The coronal plane was recommended for evaluation of all MRI features, except synovitis, for which the axial plane was recommended. Both coronal and sagittal planes were used for assessment of osteophytes. For evaluation of erosions, using 1 other plane in addition to the coronal plane was considered ideal but not mandatory. For the majority of features, the T1-weighted images were the preferred sequence to be used for evaluation of pathology, except BML, which was evaluated using STIR images. Postcontrast images were recommended for assessment of synovitis.

Reliability exercises. The demographic and clinical variables for the patients whose images were included in the reliability exercises are presented in Table 2.

In the first reliability exercise (6 readers), we demonstrated good to very good interreader reliability for most of the MRI features, except for a fair AvmICC for synovitis. Close agreement was found in > 62.5% of the joints for all MRI features, whereas the exact agreement was generally low (Table 3).

In the second reliability exercise (4 readers), the readers used an atlas to facilitate scoring and preferably the axial plane for assessment of synovitis. Interreader reliability was

Table 1. The definitions and scaling of features included in the preliminary OMERACT Hand Osteoarthritis Magnetic Resonance Imaging Score.

	Definition	Scaling
Synovitis	Thickened synovium with enhancement of gadolinium	0 = normal; 1 = mild; 2 = moderate; 3 = severe. The 1–3 scores are defined by thirds of the presumed maximum volume of enhancing tissue in the synovial compartment
Erosive damage	Subchondral bone loss, including erosions (sharply marginated bone lesion with cortical break in at least one plane), bone attrition (flattening or depression of the joint plate) and bone resorption	0 = normal; 1 = mild, $\leq 10\%$ of bone volume or $\leq 25\%$ joint surface affected*; 2 moderate, 11–20% of bone volume and/or 26–50% of joint surface affected*; 3 = severe, $> 20\%$ of bone volume and/or $> 50\%$ of joint surface affected*
Cyst	A sharply marginated bone lesion with typical signal characteristics* without a cortical break. * Increased signal intensity on T1w fs, T2w fs or STIR images	0 = normal; 1 = mild, $\leq 10\%$ of bone volume*; 2 = moderate, 11–20% of bone volume*; 3 = severe, $> 20\%$ of bone volume*
Osteophyte	Abnormal bone protuberance at joint margins or surfaces	0 = normal, no osteophytes; 1 = mild, 1–3 small osteophytes; 2 = moderate, ≥ 4 small osteophytes and/or ≥ 1 moderate osteophyte(s); 3 = severe, at least 1 large osteophyte
Cartilage space loss	Loss of the cartilage space based on the inter-bone distance	0 = normal; 1 = mild, loss of cartilage space without bone-to-bone contact; 2 = moderate, focal complete loss of cartilage space; 3 = severe, complete cartilage space loss affecting $> 50\%$ of the articulating joint area
Malalignment	Angulation or subluxation of joints in the frontal plane	0 = no apparent malalignment; 1 = mild, $\leq 10^\circ$ angulation and/or not congruent joint surfaces; 2 = moderate: 11–20° angulation and/or or moderate subluxation without crossing the midline; 3 = severe: $> 20^\circ$ angulation and/or subluxation with crossing the midline
Bone marrow lesions	A lesion within the trabecular bone with signal characteristic consistent with increased water content* and with ill-defined margins. * High signal intensity on STIR or T2w fs images	0 = normal; 1 = mild, 1–33% of bone volume*; 2 = moderate, 34–66% of bone volume*; 3 = severe, 67–100% of bone volume*

* Bone volume and joint surface refer to the proximal and distal part of the joint combined. The assessed bone volume should extend from the articular surface to a depth of 0.5 cm for distal interphalangeal joints and 1.0 cm for proximal interphalangeal joints on either side; # 0.5 increments for the assessment of synovitis, erosions, and bone marrow lesions. T1w: T1-weighted, T2w: T2-weighted, fs: fat-suppressed, STIR: short-tau inversion recovery.

Table 2. Clinical and demographic characteristics of patients with hand osteoarthritis included in the reliability exercises.

Characteristic	Patients in Exercise 1, n = 8	Patients in Exercise 2, n = 8
Women, n (%)	8 (100)	n = 8 (100)
Age, mean (SD) yrs	66.7 (4.7)	68.5 (6.2)
Body mass index, mean (SD) kg/m ² *	25.3 (2.3)	28.9 (5.9)
ACR criteria for hand OA, n (%)	8 (100)	8 (100)
KL sum (DIP/PIP dominant hand), mean (SD) [range 0–32]	15.5 (4.6)	14.8 (10.3)
AUSCAN pain, mean (SD) [0–20]	10.1 (2.5)	6.8 (5.1)
AUSCAN physical, mean (SD) [0–36]	21.5 (6.9)	15.8 (8.5)
Grip strength dominant hand, mean (SD) kg	14.6 (5.5)	20.0 (11.0)

* Body mass index data on 7 patients in each exercise only. Ranges are listed in brackets. ACR: American College of Rheumatology, KL: Kellgren-Lawrence, DIP: distal interphalangeal, PIP: proximal interphalangeal, AUSCAN: Australian/Canadian Hand Index.

generally higher, especially for synovitis, compared to the first reliability exercise (except for BML). All features showed good to very good ICC values, and higher PCA and PEA values compared with the first exercise (Table 4).

We also calculated the ICC values in the first round excluding the 2 readers who did not attend the second exercise. For most features, the ICC values in the first round were lower when examining 4 readers only, making the

difference from the first and second exercise even larger (data not shown).

DISCUSSION

In our study, the OMERACT MRI Task Force developed a preliminary OMERACT HOAMRIS and tested its inter-reader reliability in 2 iterative scoring exercises. The scoring system constituted 6 MRI features, all scored on 0–3 scales

Table 3. Interreader reliability of magnetic resonance imaging features in the first exercise (6 readers).

	Median (IQR)	PEA, %	PCA, %	AvmICC (95% CI)
Synovitis	4 (4, 6)	28.1	70.3	0.34 (−0.14, 0.79)
Erosive damage	4 (2.5, 6)	37.5	65.6	0.83 (0.57, 0.96)
Cysts	2 (1, 4)	31.3	82.8	0.64 (0.21, 0.91)
Osteophytes	10 (5.5, 11)	6.3	67.2	0.88 (0.68, 0.97)
Cartilage space loss	9 (6, 11)	15.6	62.5	0.88 (0.65, 0.97)
Malalignment	3 (2, 4)	39.1	84.4	0.75 (0.39, 0.94)
Bone marrow lesions	2 (1, 3)	43.8	81.3	0.94 (0.84, 0.99)

IQR: interquartile range; PEA: percentage exact agreement; PCA: percentage close agreement; AvmICC: average measure intraclass correlation coefficients.

Table 4. Interreader reliability of magnetic resonance imaging features in the second exercise (4 readers with use of atlas).

	Median (IQR)	PEA, %	PCA, %	AvmICC (95% CI)
Synovitis	5 (3, 8)	35.9	79.7	0.74 (0.30, 0.94)
Erosive damage	5 (1, 11)	46.9	78.1	0.96 (0.90, 0.99)
Cysts	3 (2, 4.5)	54.7	85.9	0.83 (0.52, 0.96)
Osteophytes	10 (6, 16)	17.2	89.1	0.97 (0.90, 0.99)
Cartilage space loss	13 (6, 17)	31.3	85.9	0.98 (0.93, 0.99)
Malalignment	2.5 (2, 4)	57.8	89.1	0.93 (0.79, 0.98)
Bone marrow lesions	2 (1, 4.5)	56.3	85.9	0.79 (0.39, 0.95)

IQR: interquartile range; PEA: percentage exact agreement; PCA: percentage close agreement; AvmICC: average measure intraclass correlation coefficients.

for normal to severe. When the readers used an atlas with examples of images, good to very good reliability was demonstrated for all the MRI features, suggesting that MRI can reliably assess structural and inflammatory features in the DIP and PIP joints in patients with hand OA.

The Oslo hand OA MRI scoring system for hand OA was used as a starting point³, and the experience and results from the validation studies performed in Oslo were taken into account during the development of the HOAMRIS^{4,5}. Haugen, *et al* found that flexor tenosynovitis had only moderate reliability and was not related to OA severity, or associated with pain in the same joint^{3,4}. Concerns were also raised regarding a potential magic angle phenomenon (i.e., increase in signal intensity occurring when collagen fibers are oriented 55 degrees relative to the static magnetic field) for the assessment of collateral ligament discontinuation, which may result in false positive reading of collateral ligament pathology¹⁰. Consequently, we chose to exclude flexor tenosynovitis and collateral ligament pathology from the scoring system.

The definitions of the features were reassessed and modified. In the Oslo system, the definition of erosion was based mainly on loss of bone volume, and subchondral bone attrition (flattening or depression of the joint plate, leading to small loss of bone volume) was scored as grade 1 only³. We felt the severity of bone attrition was not captured by the Oslo system. In the HOAMRIS system, the grading of

erosions was based on both the volume of the erosions and the extent to which the joint surface was affected by bone damage. With this definition, the severity of OA central erosions with bone attrition is expected to be better captured.

When scoring erosions, cysts, and bone marrow lesions, one should assess the total volume from both the distal and proximal side. Hence, 2 erosions estimated at 15% bone loss on both the proximal and distal side (in total 30%) and an erosion estimated at 30% on 1 side should both be scored as grade 3.

Terminology was also changed, and the term for cartilage loss assessment was changed from “joint space narrowing” to “cartilage space loss” to have a more precise wording of the feature. The thickness of the cartilage in the small DIP and PIP joints is difficult to assess directly owing to current image resolution, and the definition was therefore based on the interbone distance.

All features were scored on 0–3 scales with increments of 0.5 for synovitis, erosive damage, and BML. These features were considered to be most important in future clinical trials; thus, an increment of 0.5 was chosen to improve sensitivity to detect more subtle alterations, cross-sectionally and longitudinally. An increment of 0.5 should be used when a reader is uncertain about 2 adjacent categories, e.g., grade 1.5 if uncertain about whether to assign grade 1 or 2. In longitudinal studies or clinical trials,

a 0.5 increment can be used when there is increase/decrease without change of category. The use of 0.5 increments may potentially lead to reduced reliability, and there is therefore a need for further evaluation of this in future reliability studies. Training sessions between the readers and more example images in the atlas may improve the reliability between readers.

Our results clearly show the importance of calibration between readers. Higher reliability was found in the second exercise, in which the readers used an atlas with example images. The atlas that was used in the second exercise was based on images from the first exercise, and did not include example images of all categories for all features. More work is therefore needed to improve and complete the atlas. Only a few validation studies of MRI have been performed in hand OA^{4,5,11}, and they are limited by the lack of a gold standard as reference. MRI can detect more erosions and osteophytes than conventional radiography⁴, but these results need to be confirmed in studies using computed tomography (CT) or histology to show that these MRI features represent true findings. Suboptimal resolution of the MRI also makes the distinction between marginal erosions and cysts as well as cysts and bone marrow lesions difficult, emphasizing the need for validation studies using CT and/or histology. Longitudinal studies will also show the temporal relationship between, for example, bone marrow lesions and cysts, which have been shown to be temporally related in knee OA¹².

In the current reliability exercise we used MRI from patients in the Oslo hand OA cohort³. In this cohort the patients had MRI of their dominant hand. However, we have not added any recommendations for which hand to scan, as future studies may want to use other inclusion criteria, such as the hand with most symptoms or radiographic criteria.

One limitation is that the current scoring system does not include the base of the thumb. Thumb base involvement is important for both pain and function in patients with hand OA¹³. In the Oslo hand OA cohort, only the DIP and PIP joints were scanned, since inclusion of the thumb base joint would have required a separate acquisition. In future studies, we need to examine whether we can use the same definitions for the thumb base joint as for the DIP and PIP joints. Further, the interpretation of the ICC values is dependent on the range of the measuring scale¹⁴; the wider the range, the better the results. The patient variation was larger in the second exercise (i.e., broader IQR) for some of the features. Larger patient variation may lead to larger ICC as an artefact, i.e., patient variance being a larger component of the total variance (the denominator). However, also PEA and PCA agreement was higher in the second exercise, supporting a true improvement of reliability. We also found good reliability for features that were less frequently present, such as cysts, malalignment, and BML. A potential limitation is the use of contrast-enhanced images for evalu-

ation of synovitis in a typically elderly population with patients with OA. Future studies should look into whether MRI without contrast have similar sensitivity and specificity in detection of synovitis in hand OA as contrast-enhanced MRI. In both exercises, we used MRI from 8 patients. However, a greater number of patients would have allowed us to look into a subset of patients to explore how different patients affect the reliability.

The next step in the validation of the preliminary OMERACT hand OA MRI score is to examine the feasibility of the scoring system and whether it is sensitive to change. At this time, the OMERACT MRI Task Force does not have available data from any observational longitudinal cohorts or clinical trials using MRI as an outcome measure. However, we expect that within a few years we will have such data available to explore responsiveness.

The OMERACT MRI Task Force has developed a preliminary hand scoring system, the OMERACT HOAMRIS, and tested its interreader reliability. Our results suggest that MRI can reliably assess OA features when readers use an atlas with example images. However, further validation of MRI features and assessment of sensitivity to change must be tested before MRI can be recommended as an outcome measure in clinical hand OA trials.

ACKNOWLEDGMENT

We thank Barbara Slatkowsky-Christensen for her contribution to the data collection in the Oslo hand OA cohort, and Tore K. Kvien, Sølve Sesseng, and Désirée van der Heijde for their important contributions to the development of the Oslo hand OA MRI scoring system.

REFERENCES

1. Haugen IK, Englund M, Aliabadi P, Niu J, Clancy M, Kvien TK, et al. Prevalence, incidence and progression of hand osteoarthritis in the general population: The Framingham Osteoarthritis Study. *Ann Rheum Dis* 2011;70:1581-6.
2. Slatkowsky-Christensen B, Mowinckel P, Loge JH, Kvien TK. Health-related quality of life in women with symptomatic hand osteoarthritis: a comparison with rheumatoid arthritis patients, healthy controls, and normative data. *Arthritis Rheum* 2007;57:1404-9.
3. 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.
4. Haugen IK, Bøyesen P, Slatkowsky-Christensen B, Sesseng S, Bijsterbosch J, van der Heijde D, et al. Comparison of features by MRI and radiographs of the interphalangeal finger joints in patients with hand osteoarthritis. *Ann Rheum Dis* 2012;71:345-50.
5. Haugen IK, Bøyesen P, Slatkowsky-Christensen B, Sesseng S, Van Der Heijde D, Kvien TK. Associations between MRI-defined synovitis, bone marrow lesions and structural features and measures of pain and physical function in hand osteoarthritis. *Ann Rheum Dis* 2012;71:899-904.
6. Østergaard M, Peterfy C, Conaghan P, McQueen F, Bird P, Ejbjerg B, et al. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system. *J Rheumatol* 2003;30:1385-6.

7. 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.
8. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol* 1998;25:198-9.
9. Altman DG, ed. Some common problems in medical research. *Practical Statistics for Medical Research*. London: Chapman & Hall; 1991:404.
10. Peterfy CG, Janzen DL, Tirman PF, van Dijke CF, Pollack M, Genant HK. "Magic-angle" phenomenon: a cause of increased signal in the normal lateral meniscus on short-TE MR images of the knee. *AJR Am J Roentgenol* 1994;163:149-54.
11. Wittoek R, Jans L, Lambrecht V, Carron P, Verstraete K, Verbruggen G. Reliability and construct validity of ultrasonography of soft tissue and destructive changes in erosive osteoarthritis of the interphalangeal finger joints: a comparison with MRI. *Ann Rheum Dis* 2011;70:278-83.
12. Crema MD, Roemer FW, Zhu Y, Marra MD, Niu J, Zhang Y, et al. Subchondral cystlike lesions develop longitudinally in areas of bone marrow edema-like lesions in patients with or at risk for knee osteoarthritis: detection with MR imaging — the MOST study. *Radiology* 2010;256:855-62.
13. Bijsterbosch J, Visser W, Kroon HM, Stamm T, Meulenbelt I, Huizinga TW, et al. Thumb base involvement in symptomatic hand osteoarthritis is associated with more pain and functional disability. *Ann Rheum Dis* 2010;69:585-7.
14. Müller R, Büttner P. A critical discussion of intraclass correlation coefficients. *Stat Med* 1994;13:2465-76.