
Universiteit
Leiden
The Netherlands

Ultrasonography of inflammatory and structural lesions in hand osteoarthritis: an outcome measures in rheumatology agreement and reliability study

Mathiessen, A.; Hammer, H.B.; Terslev, L.; Kortekaas, M.C.; D'Agostino, M.A.; Haugen, I.K.; ... ; OMERACT Ultrasound Working Grp

Citation
Mathiessen, A., Hammer, H. B., Terslev, L., Kortekaas, M. C., D'Agostino, M. A., Haugen, I. K., ... Ellegaard, K. (2022). Ultrasonography of inflammatory and structural lesions in hand osteoarthritis: an outcome measures in rheumatology agreement and reliability study. *Arthritis Care & Research*, *74*(12), 2005-2012. doi:10.1002/acr.24734

Version:	Publisher's Version
License:	Creative Commons CC BY-NC-ND 4.0 license
Downloaded from:	https://hdl.handle.net/1887/3458829

Note: To cite this publication please use the final published version (if applicable).

Check for updates

Arthritis Care & Research
Vol. 74, No. 12, December 2022, pp 2005–2012
DOI 10.1002/acr.24734
© 2021 The Authors. *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

American College of Rheumatology
Empowering Rheumatology Professionals

Ultrasonography of Inflammatory and Structural Lesions in Hand Osteoarthritis: An Outcome Measures in Rheumatology Agreement and Reliability Study

Alexander Mathiessen,[1] Hilde B. Hammer,[2] Lene Terslev,[3] Marion C. Kortekaas,[4] Maria A. D'Agostino,[5] Ida K. Haugen,[1] George A. Bruyn,[6] Georgios Filippou,[7] Emilio Filippucci,[8] Margreet Kloppenburg,[4] Luana Mancarella,[9] Peter Mandl,[10] Ingrid Möller,[11] Mohamed A. Mortada,[12] Esperanza Naredo,[13] Andrea Delle Sedie,[14] Joseph Sexton,[1] Ruth Wittoek,[15] Annamaria Iagnocco,[16] and Karen Ellegaard,[17] on behalf of the OMERACT Ultrasound Working Group

Objective. To standardize and assess the reliability of ultrasonographic assessment of inflammatory and structural lesions in patients with hand osteoarthritis (OA).

Methods. The Outcome Measures in Rheumatology Ultrasound Working Group selected synovial hypertrophy (SH), joint effusion (JE), and power Doppler (PD) signals as the main inflammatory lesions in hand OA, and suggested osteophytes in the scapho-trapezio-trapezoid (STT) and cartilage defects in the proximal interphalangeal (PIP) joints as novel additions to previous structural scoring systems. A complementary imaging atlas provided detailed examples of the scores. A reliability exercise of static images was performed for the inflammatory features, followed by a patient-based exercise with 6 sonographers testing inflammatory and structural features in 12 hand OA patients. We used Cohen's kappa for intrareader and Light's kappa for interreader reliability for all features except PD, in which prevalence-adjusted bias-adjusted kappa (PABAK) was applied. Percentage agreement was also assessed.

Results. The web-based reliability exercise demonstrated substantial intra- and interreader reliability for all inflammatory features ($\kappa > 0.64$). In the patient-based exercise, intra- and interreader reliability, respectively, varied: SH $\kappa = 0.73$ and 0.45; JE $\kappa = 0.70$ and 0.55; PD PABAK = 0.90 and 0.88; PIP joint cartilage $\kappa = 0.56$ and 0.45; and STT osteophytes $\kappa = 0.62$ and 0.36. Percentage close agreement was high for all features (>85%).

Conclusion. With ultrasound, substantial to excellent intrareader reliability was found for inflammatory features of hand OA. Interreader reliability was moderate, but overall high close agreement between readers suggests that better reliability is achievable after further training. Assessment of osteophytes in the STT joint and cartilage in the PIP joints achieved less reliability and the latter is not endorsed.

INTRODUCTION

Hand osteoarthritis (OA) is a common musculoskeletal disorder causing considerable pain and disability, with still largely unknown etiology (1,2). Inflammation is frequently present in hand OA (3–5), and inflammatory features, as detected by magnetic resonance imaging (MRI) and ultrasound, are associated with clinical features and progression of structural abnormalities

Supported by AbbVie with an unrestricted grant to complete the patient-based exercise. GE provided ultrasound machines for the study.

[1]Alexander Mathiessen, MD, PhD, Ida K. Haugen, MD, PhD, Joseph Sexton, PhD: Diakonhjemmet Hospital, Oslo, Norway; [2]Hilde B. Hammer, MD, PhD: Diakonhjemmet Hospital and University of Oslo, Oslo, Norway; [3]Lene Terslev, MD, PhD: Rigshospitalet, Copenhagen, Denmark; [4]Marion C. Kortekaas, MD, PhD, Margreet Kloppenburg, MD, PhD: Leiden University Medical Center, Leiden, The Netherlands; [5]Maria A. D'Agostino, MD, PhD: APHP, Hôpital Ambroise Paré, Boulogne-Billancourt, France; [6]George A. Bruyn, MD, PhD: MC Groep Hospitals, Lelystad, The Netherlands; [7]Georgios Filippou, MD, PhD: Luigi Sacco University Hospital, Milan, Italy; [8]Emilio Filippucci, MD, PhD: Università Politecnica delle Marche, Jesi (Ancona), Italy; [9]Luana Mancarella, MD: IRCCS Istituto Ortopedico Rizzoli, Bologna, Italy; [10]Peter Mandl, MD, PhD: Medical University of Vienna, Vienna, Austria; [11]Ingrid Möller, MD, PhD: University of Barcelona, Barcelona, Spain; [12]Mohamed A. Mortada, MD, PhD: Zagazig University, Zagazig, Egypt; [13]Esperanza Naredo, MD, PhD: Hospital Universitario Fundación Jiménez Díaz, Madrid, Spain; [14]Andrea Delle Sedie, MD, PhD: Azienda Ospedaliero-Universitaria Pisana, Pisa, Italy; [15]Ruth Wittoek, MD, PhD: Ghent University Hospital, Ghent University, Ghent, Belgium; [16]Annamaria Iagnocco, MD, PhD: Università degli Studi di Torino, Turin, Italy; [17]Karen Ellegaard, MSc, PhD: Copenhagen University Hospital Bispebjerg-Frederiksberg, Frederiksberg, Denmark.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.24734&file=acr24734-sup-0001-Disclosureform.pdf.

Address correspondence to Alexander Mathiessen, MD, PhD, Diakonhjemmet Hospital, Department of Rheumatology, Box 23 Vinderen, Oslo, Norway 0319. Email: alexander_mathiessen@hotmail.com.

Submitted for publication December 23, 2020; accepted in revised form June 15, 2021.

SIGNIFICANCE & INNOVATIONS

- Based on previous work and definitions on ultrasonographic lesions in hand osteoarthritis, we present data from a real-life reliability exercise on synovial hypertrophy, effusion, and power Doppler signals, a new scoring system for cartilage and exploration of osteophyte assessment of the thumb base joint.
- In this exercise, reliability for inflammation was shown to be moderate to excellent, for both intra- and interreader reliability, whereas for ultrasonography structural damage scoring systems, intra- and interreader reliability was fair to moderate.
- The complementary ultrasonographic imaging atlas is expected to enhance unified interpretations of the grading scales between sonographers, departments, and countries.

(6–11). Some OA patients may benefit structurally and clinically from antiinflammatory interventions (12). Moreover, a recent proof-of-concept study showed that 6 weeks of prednisolone treatment improved pain and function in hand OA patients with concurrent joint inflammation (13).

Due to these developments, numerous randomized controlled trials on disease-modifying OA drugs and other treatment strategies for OA will be performed or are already in the pipeline. Ultrasound is feasible, readily available, noninvasive, and inexpensive, and therefore inflammatory and structural ultrasonographic scoring systems could be suitable instruments for such trials (14).

Unfortunately, there is no consensus yet on ultrasonographic scoring systems of most elementary lesions in hand OA. In 2008, a preliminary scoring system was developed and included semiquantitative assessments of the elementary lesions synovitis (grayscale synovial hypertrophy [SH] and joint effusion [JE] combined), power Doppler (PD) signals, and osteophytes (15). However, this scoring system was not further developed, and since then, various modifications and other scoring systems have been used for hand OA research (16,17), making comparison of research outcomes difficult.

The Outcome Measures in Rheumatology (OMERACT) Ultrasound Working Group, subgroup OA, has therefore started developing ultrasonographic scoring systems for structural and inflammatory abnormalities in hand OA. This work has resulted in the definition of the ultrasonographic scoring system for structural damage, comprising the elementary lesions osteophytes and cartilage, as well as the development of an ultrasonography atlas as a reference (18). Subsequent reliability testing showed good reliability of osteophyte semiquantitative scoring, but the reliability of the cartilage semiquantitative scoring system in the metacarpophalangeal (MCP) joints was disappointing, and only a dichotomous scoring could therefore be endorsed. Furthermore, the scaphotrapezio-trapezoid (STT) joint was not included. The STT joint is, however, often affected by OA on radiographs and was therefore included in the recent OMERACT thumb base OA MRI scoring system (19). An association between radiographic OA damage and pain in the thumb base was recently demonstrated, and in contrast to finger OA studies, this association seemed more important in predicting thumb base pain than inflammatory features (20).

The aim of the current study by the OMERACT Ultrasound Working Group was 1) to develop an ultrasonographic scoring system for the inflammatory lesions SH, JE, and PD signals in hand OA, 2) to introduce a novel scoring system for cartilage in the palmar aspect of the proximal interphalangeal (PIP) joints, and 3) to extend the osteophyte scoring system (that has already been defined and found reliable) to also include the STT joint. Finally, the scoring systems were tested in a web-based and patient-based exercise.

PATIENTS AND METHODS

Delphi survey. Based on the literature and already existing ultrasound definitions of OA pathologies, a Delphi survey was carried out to develop scoring systems. These were subsequently tested in web-based and patient-based exercises. The Delphi survey was performed to agree on which elementary lesions to include in the scoring systems and which joints and scans were relevant when examining hand OA with ultrasound. An initial round of questionnaires for the level of agreement according to a Likert scale (1 = strongly disagree to 5 = strongly agree) was distributed to 22 OMERACT participants (subgroup hand OA; participants listed in Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24734). Based on the results and comments, a steering group prepared and distributed a second survey. Each survey was considered valid when ≥10 experts responded, and consensus on each statement was achieved when ≥75% agreed to a score of 4 (agree) or 5 (strongly agree).

Ultrasound imaging atlas. Based on the Delphi survey, a preliminary ultrasound atlas was developed and made available for the web-based exercise, and later edited according to feedback from the experts and used in the patient-based exercise. Anonymized images were collected of hand OA patients in the rheumatologic outpatient clinic at Diakonhjemmet Hospital (Oslo, Norway) and participants of hand OA studies from the Parker Institute (Copenhagen, Denmark) and Leiden University Medical Hospital (The Netherlands).

Web-based exercise. A web-based reliability exercise was performed on the inflammatory features SH, JE, and PD signals using the developed ultrasound atlas. A pool of 99 static and anonymized ultrasound images was selected by the author AM to represent all degrees of pathology. These were distributed to the expert panel and scored semiquantitatively (0–3) for

21514658, 2022, 12, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.24734 by Leiden University Libraries, Wiley Online Library on [09/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

interreader reliability. For intrareader reliability, 40 images were randomly chosen and redistributed 2 weeks after the first round.

Patient-based exercise. A training session for the sonographers was held before the patient-based reliability exercise. Six sonographers and 3 facilitators met in Copenhagen (Denmark) as well as 3 experts who participated through a videoconference. The group agreed to assess SH and JE separately in addition to PD signals, all on a semiquantitative scale (0–3) (Table 1). Osteophytes in the STT joint were deemed feasible to score on a 0–3 scale (Table 1). Finally, the group decided to also include an assessment of cartilage. However, compared to previous work with dorsal and longitudinal scans of the MCP joints (18), the group instead suggested a transverse scan of the palmar aspect of the PIP joints (Table 1). Due to many joints and lesions, we limited the number of joints in the patient-based exercise to the following: 1) inflammation in the 2nd to 5th PIP and distal interphalangeal (DIP) joints of the dominant hand, assessing SH and JE separately in both dorsal and palmar aspect of the joints (grade 0–3) but PD activity on the dorsal side only (grade 0–3); 2) osteophytes in the STT joint bilaterally (grade 0–3), since good reliability was previously demonstrated in the other finger joints (18); and 3) cartilage defects (partial or complete loss of cartilage, or loss of interphase sharpness) on the palmar side of the

Table 1. Ultrasonographic assessment of the elementary lesions in hand osteoarthritis included in the current work*

Ultrasonographic lesion	Joint (projection)	Morphologic description	Scoring
Joint effusion	CMC 1 (radiopalmar), MCP 1–5 (dorsal), IP 1 (dorsal + palmar), PIP 2–5 (dorsal + palmar),† DIP 2–5 (dorsal + palmar)†	Abnormal hypoechoic or anechoic (relative to subdermal fat, but sometimes may be isoechoic or hyperechoic) intraarticular material that is displaceable and compressible, but does not exhibit Doppler signal (28)	0–3, scored relative to the maximal size of effusion that can be seen in the respective joint group (see Supplementary Figures 1 and 2 and Supplementary Appendix B, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24734): 0 = none, 1 = minimal, 2 = moderate, 3 = severe
Synovial hypertrophy	CMC 1 (radiopalmar), MCP 1–5 (dorsal), IP 1 (dorsal + palmar), PIP 2–5 (dorsal + palmar),† DIP 2–5 (dorsal + palmar)†	Abnormal hypoechoic (relative to subdermal fat, but sometimes may be isoechoic or hyperechoic) intraarticular tissue that is nondisplaceable and poorly compressible and which may exhibit Doppler signal (28)	0–3: 0 = none, 1 = minimal (up to the level of the horizontal line connecting bone surfaces of the joint), 2 = moderate (extending beyond joint line but with upper surface concave or flat), 3 = severe (extending beyond joint line but with upper surface convex)
Doppler signals	CMC 1 (radiopalmar), MCP 1–5 (dorsal), IP 1 (dorsal), PIP 2–5 (dorsal),† DIP 2–5 (dorsal)†	Flow signal in the synovium; must be detected within synovial hypertrophy to be considered as a sign of synovitis (26,27)	0–3: 0 = no flow in the synovium 1 = minor (single vessel signals ≥1), 2 = moderate (confluent vessel signals in less than half of the area of the synovium), 3 = major (vessel signals in more than half of the area of the synovium)
Osteophytes	STT (radiopalmar),† CMC 1 (radiopalmar), MCP 1–5 (dorsal), IP 1 (dorsal), PIP 2–5 (dorsal), DIP 2–5 (dorsal)	A clear, step-up cortical prominence at the bony margin that is visible in 2 perpendicular planes (29)	0–3, severity scored relative to the respective joint group (see Supplementary Figures 1 and 2 and Supplementary Appendix B, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24734); proximal and distal margin assessed together, and the largest osteophyte is scored: 0 = none, 1 = minor, 2 = moderate, 3 = major
Cartilage defects	PIP 2–5 (palmar)†	Normal cartilage has a sharp interphase (white band) on its margins perpendicular to the probe; loss of sharpness occurs when cartilage interphase is not visible; complete loss when cartilage cannot be visualized	0–2: 0 = normal cartilage (anechoic structure with visible cartilage interface), 1 = focal or complete thinning of cartilage, or loss of sharpness of at least 1 cartilage margin, 2 = focal or complete loss of cartilage

* CMC = carpometacarpal; DIP = distal interphalangeal; IP = interphalangeal; MCP = metacarpophalangeal; PIP = proximal interphalangeal; STT = scapho-trapezio-trapezoid.
† Joints and projections included in the patient-based reliability exercise.

21514658, 2022, 12, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.24734 by Leiden University Libraries, Wiley Online Library on [09/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2nd to 5th PIP joints bilaterally (grade 0–2), with the fingers fully extended and the probe in a transverse view.

Six experienced sonographers from 5 European countries performed the ultrasound examination on 12 hand OA patients (11 female patients; mean ± SD age 73.8 ± 7.8 years) recruited from the Parker institute (Bispebjerg-Frederiksberg Hospital, Copenhagen, Denmark). The participants fulfilled the American College of Rheumatology clinical criteria of hand OA (21) and inflammatory joint diseases were excluded. Written consent was obtained before the exercise.

Six high-end ultrasound machines (GE Logiq E9) were used, all equipped with 2 multifrequency linear probes operating at a frequency of 15 MHz (for inflammation and osteophytes) and 18 MHz (for cartilage). The same settings (15 MHz probe: GS frequency 15 MHz, GS gain 51, Doppler frequency 7.5 MHz, pulse repetition frequency 0.4 kHz, Doppler gain 18.5; 18 MHz probe: GS frequency 15 MHz, GS gain 45) were used on all units and each sonographer was allowed to modify only the depth and focus. The patients were positioned in separate rooms with their hands resting on a small table close to the ultrasound machines. The assessments were performed on fully extended fingers, but when in doubt during the scoring of synovial thickening and effusion in the PIP joints, the joint could be slightly flexed to identify the extensor complex correctly. The sonographers rotated between the rooms and were given 20 minutes to complete each evaluation. They examined each patient twice with at least a 2-hour interval.

Statistical analysis. Intrareader reliability was calculated by Cohen's kappa with quadratic weighting (22). Interreader reliability was calculated as the average of all possible n(n–1)/2 2-rater Cohen's kappa, i.e., Light's kappa (23). The 95% confidence intervals (95% CIs) were based on patient resampling by bootstrapping. To account for bias through low prevalence (relative probability) and difference in reported frequencies between raters (marginal distribution), we instead calculated prevalence-adjusted bias-adjusted kappa (PABAK) for PD signals (24). All of the kappa coefficients were evaluated using the guideline outlined by Landis and Koch (25), with the following strength of the kappa coefficients: 0.01–0.20 poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1.00 excellent. The percentage of exact agreement (i.e., percentage of observations that obtained the same score) and the percentage of close agreement (i.e., a score difference of ±1) between all possible pairs of raters, as well as prevalence of the observed lesions, were also calculated. Analyses were performed using R: A Language and Environment for Statistical Computing, version 3.4.4.

RESULTS

The Delphi survey was completed by 20 (round 1) and 18 (round 2) participants from a total of 22 invited experts. The first survey included 15 statements for voting (see Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24734), of which 10 items reached consensus. The remaining statements were revised and modified according to the experts' comments and suggestions and proposed again in the second round, with 7 new statements for voting (see Supplementary Appendix A), of which 6 items reached consensus. Three inflammatory features (SH, JE, and PD signals) and 2 structural features (osteophytes and cartilage) were selected as core elements for ultrasound assessment of hand OA. The ultrasound characteristics of these features, except for cartilage defects, were based on previous definitions (26–29). Final ultrasound methodology, morphologic description, and scoring systems of these lesions are summarized in Table 1.

Ultrasound imaging atlas development. An ultrasonographic imaging atlas of elementary lesions was developed, with a comprehensive version available for the reliability exercise (see Supplementary Figures 1 and 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24734) and an extended version for publication (see Supplementary Appendix B, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24734).

Web-based reliability exercise. In total, 13 experts completed the first round and 11 the second round of the web-based reliability exercise of the scoring systems. SH and PD severity were evenly distributed across grades 0 to 3, whereas few joints were assessed as having JE grade 3 (1.5%) (Table 2). The intra- and interreader kappa coefficients were excellent for scoring PD activity (Table 2), and substantial to excellent agreement was

Table 2. Reader agreement in the web-based reliability exercise of inflammatory ultrasound features

	Grade mean prevalence, %					
	0	1	2	3	Intrareader*	Interreader†
Synovial hypertrophy	28.7	28.5	26.3	16.5	0.78 (0.46–0.95)	0.83 (0.77–0.89)
Joint effusion	47.4	35.0	16.1	1.5	0.79 (0.54–0.97)	0.64 (0.50–0.78)
Power Doppler signals	30.2	25.0	25.0	19.8	0.94 (0.85–1.00)	0.86 (0.72–1.00)

* Intrareader agreement according to mean (range) Cohen's kappa with quadratic weighting.
† Interreader agreement according to Light's kappa (95% confidence interval), i.e., mean Cohen's kappa with quadratic weighting between all pairs of readers.

21514658, 2022, 12, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.24734 by Leiden University Libraries, Wiley Online Library on [09/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3. Mean prevalence of observed lesions between the 6 sonographers in the patient-based reliability exercise, including 12 patients*

Feature and joints (projection)	Joints, no.	First scan, grade				Second scan, grade			
		0	1	2	3	0	1	2	3
SH									
PIP + DIP (dorsal)	576	54.9	31.6	10.6	3.0	56.1	30.6	10.9	2.4
PIP + DIP (palmar)	576	66.8	23.3	8.7	1.2	65.0	26.8	6.8	1.4
JE									
PIP + DIP (dorsal)	576	59.5	30.9	8.0	1.6	61.4	28.3	8.5	1.7
PIP + DIP (palmar)	576	69.7	20.2	8.3	1.7	68.9	22.7	6.8	1.6
PD: PIP + DIP (dorsal)	576	94.1	3.8	1.9	0.2	93.6	4.9	1.4	0.2
Cartilage defects: PIP (palmar)	288	18.8	43.9	37.2	NA	17.8	43.0	39.2	NA
Osteophytes: STT (palmar)	144	38.9	35.4	18.8	6.9	29.2	45.1	19.4	6.3

* Values are the mean prevalence percentage unless indicated otherwise. DIP = distal interphalangeal; JE = joint effusion; NA = not methodologically applicable; PD = power Doppler; PIP = proximal interphalangeal; SH = synovial hypertrophy; STT = scapho-trapezio-trapezoid.

demonstrated for SH and JE scoring on 0–3 semiquantitative scales (Table 2).

Patient-based reliability exercise. All ultrasound features were present across the whole spectrum of severity (Table 3). However, for all features except cartilage, there was a low prevalence of the highest score. SH and JE were frequently observed in the interphalangeal joints (Table 3), but both features were slightly more prevalent in the dorsal (approximately 45% and approximately 40%, respectively) than the palmar aspects (approximately 34% and approximately 30%, respectively) of the PIP and DIP joints.

The mean kappa coefficients for intrareader reproducibility were substantial for both SH (κ = 0.73) and JE (κ = 0.70) on the dorsal side of the interphalangeal joints (Table 4), whereas intrareader agreement on the palmar side was only moderate (κ = 0.56 and 0.54, respectively). All 6 readers achieved substantial or excellent intrareader agreement on dorsal SH assessment, and 4 readers achieved the same for dorsal JE (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24734).

The interreader agreement was more variable. The percentage of close agreement between all possible reader pairs was very high for both features (>92%), but the percentage of exact agreement was lower (52–64%). The kappa coefficients were better on the dorsal than the palmar side for both SH and JE, but interreader agreement was moderate at best, with κ = 0.45–0.55 on dorsal scans (Table 4).

In contrast to the more common greyscale inflammatory features, PD signals were only reported in 6% of the interphalangeal joints (Table 3). This finding explains the striking improvement in reliability, from low kappa coefficients to high PABAK that was only seen for PD signals and not the other features; intra- and interreader agreement were excellent (0.88–0.90) in the prevalence-adjusted PABAK analyses (Table 4). The percentage of exact agreement between readers was noticeably higher for PD (91%) than the other features.

Table 4. Reader agreement in the patient-based reliability exercise*

Feature and joints (projection)	Intrareader agreement†	Interreader agreement, first scan‡			Interreader agreement, second scan‡		
		Kappa (95% CI)	PEA, %	PCA, %	Kappa (95% CI)	PEA, %	PCA, %
SH							
PIP + DIP (dorsal)	0.73 (0.64–0.83)§	0.45 (0.33–0.57)	53	94	0.45 (0.33–0.57)	52	94
PIP + DIP (palmar)	0.56 (0.48–0.69)	0.31 (0.17–0.45)	59	92	0.35 (0.23–0.47)	57	94
JE							
PIP + DIP (dorsal)	0.70 (0.55–0.81)§	0.52 (0.40–0.64)	63	96	0.55 (0.43–0.67)	64	96
PIP + DIP (palmar)	0.54 (0.36–0.76)	0.33 (0.23–0.43)	63	92	0.31 (0.21–0.41)	62	93
PD: PIP + DIP (dorsal)	0.90 (0.75–0.96)¶	0.88 (0.82–0.94)¶	91	98	0.88 (0.80–0.96)¶	91	99
Cartilage defects: PIP (palmar)	0.56 (0.42–0.81)	0.44 (0.34–0.54)	53	NA	0.45 (0.35–0.55)	56	NA
OP: STT (palmar)	0.62 (0.37–0.80)§	0.36 (0.20–0.52)	44	85	0.27 (0.09–0.45)	36	87

* DIP = distal interphalangeal; JE = joint effusion; NA = not methodologically applicable; OP = osteophytes; PABAK = prevalence-adjusted bias-adjusted kappa; PCA = percentage close agreement, i.e., ±1 grade; PEA = percentage of exact agreement; PD = power Doppler; PIP = proximal interphalangeal; SH = synovial hypertrophy; STT = scapho-trapezio-trapezoid.
† Intrareader agreement according to Cohen's kappa (range), with quadratic weighting or PABAK.
‡ Interreader agreement according to Light's kappa (95% confidence interval), i.e., mean Cohen's kappa (with quadratic weighting) or PABAK between all pairs of readers and percentage of exact and close (±1 grade) agreement between all readers.
§ Substantial to excellent agreement.
¶ PD kappa reported as PABAK, with substantial to excellent agreement.

21514658, 2022, 12, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.24734 by Leiden University Libraries, Wiley Online Library on [09/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Morphologic cartilage abnormalities were found in 82% of the palmar aspect of the PIP joints according to the newly proposed 0–2 scoring system (Table 3). Both intra- and interreader agreement were moderate according to kappa coefficients (Table 4), but 2 readers achieved substantial and excellent intrareader reproducibility (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24734).

The largest discrepancy between the first and second reading among all the features was observed for the lower spectrum of osteophytes in the STT joints, in which many readers shifted their assessment from grade 0 to grade 1 (Table 3). Still, intrareader reproducibility was fairly good, with a mean coefficient of $\kappa = 0.62$. However, interreader reliability was quite low, with coefficients of $\kappa = 0.36$ and the percentages of exact agreement of 44% at best (Table 4), but percentage of close agreements between all reader pairs were satisfactory (85–87%).

DISCUSSION

This study has developed and assessed the first consensus-based scoring systems for inflammatory and structural lesions with ultrasonography of hand OA using the OMERACT methodology (30). Based on previous ultrasonographic definitions of elementary lesions, we scored the inflammatory abnormalities SH and JE separately in addition to PD signals. We also introduced a novel scoring system of cartilage in the PIP joints and added the STT joint to the osteophyte scoring system that has already been defined and found reliable (18). Semiquantitative ultrasonographic assessment of hand OA as an outcome measure is a target area for OMERACT, as numerous trials on potential structure-modifying or cartilage-protective treatments and other management strategies for OA are anticipated. Our scoring systems can be instruments for domains determined by the OMERACT hand OA working group to be assessed in clinical trials of hand OA (14,31). Finally, the complementary ultrasonographic imaging atlas is expected to enhance unified interpretations of the grading scales between sonographers, departments, and countries.

The inflammatory lesions were tested in a web-based exercise with substantial to excellent agreement. This result may be because images selected for web-based exercises are of high quality and do not include acquisition (32). In contrast to a preliminary ultrasonographic scoring system for hand OA that combined SH and JE into 1 greyscale synovitis score (15), we demonstrated that SH and JE can be scored separately on 0–3 scales. This finding is similar to recent OMERACT studies of rheumatoid arthritis (RA) of the hand (33) and OA of the foot (34), except that those studies scored JE as absent/present (0–1). In our study, both features were tested in the patient-based exercise and reached substantial intrareader agreement in dorsal scans of the interphalangeal joints, and moderate agreement between readers. The percentage of exact agreement ranged 52–64% and was higher than previously demonstrated for other inflammatory features in the Oslo Hand OA MRI score: synovitis (46%) and flexor tenosynovitis (36%) (35). The benefit of an ultrasonographic semiquantitative JE score in contrast to a binary score needs further exploration, but all grades were present in the current exercise and evidently possible to score. The role of effusion in hand OA is yet not clear, and it might not be the same as in RA, where effusion has been shown to have little relevance for the disease (36). A high prevalence of JE also exists in healthy subjects (37). The current semiquantitative score may be more helpful in elucidating this role in hand OA and other diseases compared to binary scores alone.

In general, some variation in intrareader reliability and a significant discrepancy between intra- and interreader reliability is probably due to the initial difficulty applying new definitions in a real-life scanning. We applied a free longitudinal scan for more accurate detection of the real amount of inflammation, although standardized alignment of the probe in the midline has been found to improve reliability when assessing small joints (38,39) and may be applied in future studies for reliability purposes. Furthermore, better reliability and higher frequency of SH and JE were demonstrated for dorsal scans of the interphalangeal joints compared to palmar scans. This finding, and the additional time required of a palmar scan, favors a dorsal ultrasound approach to finger joints in OA, similar to RA (38). However, large osteophytes may dominate the dorsal joint aspect, and future study protocols may opt for a palmar scan of SH and JE if in doubt, and only report the highest dorsal or palmar score.

Similar to previous studies on hand and foot OA (9,34), we demonstrated excellent reliability for assessment of PD activity on still images and we reproduced this reliability in the patient-based exercise after adjusting for low prevalence and bias. PABAK was used since PD activity was rarely seen, and a low prevalence may give misleadingly low Cohen's kappa values (34,40). At the same time, the high percentage of exact agreement and the percentage of close agreement should be interpreted with caution due to a high number of joints with absent PD signals.

A previous attempt to develop a semiquantitative 0–3 scoring system for cartilage in hand OA found moderate intrareader and only fair interreader agreement (18), and because the proposed definitions could not help to sufficiently discriminate between intermediary grades, a 0–2 score was suggested as more suitable. Another recent study on cartilage in RA patients simplified the scoring to a 0–2 scale and found moderate to excellent reliability in the MCP joints but only poor reliability for the PIP joints (41). We opted for a 0–2 semiquantitative scoring system based on the morphologic integrity of the superficial interphase of the cartilage and the cartilage thickness. We also changed to a palmar scan of cartilage since osteophytes may cover the dorsal joint space and chose the PIP instead of the MCP joints due to higher prevalence and incidence of OA (42). With this approach,

21514658, 2022, 12, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.24734 by Leiden University Libraries, Wiley Online Library on [09/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

moderate to excellent intrareader reliability but only moderate interreader reliability was found. As with previous attempts, the current study suggests that there are technical and interpretational pitfalls of a semiquantitative cartilage assessment that we have yet not overcome, and the current scoring system is not endorsed.

In the current study, we explored assessment of osteophytes in the STT joint as a supplement to the OMERACT osteophyte scoring system for hand OA that has already been defined and found reliable (18). Encouragingly, we found substantial intrareader reproducibility, although with a larger variation than previously found for other hand joints (18), and only fair to moderate interreader agreement. The divergent prevalence between the first and second round for grades 0 and 1 indicates the difficulty in assessing this joint, and the group recommend both probe position and image interpretation as areas of improvement.

To complete the current real-life reliability study, relevant joints were omitted from the exercise, especially inflammatory features of the carpometacarpal, MCP, and DIP joints. These joints should be examined for domains reflecting structural change and inflammation in future studies (15). Bone erosions were also omitted. Imaging studies applying MRI and ultrasound have found erosive changes in the majority of patients with hand OA, including those without signs of erosions at conventional radiographs (9,43–46). However, scoring of centrally located erosions in hand OA with ultrasound is difficult due to osteophytes that limit the acoustic window (15). The only proposed scoring system for ultrasound-detected bone erosions demonstrated erosions more frequently, but not specifically, for RA compared to OA, psoriatic arthritis, gout, or healthy controls (47). We propose a systematic literature review on ultrasonography of erosions in OA, followed by a discussion of whether this topic should be a focus area for future work.

In conclusion, OMERACT consensus-based semiquantitative scoring systems for SH, JE, and PD activity in hand OA were developed using a complementary ultrasonographic imaging atlas with detailed examples of all scores. We found moderate to substantial agreement for SH and JE as well as excellent PABAK for PD activity, supporting scoring of inflammatory pathologies with US in hand OA. Osteophyte assessment in the STT joints achieved fair to substantial agreement, whereas cartilage assessment of the palmar PIP joints was only moderately reproducible and is therefore not endorsed.

ACKNOWLEDGMENTS

We thank the participants of the patient-based exercise for their time in this project and The Parker Institute (Copenhagen University Hospital Bispebjerg-Frederiksberg, Frederiksberg, Denmark) for facilitating the exercise. We would also like to thank Drs. Frederique Gandjbakhch, Helen Keen, and Marianna Vlychou for participating in the Delphi exercise.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Mathiessen had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Mathiessen, Hammer, Terslev, Kortekaas, D'Agostino, Haugen, Bruyn, Iagnocco, Ellegaard.

Acquisition of data. Mathiessen, Hammer, Terslev, Kortekaas, D'Agostino, Bruyn, Filippou, Filippucci, Kloppenburg, Mancarella, Mandl, Möller, Mortada, Naredo, Delle Sedie, Wittoek, Iagnocco, Ellegaard.

Analysis and interpretation of data. Mathiessen, Hammer, Terslev, Kortekaas, Sexton, Ellegaard.

REFERENCES

1. Kloppenburg M, Kwok WY. Hand osteoarthritis: a heterogeneous disorder. Nat Rev Rheumatol 2011;8:22–31.
2. Zhang Y, Jordan JM. Epidemiology of osteoarthritis. Rheum Dis Clin North Am 2008;34:515–29.
3. Keen HI, Wakefield RJ, Grainger AJ, Hensor EM, Emery P, Conaghan PG. An ultrasonographic study of osteoarthritis of the hand: synovitis and its relationship to structural pathology and symptoms. Arthritis Rheum 2008;59:1756–63.
4. Vlychou M, Koutroumpas A, Malizos K, Sakkas LI. Ultrasonographic evidence of inflammation is frequent in hands of patients with erosive osteoarthritis. Osteoarthritis Cartilage 2009;17:1283–7.
5. Wittoek R, Carron P, Verbruggen G. Structural and inflammatory sonographic findings in erosive and non-erosive osteoarthritis of the interphalangeal finger joints. Ann Rheum Dis 2010;69:2173–6.
6. Kortekaas MC, Kwok WY, Reijnierse M, Watt I, Huizinga TW, Kloppenburg M. Pain in hand osteoarthritis is associated with inflammation: the value of ultrasound. Ann Rheum Dis 2010;69:1367–9.
7. Haugen IK, Slatkowsky-Christensen B, Boyesen P, Sesseng S, van der Heijde D, Kvien TK. Increasing synovitis and bone marrow lesions are associated with incident joint tenderness in hand osteoarthritis. Ann Rheum Dis 2016;75:702–8.
8. Haugen IK, Boyesen 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.
9. Mancarella L, Addimanda O, Pelotti P, Pignotti E, Pulsatelli L, Meliconi R. Ultrasound detected inflammation is associated with the development of new bone erosions in hand osteoarthritis: a longitudinal study over 3.9 years. Osteoarthritis Cartilage 2015;23:1925–32.
10. Mathiessen A, Slatkowsky-Christensen B, Kvien TK, Hammer HB, Haugen IK. Ultrasound-detected inflammation predicts radiographic progression in hand osteoarthritis after 5 years. Ann Rheum Dis 2016;75:825–30.
11. Kortekaas MC, Kwok WY, Reijnierse M, Kloppenburg M. Inflammatory ultrasound features show independent associations with progression of structural damage after over 2 years of follow-up in patients with hand osteoarthritis. Ann Rheum Dis 2015;74:1720–4.
12. Conaghan PG, Cook AD, Hamilton JA, Tak PP. Therapeutic options for targeting inflammatory osteoarthritis pain. Nat Rev Rheumatol 2019;15:355–63.
13. Kroon FP, Kortekaas MC, Boonen A, Böhringer S, Reijnierse M, Rosendaal FR, et al. Results of a 6-week treatment with 10 mg prednisolone in patients with hand osteoarthritis (HOPE): a double-blind, randomised, placebo-controlled trial. Lancet 2019;394:1993–2001.
14. Kloppenburg M, Boyesen P, Visser AW, Haugen IK, Boers M, Boonen A, et al. Report from the OMERACT Hand Osteoarthritis

21514658, 2022, 12, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.24734 by Leiden University Libraries, Wiley Online Library on [09/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Working Group: set of core domains and preliminary set of instruments for use in clinical trials and observational studies. J Rheumatol 2015;42:2190–7.

15. Keen HI, Lavie F, Wakefield RJ, D'Agostino MA, Hammer HB, Hensor E, et al. The development of a preliminary ultrasonographic scoring system for features of hand osteoarthritis. Ann Rheum Dis 2008;67:651–5.
16. Oo WM, Linklater JM, Daniel M, Saarakkala S, Samuels J, Conaghan PG, et al. Clinimetrics of ultrasound pathologies in osteoarthritis: systematic literature review and meta-analysis. Osteoarthritis Cartilage 2018;26:601–11.
17. Saltzherr MS, Selles RW, Bierma-Zeinstra SM, Muradin GS, Coert JH, van Neck JW, et al. Metric properties of advanced imaging methods in osteoarthritis of the hand: a systematic review. Ann Rheum Dis 2014; 73:365–75.
18. Hammer HB, Iagnocco A, Mathiessen A, Filippucci E, Gandjbakhch F, Kortekaas MC, et al. Global ultrasound assessment of structural lesions in osteoarthritis: a reliability study by the OMERACT ultrasonography group on scoring cartilage and osteophytes in finger joints. Ann Rheum Dis 2016;75:402–7.
19. 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.
20. Kroon F, van Beest S, Ermurat S, Kortekaas M, Bloem J, Reijnierse M, et al. In thumb base osteoarthritis structural damage is more strongly associated with pain than synovitis. Osteoarthritis Cartilage 2018;26: 1196–202.
21. Altman R, Alarcón G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hand. Arthritis Rheum 1990;33:1601–10.
22. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.
23. Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. Psychol Bull 1971;76:365.
24. Mak HK, Yau KK, Chan BP. Prevalence-adjusted bias-adjusted kappa values as additional indicators to measure observer agreement. Radiology 2004;232:302–3.
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
26. Koski JM, Saarakkala S, Helle M, Hakulinen U, Heikkinen JO, Hermunen H. Power Doppler ultrasonography and synovitis: correlating ultrasound imaging with histopathological findings and evaluating the performance of ultrasound equipments. Ann Rheum Dis 2006; 65:1590–5.
27. D'Agostino MA, Terslev L, Aegerter P, Backhaus M, Balint P, Bruyn GA, et al. Scoring ultrasound synovitis in rheumatoid arthritis: a EULAR-OMERACT ultrasound taskforce. Part 1: definition and development of a standardised, consensus-based scoring system. RMD Open 2017;3:e000428.
28. Wakefield RJ, Balint PV, Szkudlarek M, Filippucci E, Backhaus M, D'Agostino MA, et al. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. J Rheumatol 2005;32:2485–7.
29. Bruyn GA, Iagnocco A, Naredo E, Balint PV, Gutierrez M, Hammer HB, et al. OMERACT definitions for ultrasonographic pathologies and elementary lesions of rheumatic disorders 15 years on. J Rheumatol 2019;46:1388–93.
30. Terslev L, Naredo E, Keen HI, Bruyn GA, Iagnocco A, Wakefield RJ, et al. The OMERACT stepwise approach to select and develop imaging outcome measurement instruments: the musculoskeletal ultrasound example. J Rheumatol 2019;46:1394–400.
31. Maheu E, Altman RD, Bloch DA, Doherty M, Hochberg M, Mannoni A, et al. Design and conduct of clinical trials in patients with osteoarthritis of the hand: recommendations from a task force of the Osteoarthritis Research Society International. Osteoarthritis Cartilage 2006;14: 303–22.
32. Terslev L, Gutierrez M, Christensen R, Balint PV, Bruyn GA, Delle Sedie A, et al. Assessing elementary lesions in gout by ultrasound: results of an OMERACT patient-based agreement and reliability exercise. J Rheumatol 2015;42:2149–54.
33. Terslev L, Iagnocco A, Bruyn GA, Naredo E, Vojinovic J, Collado P, et al. The OMERACT Ultrasound Group: a report from the OMERACT 2016 meeting and perspectives. J Rheumatol. 2017;44:1740–3.
34. Zabotti A, Filippou G, Canzoni M, Adinolfi A, Picerno V, Carrara G, et al. OMERACT agreement and reliability study of ultrasonographic elementary lesions in osteoarthritis of the foot. RMD Open 2019;5: e000795.
35. 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.
36. Aletaha D, Smolen JS. Diagnosis and management of rheumatoid arthritis: a review. JAMA 2018;320:1360–72.
37. Padovano I, Costantino F, Breban M, D'Agostino MA. Prevalence of ultrasound synovial inflammatory findings in healthy subjects. Ann Rheum Dis 2016;75:1819–23.
38. Witt MN, Mueller F, Weinert P, Nigg AP, Reindl CS, Proft F, et al. Ultrasound of synovitis in rheumatoid arthritis: advantages of the dorsal over the palmar approach to finger joints. J Rheumatol 2014;41: 422–8.
39. Terslev L, Naredo E, Aegerter P, Wakefield RJ, Backhaus M, Balint P, et al. Scoring ultrasound synovitis in rheumatoid arthritis: a EULAR-OMERACT ultrasound taskforce. Part 2: reliability and application to multiple joints of a standardised consensus-based scoring system. RMD Open 2017;3:e000427.
40. Cheung PP, Kong KO, Chew LC, Chia FL, Law WG, Lian TY, et al. Achieving consensus in ultrasonography synovitis scoring in rheumatoid arthritis. Int J Rheum Dis 2014;17:776–81.
41. Mandl P, Studenic P, Filippucci E, Bachta A, Backhaus M, Bong D, et al. Development of semiquantitative ultrasound scoring system to assess cartilage in rheumatoid arthritis. Rheumatology (Oxford) 2019;58:1802–11.
42. 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.
43. Punzi L, Frigato M, Frallonardo P, Ramonda R. Inflammatory osteoarthritis of the hand. Best Pract Res Clin Rheumatol 2010;24:301–12.
44. Haugen IK, Boyesen P. Imaging modalities in hand osteoarthritis and perspectives of conventional radiography, magnetic resonance imaging, and ultrasonography. Arthritis Res Ther 2011;13:248.
45. 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.
46. Vlychou M, Koutroumpas A, Alexiou I, Fezoulidis I, Sakkas LI. High-resolution ultrasonography and 3.0 T magnetic resonance imaging in erosive and nodal hand osteoarthritis: high frequency of erosions in nodal osteoarthritis. Clin Rheumatol 2013;32:755–62.
47. Zayat AS, Ellegaard K, Conaghan PG, Terslev L, Hensor EM, Freeston JE, et al. The specificity of ultrasound-detected bone erosions for rheumatoid arthritis. Ann Rheum Dis 2015;74:897–903.

21514658, 2022, 12, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.24734 by Leiden University Libraries, Wiley Online Library on [09/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License