






ORIGINAL ARTICLE

OMERACT agreement and reliability study of ultrasonographic elementary lesions in osteoarthritis of the foot

Alen Zabotti ¹, Georgios Filippou,² Marco Canzoni,³ Antonella Adinolfi,⁴ Valentina Picerno,⁵ Greta Carrara,⁶ Peter Balint,⁷ George A Bruyn,⁸ Maria Antonietta D'Agostino ⁹, Nemanja Damjanov,¹⁰ Andrea Delle Sedie,¹¹ Emilio Filippucci,¹² Maria Luz Gonzalez Fernandez,¹³ Hilde Berner Hammer,¹⁴ Zunaid Karim,¹⁵ Peter Mandl ¹⁶, Ingrid Moller,¹⁷ Maria Rosario Morales Lozano,¹⁸ Esperanza Naredo,¹⁹ Francesco Porta,²⁰ Garifallia Sakellariou ²¹, Lene Terslev,²² Carlo Alberto Scirè ^{2,6}, Annamaria Iagnocco,²³ The OMERACT Ultrasound Task Force members

To cite: Zabotti A, Filippou G, Canzoni M, *et al*. OMERACT agreement and reliability study of ultrasonographic elementary lesions in osteoarthritis of the foot. *RMD Open* 2019;5:e000795. doi:10.1136/rmdopen-2018-000795

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/rmdopen-2018-000795>).

AZ, GF and MC contributed equally.

Received 31 July 2018
Revised 22 January 2019
Accepted 8 February 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Professor Annamaria Iagnocco; annamaria.iagnocco1@gmail.com

ABSTRACT

Objective To evaluate the level of agreement on ultrasonographic (US) lesions among highly experienced sonographers as well as the intraobserver and interobserver reliability of inflammatory and structural US lesions in patients with osteoarthritis (OA) of the foot.

Methods After a systematic literature review, a Delphi survey was performed to test definitions of US lesions in OA of the foot, including inflammatory lesions (ie, synovial hypertrophy [SH], joint effusion [JE], power Doppler signal [PD]), and structural abnormalities (ie, cartilage damage [CD] and osteophytes). Subsequently, the reliability of US in assessing the aforementioned lesions was tested on static images as well as during a live exercise. Reliability was assessed by kappa analyses and prevalence-adjusted bias-adjusted kappa (PABAK) on a dichotomous and an ordinal scale.

Results Intraobserver and interobserver reliability for SH and JE evaluated by binary scoring was good for both components, while the intraobserver reliability for semiquantitative scoring of SH ranged from moderate in the web-based exercise (PABAK 0.49) to good (PABAK 0.8) in the live exercise. Reliability for CD and PD assessments were respectively good and excellent in all exercises (ranged from PABAK 0.61 to 0.79 for CD and 0.88 to 0.95 for PD). The interobserver reliability for the semiquantitative scoring of osteophytes was fair in the live exercise (PABAK 0.36) and moderate in the static exercise (PABAK 0.60).

Conclusions Consensual US definitions were found to be reliable for assessing inflammatory lesions in OA of the foot, while the use of US to assess structural damage requires further studies.

INTRODUCTION

Osteoarthritis (OA) is a degenerative joint disease characterised by cartilage breakdown, growth of osteophytes and subsequent

Key messages**What is already known about this subject?**

- Foot is a target area in osteoarthritis (OA) and it could significantly impact on patients' quality of life.
- EULAR recommendations on the use of imaging in OA have highlighted that imaging studies in foot are scarce.

What does this study add?

- The study demonstrated that ultrasonography may be a reliable tool for assessing inflammatory lesions in OA of the foot.
- Ultrasonography seems to be a promising tool to be further tested in diagnostic, prognostic and follow-up studies.

low-grade inflammation of the synovial membrane.¹ OA is common in the middle-aged to elderly population and may lead to significant disability and pain. The cornerstones of the therapy of OA are symptomatic treatment as well as measures aimed at preserving physical function. However, recent therapeutic developments which address specific molecular pathways may change the way OA is treated in the future.^{2,3} For this reason, a renewed interest on valid tools assessing disease activity and damage in OA has emerged, with several imaging techniques identified as potential candidates to monitor the impact of new treatments.⁴ In this context, there is a growing interest in the use of ultrasonography (US) for the assessment of OA, as US findings are in good agreement with conventional radiography in detecting typical elementary lesions of OA (eg, central joint

erosions, osteophytes).^{5,6} The foot is recognised as a target region for OA, and this involvement could significantly impact on patients' quality of life.⁷ Despite extensive literature on the use of imaging in OA, it has been shown that only a minority of studies focused on the foot, and this applies to US also.⁵ Therefore, foot and ankle imaging studies should be prioritised, favouring patient management.⁵ The validation of US as an outcome measure for evaluating foot OA is an area of interest for the Outcome Measures in Rheumatology (OMERACT) Ultrasound Group. Exploring the reliability of inflammatory elementary lesions (eg, effusion, synovial hypertrophy) and of structural changes (eg, cartilage abnormalities and osteophytes) is an essential step to include US in trials and clinical practice. For this purpose, the OA task force of the OMERACT US group decided to evaluate the level of agreement among highly experienced sonographers as well as the intraobserver and interobserver reliability of US on inflammatory and structural US lesions in patients with OA of the foot.

MATERIALS AND METHODS

Design of the study

Following the OMERACT methodology,⁸ a systematic literature review (SLR) on US in OA of the foot was performed. Based on these results, a Delphi survey on the definition and characteristics of US lesions in patients with OA was circulated among a group of experts in the field of US and OA selected from the OMERACT special interest group on US. Subsequently, a web-based as well as a patient-based exercise was performed, with the aim of testing the reliability of US in the detection of inflammatory and structural US lesions. Before starting the exercise on patients, a training session was performed on US images of OA abnormalities in the foot and discussions among the experts participating in the meeting took place. The methods and results in our manuscript follows previously published guidelines.⁹ The study was reported to the local ethics committee and no further approval has been deemed necessary.

Systematic literature review

A systematic literature review was performed by one of the authors (GS). MEDLINE via PubMed and Embase were searched from inception to 31 January 2016. Eligible studies had to involve patients with foot (midfoot or metatarsophalangeal [MTP] joints) OA and undergoing US; possible comparators were other imaging techniques or histology. The outcome of interest was the definition of pathology in both greyscale (GS) and power Doppler (PD). All study types excluding narrative reviews were eligible. Search strategies including terms addressing OA and US were applied in both databases (Table S1); prespecified forms were used for data extraction. After screening the title and abstract of 83 studies and the full-text of 11 studies (online supplementary figure S1), 4 studies were finally included (online supplementary

figure S1 and table S2). The hand search of the references of the included studies did not lead to further inclusions.

Delphi

A preliminary questionnaire was circulated to present the results of the SLR to all participants and to collect their comments and suggestions on the items to be included in the Delphi survey. In the first round, the Delphi survey consisted of 15 statements and 19 participants rated their level of agreement for each according to a Likert scale (1=strongly disagree to 5=strongly agree) and gave their comments. Based on the results and comments obtained, the survey was modified and proposed again to the participants until agreement was reached. Group agreement was considered achieved with a total cumulative agreement of 75% or more (a score of 4 or 5 in the Likert scale). Statements that did not reach this cut-off were eliminated from subsequent rounds while statements that achieved agreement were proposed again for voting only in the case of the presence of new statements that were formulated according to the panel's suggestions. If no statement achieved 75% of agreement, those that reached 60% or more, plus new statements were proposed again for voting to avoid missing value in the definitions.

Web-based exercise

A pool of 110 US images (from 83 patients with OA) of the anatomical sites under examination were collected from a personal database of three collaborators (FF, IR, CS) who did not participate in the exercise. Images from patients with foot OA and healthy controls were chosen in order to have both images of normal and abnormal joints. A total of 20 experts were invited to participate in the exercise and each of them rated the images according to the definitions approved in the Delphi survey. The whole Delphi process and the web-based agreement exercise were carried out on a web-based platform (RedCap).¹⁰ Only the facilitator and the epidemiologists of the study had access to the online data and were responsible for the upload and preparation of the Delphi rounds and the web-based exercise.

Training session

Prior to the patient-based reliability exercise, the US methodology was clarified among sonographers and a consensus was obtained on both the scanning protocol and on image interpretation of normal and pathological US findings.

Patient-based exercise

The patient-based exercise was performed on 12 patients (10 female and 2 male, mean age of 67.75 years) recruited if they reported foot pain on weight-bearing and had a diagnosis of OA of the foot based on clinical examination and on the presence of radiographic criteria of OA in at least one foot joint.¹¹ Patients were located in a comfortable examining room and they were lying on an examination bed. The single seats were placed at a

distance that permitted a blinded and separate evaluation by the sonographers, each of whom was seated in front of a single patient. The time frame between the two rounds was 3 hours (first round in the morning and second round in the afternoon of the same day). Twelve high-level US units (six Esaote MyLab ClassC; six General Electric Logiq e9) were used, all equipped with multifrequency linear probes operating at a frequency of 18 MHz (Esaote) and 15 MHz (General Electric). The same settings (GS frequency 18 MHz Esaote and 15 MHz General Electric; GS gain 50% Esaote and 48% General Electric; PD frequency 8.3 MHz Esaote and 7.7 MHz General Electric; pulse repetition frequency (PRF) 0.5 Hz; PD gain 50% Esaote and 30% General Electric) were used on all units and each sonographer was allowed to modify only one basic function (depth).

Statistical analysis

Intraobserver and interobserver reliability were calculated using the kappa coefficient. Intraobserver reliability was assessed by Cohen's kappa. Interobserver reliability was studied by calculating the mean kappa on all pairs (ie, Light's kappa). Kappa coefficients were interpreted according to Landis and Koch. Kappa values of 0–0.20 were considered poor, 0.20–0.40 fair, 0.40–0.60 moderate, 0.60–0.80 good and 0.80–1.00 excellent.^{12 13} The percentage of observed agreement (ie, percentage of observations that obtained the same score), prevalence of the observed lesions and prevalence-adjusted bias-adjusted kappa (PABAK) were also calculated. Analyses were performed using R Statistical Software (Foundation for Statistical Computing).

RESULTS

Delphi survey

All 19 participants responded to all rounds of the Delphi survey. At the preliminary questionnaire, the definitions extrapolated from the SLR were elaborated and

presented to the panel. The first Delphi round included 15 statements for voting (online supplementary table S3). In the first round, 10 statements reached agreement; the remaining statements and one, modified according to the comments received by the experts, were proposed again for voting in the second round and third rounds, reaching agreement only for one more statement (online supplementary table S4). A summary of the results of the Delphi survey can be seen in [table 1](#). Furthermore, based on the need to assess osteophytes ([table 1](#)), the statement of the Delphi on osteophytes scoring with best agreement was selected (ie, semiquantitative 0–3) (online supplementary table S4).

Web-based exercise

The web-based exercise was successfully completed in two rounds by 13 participants. Interobserver reliability, including both rounds and together MTP and midfoot joints, ranged from 0.50 for synovial hypertrophy (SH) to 0.89 for PD score ([table 2](#)), while considering only midfoot it ranged from 0.51 for SH score to 0.84 for PD score ([table 3](#)), and only MTP joints it ranged from 0.49 for SH to 0.89 for PD score ([table 4](#)). Intraobserver reliability ranged from a minimum value of 0.48 for SH score (0.54 for midfoot and 0.60 for MTP joints, [tables 3–4](#)) to a 0.9 for PD score ([table 2](#)). Adjusting kappa values for the prevalence of the observed lesions, no significant differences were noted.

Training session

The sonographers agreed to use the previously described semiquantitative scoring system for grading SH, joint effusion (JE), PD signal and osteophyte evaluation.

- *SH, JE and PD signal.* During the training session, the sonographers agreed to score SH and JE as absent/present (0–1) and to use for SH also a semiquantitative score (0–3).^{14 15} PD was evaluated with a semiquantitative score (0–3).¹⁵

Table 1 Elementary lesions in foot osteoarthritis: final results of the Delphi survey

Elementary lesions in foot osteoarthritis	Agreement (%)
Midfoot joints must be assessed separately for structural and inflammatory abnormalities in foot OA	84.2
I MTP joint must be assessed separately for structural and inflammatory abnormalities in foot OA	100.0
Joint inflammation and structural changes must be assessed separately in foot OA	94.7
Joint synovial hypertrophy (with or without Doppler signal) should always be assessed in foot OA	100.0
Joint effusion should always be assessed in foot OA	89.5
Synovial hypertrophy can be scored semiquantitatively from 0 to 3 (ie, 0=no; 1=mild; 2=moderate; 3=severe)	84.2
Doppler can be scored semiquantitatively from 0 to 3 (ie, 0=no; 1=mild; 2=moderate; 3=severe)	89.5
Synovial hypertrophy can also be scored dichotomously (ie, 0=absent; 1=present)	78.9
Osteophytes should always be assessed for joint structural changes in foot OA	100.0
Cartilage damage of the first metatarsal head should always be assessed for joint structural changes in foot OA	78.9
II to V MTP joints must be assessed separately for structural and inflammatory abnormalities in foot OA	94.7

MTP, metatarsophalangeal; OA, osteoarthritis.

Table 2 Intraobserver and interobserver results of the static, web-based, exercise considering together results of midfoot and MTP joints (strength of agreement: <0.20 poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, 0.81–1.00 excellent)

	Intraobserver agreement of the static exercise				Interobserver agreement of the static exercise							
	Prevalence mean range (%)	Observed agreement mean, range (%)	Kappa mean range	PABAK mean range	Prevalence mean I round (%)	Prevalence mean II round (%)	Observed agreement mean, I round (%)	Observed agreement mean, II round (%)	Kappa mean I round	Kappa mean II round	PABAK mean I round	PABAK mean II round
Synovial hypertrophy (0–1)	61 (40–78)	83 (51–94)	0.64 (0.00–0.87)	0.66 (0.03–0.87)	59	63	75	83	0.50	0.64	0.51	0.66
Joint effusion (0–1)	59 (46–73)	84 (30–100)	0.67 (–0.39 to 1)	0.67 (–0.40 to 1)	63	56	80	80	0.80	0.61	0.61	0.61
Cartilage damage (0–1)	63 (46–79)	84 (64–100)	0.64 (0.22–1)	0.68 (0.29–1)	66	63	82	83	0.60	0.63	0.64	0.65
Synovial hypertrophy score (0–3)	33 (23–44) grade 1; 25 (19–35) grade 2; 14 (8–21) grade 3	61 (31–81)	0.48 (0.06–0.73)	0.49 (0.08–0.74)	26 (grade 1) 29 (grade 2) 19 (grade 3)	41 (grade 1) 21 (grade 2) 9 (grade 3)	63	71	0.63	0.59	0.51	0.62
Power Doppler score (0–3)	24 (21–25) grade 1; 29 (25–33) grade 2; 14 (19–23) grade 3	92 (88–100)	0.90 (0.83–1)	0.90 (0.83–1)	24 (grade 1) 29 (grade 2) 21 (grade 3)	24 (grade 1) 29 (grade 2) 21 (grade 3)	91	91	0.87	0.89	0.88	0.89
Osteophytes score (0–3)	37 (29–44) grade 1; 28 (23–37) grade 2; 21 (4–23) grade 3	73 (41–88)	0.63 (0.24–0.83)	0.65 (0.22–0.83)	37 (grade 1) 28 (grade 2) 12 (grade 3)	38 (grade 1) 29 (grade 2) 15 (grade 3)	67	70	0.54	0.58	0.55	0.60

PABAK, prevalence-adjusted bias-adjusted kappa.

Table 3 Intraobserver and interobserver results of the static, web-based, exercise on the midfoot joints (strength of agreement: <0.20 poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, 0.81–1.00 excellent)

	Intraobserver agreement of the static exercise_midfoot				Interobserver agreement of the static exercise_midfoot					
	Prevalence mean range (%)	Observed agreement mean, range (%)	Kappa mean range	PABAK mean range	Prevalence mean I round (%)	Prevalence mean II round (%)	Observed agreement mean, I round (%)	Observed agreement mean, II round (%)	Kappa mean I round	Kappa mean II round
Synovial hypertrophy (0–1)	54 (29–75)	85 (75–100)	0.70 (0–0.9)	0.71 (0–0.87)	55	54	76	86	0.52	0.72
Joint effusion (0–1)	53 (44–66)	82 (29–100)	0.65 (–0.42 to 1)	0.65 (–0.42 to 1)	57	50	79	85	0.58	0.70
Synovial hypertrophy score (0–3)	30 (13–41) grade 1; 20 (13–31) grade 2; 17 (7–25) grade 3	65 (14–94)	0.54 (–0.06 to 0.91)	0.54 (–0.14 to 0.91)	23 (grade 1) 29 (grade 2) 19 (grade 3)	37 (grade 1) 12 (grade 2) 15 (grade 3)	63	78	0.51	0.68
Power Doppler score (0–3)	51 (50–67) grade 1; 30 (25–38) grade 2; 18 (0–25) grade 3	91 (75–100)	0.85 (0.6–1)	0.88 (0.66–1)	50 (grade 1) 31 (grade 2) 19 (grade 3)	51 (grade 1) 30 (grade 2) 17 (grade 3)	90	89	0.84	0.83
Osteophytes score (0–3)	38 (28–44) grade 1; 29 (16–44) grade 2; 17 (6–34) grade 3	71 (31–94)	0.60 (0.12–0.91)	0.61 (0.08–0.91)	38 (grade 1) 26 (grade 2) 16 (grade 3)	37 (grade 1) 32 (grade 2) 18 (grade 3)	60	69	0.46	0.59

PABAK, prevalence-adjusted bias-adjusted kappa.

Table 4 Intraobserver and interobserver results of the static, web-based, exercise on the MTP joints (strength of agreement: <0.20 poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, 0.81–1.00 excellent)

Intraobserver agreement of the static exercise_MTP				Interobserver agreement of the static exercise_MTP							
	Observed agreement mean, range (%)	Kappa mean range	PABAK mean range	Prevalence mean I round (%)	Prevalence mean II round (%)	Observed agreement mean, I round (%)	Observed agreement mean, II round (%)	Kappa mean I round	Kappa mean II round	PABAK mean I round	PABAK mean II round
Synovial hypertrophy (0–1)	64 (45–80)	0.61 (0.03–0.87)	0.63 (0.05–0.87)	61	68	75	81	0.49	0.58	0.51	0.57
Joint effusion (0–1)	63 (47–77)	0.67 (–0.35 to 1)	0.68 (–0.39 to 1)	65	61	81	78	0.61	0.56	0.63	0.57
Cartilage damage (0–1)	63 (46–79)	0.64 (0.22–1)	0.68 (0.28–1)	65	63	82	83	0.60	0.63	0.64	0.65
Synovial hypertrophy score (0–3)	35 (28–45) grade 1; 28 (20–38) grade 2; 12 (5–21) grade 3	0.60 (0.19–0.79)	0.46 (0.05–0.87)	27 (grade 1) 30 (grade 2) 20 (grade 3)	44 (grade 1) 26 (grade 2) 5 (grade 3)	63	68	0.51	0.53	0.51	0.57
Power Doppler score (0–3)	19 (15–20) grade 1; 29 (25–35) grade 2; 21 (20–25) grade 3	0.90 (0.79–1)	0.90 (0.8–1)	19 (grade 1) 29 (grade 2) 22 (grade 3)	19 (grade 1) 29 (grade 2) 22 (grade 3)	91	92	0.88	0.89	0.88	0.89
Osteophytes score (0–3)	38 (19–44) grade 1; 28 (18–38) grade 2; 9 (0–25) grade 3	0.70 (0.33–1)	0.71 (0.33–1)	34 (grade 1) 32 (grade 2) 6 (grade 3)	41 (grade 1) 24 (grade 2) 9 (grade 3)	79	83	0.70	0.69	0.72	0.62

MTP; metatarsophalangeal; PABAK, prevalence-adjusted bias-adjusted kappa.

- ▶ *Training session on cartilage damage (CD).* CD was as loss of anechoic structure and/or thinning of cartilage layer¹⁶ (online supplementary figure S2). During the training session, the sonographers agreed to use a binary score for CD (absent/present, 0–1)¹⁶ and to evaluate this lesion only in the first MTP joint. Indeed, to evaluate cartilage by US, the probe has to be perpendicular to the cartilage surface and dorsal osteophytes could limit the US image of cartilage; these being the reasons for limiting the assessment of CD as outlined above.
- ▶ *Training session on osteophyte evaluation.* The sonographers agreed to use the recently published semi-quantitative scoring systems of grading osteophytes (0=none, 1=minor, 2=moderate, 3=major size of osteophytes).^{17 18}
- ▶ *Training session on midfoot joints.* During the training session, the sonographers agreed to evaluate and score the midfoot joints as a single joint and to use the same method also for analysing the images of the web-based exercise. On the patient-based exercise, only the highest score of each lesion was recorded.

Patient-based exercise

The patient-based exercise was successfully completed in two rounds lasting about 3.5 hours each, one in the morning and one in the afternoon of the same day by 11 rheumatologists from five countries. All rheumatologists were experts in US and were members of the OMERACT group. Interobserver reliability, including both rounds, ranged from 0.08 for CD to 0.51 for SH, but when PABAK was considered, it ranged from 0.36 for osteophytes to 0.93 for PD score (table 5). Evaluating the results of the midfoot and MTP separately, interobserver agreement ranged from 0.37 for osteophytes score to 0.95 for PD score (using PABAK) (table 6), while for MTP joints from 0.24 for JE to 0.74 for PD score (using PABAK) (online supplementary table S5). Intrareader reliability ranged from a minimum value of 0.41 for PD score to 0.64 for SH, considering kappa adjusted it reached higher scores ranging from 0.62 for osteophytes to 0.95 for PD score (table 5). Table 6 and online supplementary table S5 reported results divided by midfoot and MTP joints.

DISCUSSION

Foot is a target area in OA and despite the high frequency of involvement and disability, the recent EULAR recommendations on the use of imaging in OA have highlighted that imaging studies in foot are scarce. Therefore, there is a need for more research concerning the benefits of imaging in such, less commonly studied sites of OA.⁵ To our knowledge, this is the first study exploring the reliability of US in scoring inflammatory and structural lesions in OA of the foot. Considering the low prevalence of certain elementary lesions in the patient-based exercise, the reliability assessment

by Cohen's kappa could be misleading and, for this purpose, the use of PABAK values was considered to optimise the evaluation of the strength of agreement. The assessment of both inflammatory and structural damage-related lesions allowed us to globally evaluate the reliability of US in OA of the foot.

In this reliability exercise, SH and JE were evaluated separately and their detection (present/absent, 0–1) showed similar intra-agreement and inter-agreement for both the web-based and the live exercises, reaching good agreement in all assessments. As suggested by the Delphi exercise, in addition to the binary score, a semiquantitative score (0–3) for SH was used and the results, similar to studies in rheumatoid arthritis and psoriatic arthritis,¹⁹ demonstrated moderate intraobserver agreement in the web-based exercises and a good agreement for patient-based exercise.

In all grades of OA, thickening of the synovial lining cell layer, increased vascularity and inflammatory cell infiltration of the synovial membranes are the main histological features.²⁰ Furthermore, angiogenesis and inflammation are closely integrated processes and may affect disease progression and pain.¹ In this scenario, imaging of vascularisation with PD mode is important for providing a complete image of joint inflammation in OA. In this reliability exercise, a semiquantitative scoring of PD demonstrated excellent reliability on static images, confirmed also on live scans with PABAK values greater than 0.9. However, considering the low prevalence of images with PD signal on live exercise, these results need to be confirmed.

Globally, these results, both for SH and PD, show a possible relevant role of US in clinical trials in OA. Moving to foot damage, this issue could significantly impact the assessment of disability of patients with OA. Ultrasound may thus be a promising method for detecting cartilage pathology, also in early stages of OA of the foot. However, in this study, which represents the first step in this field, we decided to use a binary score (absent/present, 0–1) for evaluating CD only in the first MTP joint. This choice was due to the difficulty to image cartilage in the midfoot: indeed to evaluate cartilage by US, the probe has to be perpendicular to the cartilage surface, which could be difficult to obtain in OA of the foot, particularly for midfoot joints. Using a binary score for CD, we found good intraobserver and interobserver reliability. With regard to osteophytes, however, the results of this study differed considerably from the good to excellent intraobserver and interobserver reliability of osteophytes in hand OA.¹⁸ In our study, we could demonstrate only good intra-reliability and fair to moderate inter-reliability.

In conclusion, this study demonstrated that US may be a reliable tool for assessing inflammatory lesions in OA of the foot, while for US lesions related to damage, further studies are needed, particularly in anticipation of the application of US in clinical trials. New tools as reference atlases could be useful to improve reliability

Table 5 Intraobserver and interobserver results of the live, patient-based, exercise considering together results of midfoot and MTP joints (strength of agreement: <0.20 poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, 0.81–1.00 excellent)

	Intraobserver agreement of the live exercise				Interobserver agreement of the live exercise							
	Prevalence mean range (%)	Observed agreement mean, range (%)	Kappa mean range	PABAK mean range	Prevalence mean I round (%)	Prevalence mean II round (%)	Observed agreement mean, I round (%)	Observed agreement mean, II round (%)	Kappa mean I round	Kappa mean II round	PABAK mean I round	PABAK mean II round
Synovial hypertrophy (0–1)	17 (9–26)	89 (70–97)	0.64 (0.10–0.85)	0.78 (0.40–0.94)	17	18	85	85	0.46	0.51	0.70	0.70
Joint effusion (0–1)	15 (7–25)	91 (81–98)	0.62 (0.10–0.94)	0.81 (0.63–0.98)	16	14	85	87	0.40	0.43	0.69	0.73
Cartilage damage (0–1)	88 (46–100)	89 (75–100)	0.5 (–0.12 to 1)	0.79 (0.5–1)	88	89	81	82	0.11	0.08	0.61	0.65
Synovial hypertrophy score (0–3)	12 (5–21) grade 1; 7 (1–12) grade 2; 1 (0–3) grade 3	85 (68–95)	0.64 (0.13–0.80)	0.78 (0.58–0.94)	12 (grade 1) 6 (grade 2) 1 (grade 3)	11 (grade 1) 7 (grade 2) 1 (grade 3)	78	79	0.37	0.37	0.71	0.73
Power Doppler score (0–3)	3 (1–12) grade 1; 1 (0–3) grade 2; 0 (0–0) grade 3	96 (89–99)	0.41 (–0.01 to 0.79)	0.95 (0.86–0.98)	3 (grade 1) 1 (grade 2) 0 (grade 3)	3 (grade 1) 1 (grade 2) 0 (grade 3)	94	94	0.21	0.18	0.93	0.93
Osteophytes score (0–3)	36 (17–66) grade 1; 10 (3–25) grade 2; 2 (0–8) grade 3	72 (59–90)	0.46 (0.21–0.78)	0.62 (0.45–0.86)	36 (grade 1) 10 (grade 2) 2 (grade 3)	36 (grade 1) 10 (grade 2) 2 (grade 3)	53	52	0.19	0.21	0.36	0.36

MTP; metatarsophalangeal; PABAK, prevalence-adjusted bias-adjusted kappa.

Table 6 Intraobserver and interobserver results of the live, patient-based, exercise on the midfoot joints (strength of agreement: <0.20 poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, 0.81–1.00 excellent)

	Intraobserver agreement of the live exercise_midfoot				Interobserver agreement of the live exercise_midfoot							
	Prevalence mean range (%)	Observed agreement mean, range (%)	Kappa mean range	PABAK mean range	Prevalence mean I round (%)	Prevalence mean II round (%)	Observed agreement mean, I round (%)	Observed agreement mean, II round (%)	Kappa mean I round	Kappa mean II round	PABAK mean I round	PABAK mean II round
Synovial hypertrophy (0–1)	11 (3–18)	90 (72–99)	0.49 (0.08–0.79)	0.80 (0.43–0.97)	11	11	87	85	0.32	0.29	0.75	0.71
Joint effusion (0–1)	9 (4–18)	92 (83–100)	0.57 (0.17–1)	0.85 (0.66–1)	10	9	88	89	0.30	0.34	0.76	0.79
Synovial hypertrophy score (0–3)	8 (6–21) grade 1; 3 (1–7) grade 2; 1 (0–3) grade 3	88 (72–99)	0.48 (0.09–0.80)	0.56 (0.43–0.98)	9 (grade 1) 3 (grade 2) 1 (grade 3)	7 (grade 1) 4 (grade 2) 1 (grade 3)	82	82	0.24	0.23	0.76	0.77
Power Doppler score (0–3)	2 (0–6) grade 1; 1 (0–2) grade 2; 0 (0–0) grade 3	97 (91–100)	0.32 (–0.02 to 0.79)	0.96 (0.87–1.0)	2 (grade 1) 1 (grade 2) 0 (grade 3)	2 (grade 1) 1 (grade 2) 0 (grade 3)	97	96	0.22	0.13	0.95	0.95
Osteophytes score (0–3)	3 (14–70) grade 1; 7 (2–23) grade 2; 1 (0–5) grade 3	73 (58–89)	0.43 (0.22–0.75)	0.64 (0.43–0.85)	34 (grade 1) 8 (grade 2) (grade 3)	34 (grade 1) 0 7 (grade 2) 1 (grade 3)	54	52	0.18	0.17	0.38	0.37

PABAK, prevalence-adjusted bias-adjusted kappa.

of US scoring. Finally, based on the results of this study, US seems to be a promising tool to be further tested in diagnostic, prognostic and follow-up studies on foot OA.

Author affiliations

¹Department of Medical and Biological Sciences, Rheumatology Clinic, University Hospital Santa Maria della Misericordia, Udine, Italy

²Dipartimento di Scienze Mediche, Università degli studi di Ferrara, Clinica di Reumatologia, Ferrara, Italy

³Agenzia Sanitaria Locale Viterbo, Viterbo, Italy

⁴Department of Clinical Medicine and Immunology, Rheumatology Section, University of Siena, Siena, Italy

⁵Rheumatology Department of Lucania, Azienda Ospedaliera Regionale San Carlo, Potenza, Italy

⁶Epidemiology Unit, Italian Society for Rheumatology (SIR), Milan, Italy

⁷3rd Department of Rheumatology, National Institute of Rheumatology and Physiotherapy, Budapest, Hungary

⁸Rheumatology, Medisch Centrum Leeuwarden, Sneek, Netherlands

⁹Rheumatology, Ambroise Paré Hospital, APHP, Université Versailles-Saint-Quentin en Yvelines, Boulogne-Billancourt, France

¹⁰Institute of Rheumatology, University of Belgrade Medical School, Belgrade, Serbia

¹¹Rheumatology Unit, University of Pisa, Pisa, Italy

¹²Rheumatology, Università Politecnica delle Marche, Jesi, Italy

¹³University Podiatry Clinic, Universidad Complutense de Madrid, Madrid, Spain

¹⁴Rheumatology, Diakonhjemmet Hospital, Oslo, Norway

¹⁵Department of Rheumatology, Midyorkshire NHS Trust, Midyorkshire, UK

¹⁶Internal Medicine III, Department of Rheumatology, Medical University Vienna, Vienna, Austria

¹⁷Rheumatologia, Instituto Poal de Reumatologia, Barcelona, Spain

¹⁸Clinic of Podiatry, Faculty of E.F. and Podiatry, Universidad Complutense, Madrid, Spain

¹⁹Department of Rheumatology, Joint and Bone Research Unit, Hospital Universitario Fundación Jiménez Díaz, Madrid, Spain

²⁰Azienda Sanitaria Locale 3, ASL 3, Pistoia, Italy

²¹Chair and Division of Rheumatology, University of Pavia, IRCCS Policlinico San Matteo Foundation, Pavia, Italy

²²Center for Rheumatology and Spine Disease, Rigshospitalet-Glostrup, Copenhagen, Denmark

²³Scienze Cliniche e Biologiche, Università degli Studi di Torino, Rome, Italy

Collaborators Fabiana Figus, Iolanda Rutigliano, Chiara Scirocco.

Contributors Reported in the manuscript.

Competing interests None declared.

Patient consent for publication Obtained.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Alen Zabotti <http://orcid.org/0000-0002-0573-464X>

Maria Antonietta D'Agostino <http://orcid.org/0000-0002-5347-0060>

Peter Mandl <http://orcid.org/0000-0003-1526-4052>

Garifallia Sakellariou <http://orcid.org/0000-0002-1849-5123>

Carlo Alberto Scirè <http://orcid.org/0000-0001-7451-0271>

REFERENCES

- Sellam J, Berenbaum F. The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. *Nat Rev Rheumatol* 2010;6:625–35.
- Nelson AE. Osteoarthritis year in review 2017: clinical. *Osteoarthritis Cartilage* 2018;26:319–25.
- Lane NE, Shidara K, Wise BL. Osteoarthritis year in review 2016: clinical. *Osteoarthritis Cartilage* 2017;25:209–15.
- Mathiessen A, Cimmino MA, Hammer HB, et al. Imaging of osteoarthritis (OA): what is new? *Best Pract Res Clin Rheumatol* 2016;30:653–69.
- Sakellariou G, Conaghan PG, Zhang W, et al. EULAR recommendations for the use of imaging in the clinical management of peripheral joint osteoarthritis. *Ann Rheum Dis* 2017;76:1484–94.
- Iagnocco A, Filippucci E, Riente L, et al. Ultrasound imaging for the rheumatologist XXXV. Sonographic assessment of the foot in patients with osteoarthritis. *Clin Exp Rheumatol* 2011;29:757–62.
- Roddy E, Thomas MJ, Marshall M, et al. The population prevalence of symptomatic radiographic foot osteoarthritis in community-dwelling older adults: cross-sectional findings from the clinical assessment study of the foot. *Ann Rheum Dis* 2015;74:156–63.
- Boers M, Kirwan JR, Gossec L, et al. How to choose core outcome measurement sets for clinical trials: OMERACT 11 approves filter 2.0. *J Rheumatol* 2014;41:1025–30.
- Kottner J, Audigé L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96–106.
- Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
- Menz HB, Munteanu SE, Landorf KB, et al. Radiographic classification of osteoarthritis in commonly affected joints of the foot. *Osteoarthritis Cartilage* 2007;15:1333–8.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- HKF M, KKW Y, Chan BPL. Prevalence-adjusted bias-adjusted kappa values as additional indicators to measure observer agreement. *Radiology* 2004;232:302–3.
- Terslev L, Naredo E, Aegerter 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.
- D'Agostino M-A, Boers M, Wakefield RJ, et al. Exploring a new ultrasound score as a clinical predictive tool in patients with rheumatoid arthritis starting abatacept: results from the appraise study. *RMD Open* 2016;2:e000237.
- Iagnocco A, Conaghan PG, Aegerter P, et al. The reliability of musculoskeletal ultrasound in the detection of cartilage abnormalities at the metacarpo-phalangeal joints. *Osteoarthritis Cartilage* 2012;20:1142–6.
- Mathiessen A, Haugen IK, Slatkowsky-Christensen B, et al. Ultrasonographic assessment of osteophytes in 127 patients with hand osteoarthritis: exploring reliability and associations with MRI, radiographs and clinical joint findings. *Ann Rheum Dis* 2013;72:51–6.
- Hammer HB, Iagnocco A, Mathiessen A, 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.
- Zabotti A, Bandinelli F, Batticciotto A, et al. Musculoskeletal ultrasonography for psoriatic arthritis and psoriasis patients: a systematic literature review. *Rheumatology* 2017;56:1518–32.
- Smith MD, Triantafillou S, Parker A, et al. Synovial membrane inflammation and cytokine production in patients with early osteoarthritis. *J Rheumatol* 1997;24:365–71.