Article:

Mandl, P, Studenic, P, Filippucci, E et al. (36 more authors) (2019) Development of semiquantitative ultrasound scoring system to assess cartilage in rheumatoid arthritis. Rheumatology. ISSN 1462-0324

https://doi.org/10.1093/rheumatology/kez153

Development of semiquantitative ultrasound scoring system to assess cartilage in rheumatoid arthritis

Peter Mandl[1], Paul Studenic[1]; Emilio Filippucci[2], Artur Bachta[3], Marina Backhaus[4], David Bong[5], George AW Bruyn[6], Paz Collado[7], Nemanja Damjanov[8], Christian Dejaco[9, 10], Andrea Delle-Sedie[11], Eugenio De Miguel[12], Christina Duftner[13], Irina Gessl[1], Marwin Gutierrez[14], Hilde B Hammer[15], Cristina Hernandez-Diaz[16], Annmaria Iagnocco[17], Kei Ikeda[18], David Kane[19], Helen Keen[20], Stephen Kelly[21], Eszter Kővári[22], Ingrid Möller[5], Uffe Møller-Døhn[23], Esperanza Naredo E[24], Juan C Nieto[25], Carlos Pineda[14], Alex Platzer[1], Ana Rodriguez[26], Wolfgang A Schmidt[27], Gabriela Supp[1], Marcin Szkudlarek[28], Lene Terslev[23], Ralf Thiele[29], Richard J Wakefield[30, 31], Daniel Windschall[32], Maria-Antonietta D'Agostino[33,34], Peter V Balint[35] OMERACT Ultrasound Cartilage Task Force Group

[1]Division of Rheumatology, Medical University of Vienna, Vienna, Austria; [2]Department of Rheumatology, Università Politecnica delle Marche, Jesi, Ancona, Italy; [3]Department of Rheumatology, Military Institute of Medicine, Warsaw, Poland; [4]Department of Internal Medicine, Rheumatology and Clinical Immunology, Park-Klinik Weissensee Academic Hospital of the Charité, Berlin, Germany; [5]Instituto Poal de Reumatologia, University of Barcelona, Barcelona, Spain; [6]MC Groep, Lelystad, the Netherlands; [7]Department of Rheumatology, Hospital Universitario Severo Ochoa, Madrid, Spain; [8]University of Belgrade School of Medicine, Institute for Rheumatology, Belgrade, Serbia; [9]Department of Rheumatology, Medical University of Graz, Graz, Austria; [10]Department of Rheumatology, Hospital of Bruneck, Bruneck, Italy; [11]Rheumatology Unit, University of Pisa, Pisa, Italy; [12]Department of Rheumatology, Hospital Universitario La Paz, Madrid, Spain; [13]Department of Internal Medicine, Clinical Division of Internal Medicine II, Medical University of Innsbruck/Tirol Kliniken, Innsbruck, Austria; [14]Division of Musculoskeletal and Rheumatic Diseases, Instituto Nacional de Rehabilitación, Mexico City, Mexico; [15]Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway; [16]Instituto Nacional de Rehabilitación, Mexico City, Mexico; [17]Academic Rheumatology Center, Università degli Studi di Torino, Turin, Italy; [18]Department of Allergy and Clinical Immunology, Chiba University Hospital, Chiba, Japan; [19]Department of Rheumatology, School of Medicine, Trinity College Dublin, Ireland; [20]School of Medicine and Pharmacology Fiona Stanley Hospital Unit, University of Western Australia, Perth, Australia; [21]Rheumatology Department, Mile End Hospital, Barts Health NHS Trust, London, UK; [22]School of PhD Studies, Semmelweis University, Budapest, Hungary; [23]Center for Rheumatology and Spine Diseases, Copenhagen University Hospital Glostrup, Copenhagen, Denmark; [24]Hospital Universitario Fundación Jiménez Díaz and Autónoma University, Madrid, Spain; [25]Hospital General Universitario Gregorio Marañón and Complutense University, Madrid, Spain; [26]Hospital Ramón y Cajal, Madrid, Spain; [27]Medical Center for Rheumatology, Immanuel Krankenhaus Berlin, Berlin-Buch, Germany; [28]Department of Rheumatology, Zealand's University Hospital at Køge, Denmark; [29]University of Rochester School of Medicine and Dentistry, Rochester, USA; [30]NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds, UK; [31]Leeds Institute of Rheumatic and Musculoskeletal Medicine, Chapel Allerton Hospital, University of Leeds, Leeds, UK; [32]Department of Pediatrics, Asklepios Hospital Weissenfels, Weissenfels, Germany; [33]Department of Rheumatology, APHP, Hopital Ambroise Paré, Paris, France; [34]Department of Rheumatology, INSERM U1173, Laboratoire d'Excellence INFLAMEX, UFR Simone Veil, Versailles-Saint-Quentin University, Versailles, France; [35]3rd Rheumatology Department, National Institute of Rheumatology and Physiotherapy, Budapest, Hungary

Corresponding author: Peter Mandl; Division of Rheumatology, Medical University of Vienna, Währinger Gürtel 18-20, 1090, Vienna, Austria; email: peter.mandl@meduniwien.ac.at

Abstract

Objectives: To develop and test the reliability of a new semiquantitative scoring system for the assessment of cartilage changes by ultrasound (US) in a web-based exercise as well as a patient exercise of patients with rheumatoid arthritis (RA).

Methods: A taskforce of the Outcome Measures in Rheumatology Ultrasound Working Group performed a systematic literature review on the US assessment of cartilage in RA, followed by a Delphi survey on cartilage changes and a new semiquantitative US scoring system, and finally a web-based exercise as well as a patient exercise. For the web-based exercise, taskforce members scored a dataset of anonymized static images of metacarpophalangeal (MCP) joints in RA patients and healthy controls, which also contained duplicate images. Subsequently 12 taskforce members used the same US to score cartilage in MCP and proximal interphalangeal joints of 6 patients with RA in in a patient reliability exercise. Percentage agreement and prevalence of lesions were calculated, as intrareader reliability was assessed by weighted kappa and interreader reliability by Light's kappa.

Results: The three-grade semiquantitative scoring system demonstrated excellent intrareader reliability (kappa: 0.87 and 0.83) in the web-based exercise and the patient exercise respectively. Interreader reliability was good in the web-based exercise (kappa: 0.64) and moderate (kappa: 0.48) in the patient exercise.

Conclusions: Our study demonstrates that ultrasound is a reliable tool for evaluating cartilage changes in the MCP joints of patients with RA and supports further development of a new reliable semiquantitative ultrasound scoring system for evaluating cartilage involvement in RA.

Key words: cartilage, ultrasound, rheumatoid arthritis

Key messages:

1. Consensual definitions of elementary lesions of cartilage changes in rheumatoid arthritis were formulated.
2. An ultrasound scoring system was found to be reliable in assessing cartilage in rheumatoid arthritis.
3. Further testing is required before the scoring system can be recommended as an outcome measure.

INTRODUCTION

Joint damage in patients with rheumatoid arthritis (RA) commonly implies the loss of hyaline cartilage and peri-articular erosive changes (1). It has been shown that loss of cartilage in RA may be more clearly associated with irreversible physical disability than bony damage and therapy directed solely against the erosive process does not ensure the reduction of cartilage loss (2, 3). Particular attention should therefore be given to early detection and therapeutic interference with cartilage destruction, an early key event of disease pathogenesis (2).

The assessment of cartilage and bone damage in RA has traditionally relied on radiographic assessment in which joint space narrowing has served as a surrogate marker of cartilage loss. The most widely used measure of cartilage damage is the Sharp score and its modifications (4, 5). Although joint space narrowing (JSN) is an accepted surrogate marker for cartilage loss, it lacks precision particularly in non-weight-bearing joints and discernment of the relative contributions of damage to cartilage and other soft tissue structures within the JSN score is not possible (6). Recently, musculoskeletal ultrasonography (MSUS) has been suggested as a reliable and reproducible tool for the assessment of cartilage in RA in the small joints of the hand (7-9). A scoring system for cartilage involvement has recently been validated and added to the OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Score (10, 11).

This study reports on the work of the OMERACT Ultrasound Working Group (USWG) which focused on application of the metric properties of MSUS for detecting and evaluating cartilage damage in RA. The main objectives of the study were to develop standardized definitions for the appearance of normal hyaline cartilage on MSUS, its assessment, elementary lesions for assessing hyaline cartilage change and the grading of such changes, and to test the reliability of a consensual semiquantitative scoring system for the assessment of cartilage changes in the metacarpophalangeal (MCP) joints by ultrasound (US) in a web-based exercise as well as a patient exercise of patients with rheumatoid arthritis (RA). Secondary objectives included the testing of the impact of ultrasound machines on reliability and the testing of the semiquantitive scoring system on proximal interphalangeal (PIP) joints.

METHODS

Thirty-four international rheumatologist experts in MSUS from 17 countries (Australia, Austria, Denmark, France, Germany, Hungary, Ireland, Italy, Japan, Mexico, Netherlands, Norway, Poland, Serbia, Spain, UK and USA) formed a taskforce within the OMERACT USWG in 2015. The experts agreed upon a sequence of tasks according to the OMERACT filter 2.0 for US studies (12). As a first step, a systematic literature review (SLR) was performed on studies addressing the sonographic assessment of cartilage in patients with RA. Based on the information obtained from the SLR, the steering committee of the taskforce (PM, EF, MADA, and PVB) formulated statements, including a semiquantitative scoring system, which were agreed upon by the experts in a Delphi exercise. This was followed by testing the reliability of the scoring system first in a web-based exercise on images collected by the experts, followed by a patient-exercise.

First step: systematic literature review

A SLR was performed in the PubMed and Embase databases using the search terms: cartilage AND rheumatoid arthritis AND (ultrasound OR ultrasonography). Both original articles and reviews, as well as abstracts presented at the 2010-2016 ACR and EULAR scientific meetings were included. Titles, abstracts and full reports of articles identified were systematically screened and verified by PM and PVB with regard to inclusion and exclusion criteria. Studies published in English up to November 2016, on the use of MSUS for the imaging of cartilage in adult (≥18 years) patients with RA were included. Data with a particular focus on definitions, scanning technique, scoring of cartilage, cartilage changes and cartilage loss were extracted using a standardized template that was specifically designed for the review. The results of the SLR were used by the steering committee to develop statements for the Delphi process.

Second step: Delphi exercise

A written Delphi questionnaire was constructed on the basis of data collected from the SLR and sent to the participating experts. It consisted of 9 statements/items, including definitions for the appearance of normal hyaline cartilage on MSUS, its assessment, elementary lesions for assessing hyaline cartilage change and the grading of such changes. The panel was asked to rate each item using a level of agreement or disagreement for each statement according to a five-point Likert scale (13), which was graded as follows: 1, strongly disagree; 2, disagree; 3, neither agree nor disagree; 4, agree; 5, strongly agree.

Group agreement was defined as total cumulative agreement >75% (with a score of 4-5). Only when sentences achieved a score >75%, did we consider that the group had reached a consensus and that the statement was defined as appropriate. The answers from each Delphi questionnaire were summarized with mean scores by a facilitator (PM) and re-sent with a revised questionnaire to the panel for the next round, until agreement was reached for all statements.

Third step: web-based intra- and interreader reliability exercise

Taskforce members were instructed to acquire MSUS images of MCP joints 2-5 of healthy subjects and patients with RA using a joint position of approx. 90 degrees of flexion which exposes the largest accessible area of hyaline cartilage in the MCP joints (9) using the standardized dorsal longitudinal midline and transverse scans, according to guidelines set forth in a recent review on the sonographic imaging of cartilage, in particular ensuring an insonation angle of 90 degrees (14). The MSUS equipment used for acquiring the images included the following US units: General Electric Logiq S8, P9 and E9; Esaote Mylab XVG, 25, 70, Class C and Twice; Siemens Acuson Antares and 2000; Phillips Epiq7; Hitachi-Aloka Avius and Ascendus. After a collection period of 1 month, the images were sent by e-mail to a facilitator (PM). A randomly selected group of 25 images were displayed twice in order to assess intrareader reliability. This was sent to the participants, asking them to read each image and grade the metacarpal cartilage based on the semiquantitative scoring system which was agreed upon in the Delphi exercise.

Fourth step: patient-based intra- and inter-observer reliability exercise

Twelve taskforce members (ADS, AI, CDe, CDu, DB, EF, GB, HBH, HK, IM, MADA, PVB) participated in a patient-based intra- and interobserver reliability exercise. During this meeting, MCP and PIP joints 2-5 of six patients with RA were assessed twice on the same day by all experts using ultrasound machines (GE Logiq E9, GE S8, GE Logiq e) equipped with high-frequency transducers (L8-18i-RS ranging from 8-18 MHz and L10-22-RS ranging from 10-22 MHz) with presets calibrated for the appropriate assessment of cartilage. Participants assessed metacarpal cartilage on the dorsal aspect of the respective joints according to recent guidelines (14), ensuring an insonation angle of 90 degrees and utilizing either the standardized dorsal longitudinal midline scan using the so-called flick-view position (in full possible flexion) or the freehand or dynamic technique whereby the joint position remained the same as during the standardized scan, however the sonographer was at liberty to shift the transducer and use both longitudinal and transverse planes to assess the entire. Cartilage was scored by the semiquantitative scoring system agreed upon in the Delphi process. Cartilage in the PIP joint was examined on the dorsal aspect using only the dynamic technique outlined above. Two of the patients were examined on the same machine in the morning and the afternoon session, and four patients were examined on different machines, in order to evaluate inter-machine variability.

Statistical analysis

Intraobserver reliability was assessed by weighted kappa and interobserver reliability was assessed by Light's kappa. Kappa values were interpreted as follows: values of: 0-0.20 represent slight; 0.21-0.40 fair; 0.41-0.60 moderate; 0.61-0.80 good and >0.80 excellent reliability (15). Additionally, 95% confidence intervals (95%CI) were calculated. Percentage of observed agreement (i.e., percentage of observations that obtained the same score) and prevalence of the observed lesions were also calculated. Statistical analyses were performed using R and STATA. The ethics committee of the Medical University of Vienna approved the study, which was conducted according to the guidelines of the Declaration of Helsinki. Each patient gave written informed consent to participate.

RESULTS

Systematic literature review

A total of 198 articles were identified of which finally 9 studies reporting on original research could be included in the systematic literature review (7-9, 16-21). The flowchart of the review process is included in Supplementary File 1. Data extracted from the included studies were used to formulate statements for the Delphi process and was also shared with the participants for the patient exercise. Key data from studies selected for final review are summarized in Table 1.

Delphi exercise

A total of 27 experts were invited of whom 24 (89%) participated in both the first and second, final round of the Delphi exercise. For 7 out of 9 statements agreement was ≥75%, and for 2 statements 75%< after the first round. Wording was improved based on experts' comments and all statements were presented in the second round of the exercise, in which all statements achieved agreement. Table 2 shows the final statements and their agreement. Among others, the participants agreed upon the definition of normal hyaline cartilage on US as well as on the elementary lesion of cartilage damage: blurring of the outer margin and/or the subchondral margin under orthogonal insonation, focal or diffuse thinning of the hyaline cartilage layer as well as the incomplete or complete loss of homogeneity of the echostructure. Based on these definitions a semiquantitative scoring system ranging from 0-2 (grade 0, normal cartilage; grade 1, minimal change: focal thinning or incomplete loss of cartilage; grade 2, severe change: diffuse thinning or complete loss of cartilage) was formulated (Figure 1.). In addition, the participants agreed upon a statement on the quantitative assessment of cartilage, taking into consideration the recommendations from a recent review on the pitfalls of cartilage measurement on US, in particular the need for orthogonal insonation, inclusion of the outer margin in the measurement and correction for the higher speed of sound in hyaline cartilage as compared to soft tissue.

Web-based intra- and interreader reliability exercise

A total of 17 taskforce members sent 20 anonymized images each of MCP joints 2-5 acquired from healthy subjects and patients with RA, both in the longitudinal and transverse scans. The conveners (EF, PM and PVB) reviewed the total number of 340 images for quality and created a dataset of 123 images, consisting of 73 individual images as well as duplicates of 25 randomly selected images. The dataset was sent to participants who graded each image using the semiquantitative grading system agreed upon in the Delphi exercise. The kappa values for intrareader reliability of the web-based exercise wer 0.87 (95%CI 0.83-0.92), and for interreader reliability the kappa values were 0.64 (95%CI 0.63-0.64).

Patient intra- and interreader reliability exercise

In the patient exercise, 4 out of 6 patients were women; mean age was 64 (range: 52-67) years, mean disease duration was 15 (range: 4-31) years, 83% (5/6) of patients were rheumatoid factor and anti-citrullinated peptide antibody positive. The observed prevalence of grades of cartilage damage for both the MCP and the PIP joints are listed in Table 3.

The results of the reliability of the semiquantitative scoring system are summarized in Table 4. The intrareader agreement was 84.2% (range 64.9-100) and 76.2% (range: 57.7-95.3) for the MCP joints (standardized & dynamic respectively) and 57.1% (range: 23.7-92.9) for the PIP joints. Kappa values for intrareader reliability were 0.78 (95%CI 0.74-0.82) for the standardized scan of the MCP joints, 0.83 (95%CI 0.80-0.86) for the dynamic scan of the MCP joints and 0.66 (95%CI 0.60-0.71) for the PIP joints. The interreader agreement was 62.7% (range 28.1-79.1) and 64.3% (range: 45.8-80.2) for the MCP joints (standardized & dynamic respectively) and 44.1% (range: 18.7-69.8) for the PIP joints. Kappa values for interreader reliability were 0.44 (95%CI 0.38-0.51) for the standardized scan of the MCP joints, 0.48 (95%CI 0.41-0.54) for the dynamic scan of the MCP joints and 0.17 (95%CI 0.13-0.21) for the PIP joints.

The estimates for each individual joint are listed in Table 5. Based on the kappa values no individual joint could be selected which performed better than the others, although overall the MCP5 and PIP5 joints performed worse as compared to MCP 2-4 and PIP 2-4, respectively. Finally, the estimates for each individual patient are listed in Supplementary Table 1. Kappa values for intra- and interreader reliability varied consistently, with higher overall estimates for Patient 2 and 4, who both had relatively longer disease duration, as compared to the other patients.

Impact of different ultrasound machines on reliability

Intrareader reliability was better for examinations which took place on the same machine as compared to those performed on different machines (kappa values: 0.73 (95%CI 0.63-0.81) vs. 0.59 (95%CI 0.51-0.66) for the standard scan of the MCP joints; 0.64 (95%CI 0.53-

0.73) vs. 0.52 (95%CI 0.43-0.59) for the dynamic scan of the MCP joints and 0.59 (95%CI 0.49-0.69) vs. 0.48 (95%CI 0.40-0.56) for the PIP joints).

DISCUSSION:

The main objective of this study was to develop definitions for cartilage damage in RA and test the reliability of a semiquantitative scoring system. Standardization of changes and validated scoring system would facilitate the dissemination of this technique in daily practice and allow adequately trained sonographers to participate in multicenter research studies aiming to assess cartilage changes.

This is the first reliability study of a sonographic scoring of cartilage abnormalities in RA that was developed according to the OMERACT framework. The inclusion of different US machines in both the web-based and patient-based exercise corresponds to the real-life application of US in routine clinical practice and multicenter studies, and also allowed us to demonstrate that the use of different vs. same US machines indeed has an impact on reliability.

Although an OMERACT taskforce on hand osteoarthritis (OA) reported good agreement on definitions of cartilage damage in hand OA (22), recent attempts of developing a semiquantitative scoring system in hand osteoarthritis have found only moderate intrareader and fair interreader reliability (23). It was suggested that the poor reliability, in particular of the two intermediate scores (scores 1 and 2 on a 0-3 scale) may be explained by the fact that the proposed definitions could not help to sufficiently distinguish between intermediary grades. The systematic literature review revealed a single study by Filippucci et al. who have performed a single-center interreader reliability study on MCP 2-3 joints in RA patients using two experienced rheumatologists (8) and reported substantial reliability for a 0-4 semiquantitative scoring system. In the Delphi exercise, the taskforce opted for a simpler semiquantitative scoring system of 0-2. Using this system, we found substantial to excellent intrareader reliability and moderate to substantial interreader reliability (web-based and patient-based exercise respectively) in the MCP joints of RA patients. The dynamic or freehand scanning of the MCP joints was found to be slightly superior to the standardized view. In addition, we could also confirm that using the same machine in both the morning and afternoon round leads to improved intraobserver reliability.

At the same time, by opting for a scoring system between 0-2, while this may be reliable and useful for assessing focal or severe cartilage damage at single timepoints, based on the above-mentioned experiences in scoring it may potentially be less discriminant in studies investigating progression apart from those conducted on patients with very early disease. In addition to the semiquantitative scoring system the taskforce also agreed on a statement on quantitative grading, which may provide a more accurate evaluation of cartilage, albeit it would likely be less feasible in a multicenter study. The latter definition is in line with recent recommendations which highlighted the pitfalls of US measurement of cartilage (14).

An additional limitation of our study could be the absence of PIP images in the web-based reliability exercise, which may explain the moderate intrareader reliability and only slight interobserver reliability in the PIP joints. The latter results may also reflect technical problems associated with the visualization of cartilage in this joint, which may require further modification of the scanning technique (e.g. utilizing palmar transverse scans to visualize cartilage). Although the number of patients seems very low, the number of examined structures in total was quite high (n:96). In addition, the number of readers was also quite high (n:12). According to several reports (24, 25) focusing on improving variability of reliability studies it is important either to have an adequate number of patients or of readers. In these studies, 6 to 8 patients or 10 to 14 readers are recommended as adequate simple sizes. The number of patients utilized in our study is in the range used in previous reliability exercises on US (26, 27). Although we took care to include patients in the patient exercise that conform to an average RA population with regard to distribution of age and sex and included patients with different disease duration, due to the low patient number, which is usual for such exercises, we cannot rule out a patient selection bias, which may have affected the results.

Based on the present study, the OMERACT USWG recommends the use of the presently described semiquantitative MSUS score for assessing cartilage pathology in the MCP joints of patients with RA. Further testing of this scoring system in the MCP joints of other RA cohorts in addition to joints where cartilage can be visualized (e.g. knee, metatarsophalangeal, tibiotalar, etc.) and assessment of sensitivity of change in longitudinal studies is required before the scoring system can be recommended as an outcome measure to be used in clinical trials.

REFERENCES:

1. Pap T, Distler O. Linking angiogenesis to bone destruction in arthritis. Arthritis Rheum. 2005;52:1346-8.
2. Aletaha D, Funovits J, Smolen JS. Physical disability in rheumatoid arthritis is associated with cartilage damage rather than bone destruction. Ann Rheum Dis 2011;70:733–9.
3. Landewé R, van der Heijde D. Joint space narrowing, cartilage and physical function: are we deceived by measurements and distributions? Ann Rheum Dis 2011;70:717–8.
4. Sharp JT, Lidsky MD, Collins LC, et al. Methods of scoring the progression of radiologic changes in rheumatoid arthritis. Correlation of radiologic, clinical and laboratory abnormalities. Arthritis Rheum 1971;14:706–20.
5. van der Heijde DM. Radiographic imaging: the 'gold standard' for assessment of disease progression in rheumatoid arthritis. Rheumatology (Oxford) 2000;1:9–16.
6. Navarro-Compán V, Landewe R, Provan SA, et al. Relationship between types of radiographic damage and disability in patients with rheumatoid arthritis in the EURIDISS cohort: a longitudinal study. Rheumatology (Oxford) 2015;54:83-90.

7. Möller B, Bonel H, Rotzetter M, et al. Measuring finger joint cartilage by ultrasound as a promising alternative to conventional radiograph imaging Arthritis Rheum 2009;61:435–41.
8. Filippucci E, da Luz KR, Di Geso L, et al. Interobserver reliability of ultrasonography in the assessment of cartilage damage in rheumatoid arthritis. Ann Rheum Dis 2010;69:1845–8.
9. Mandl P, Supp G, Baksa G, et al. Relationship between radiographic joint space narrowing, sonographic cartilage thickness and anatomy in rheumatoid arthritis and control joints. Ann Rheum Dis 2015;74:2022-7.
10. Døhn UM, Conaghan PG, Eshed I, et al. The OMERACT-RAMRIS rheumatoid arthritis magnetic resonance imaging joint space narrowing score: intrareader and interreader reliability and agreement with computed tomography and conventional radiography. J Rheumatol 2014;41:392-7.
11. Glinatsi D, Lillegraven S, Haavardsholm EA, et al. Validation of the OMERACT Magnetic Resonance Imaging Joint Space Narrowing Score for the wrist in a multireader longitudinal trial. J Rheumatol 2015;42:2480-5.
12. Boers M, Kirwan JR, Wells G, Beaton D, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol. 2014;67:745–53.
13. Likert R. A Technique for the Measurement of Attitudes. Archives of Psychology 1932;140:1–55.
14. Torp-Pedersen S, Bartels EM, Wilhjelm J, et al. Articular cartilage thickness measured with US is not as easy as it appears: a systematic review of measurement techniques and image interpretation. Ultraschall Med 2011;32:54–61.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:15974.
16. Iagnocco A, Coari G, Zoppini A. Sonographic evaluation of femoral condylar cartilage in osteoarthritis and rheumatoid arthritis. Scand J Rheumatol. 1992;21:201-3.
17. Grassi W, Tittarelli E, Pirani O, et al. Ultrasound examination of metacarpophalangeal joints in rheumatoid arthritis. Scand J Rheumatol. 1993;2:243-7.
18. Riente L, Delle Sedie A, Filippucci E, et al. Ultrasound Imaging for the rheumatologist XXVII. Sonographic assessment of the knee in patients with rheumatoid arthritis. Clin Exp Rheumatol. 2010;28:300-3.
19. Yücesoy CY, Genc G, Bal A, et al. Ultrasonographic Assessment of Knee in Patients with Rheumatoid Arthritis: Is it an Effective Imaging Method for Initial Evaluation? Turk J Rheumatol 2011;26:120-126.
20. Di Geso L, Filippucci E, Riente L, et al. Ultrasound imaging for the rheumatologist XL. Sonographic assessment of the hip in rheumatoid arthritis patients. Clin Exp Rheumatol. 2012;30:464-8.
21. Onodera T, Kasahara Y, Kasemura T, et al. A Comparative Study With In Vitro Ultrasonographic and Histologic Grading of Metatarsal Head Cartilage in Rheumatoid Arthritis. Foot Ankle Int. 2015;36:774-9
22. Iagnocco A, Conaghan PG, Aegerter P, et al. The reliability of musculoskeletal ultrasound in the detection of cartilage abnormalities at the metacarpophalangeal joints. Osteoarthritis Cartilage 2012;20:1142–6.
23. 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.

24. Obuchowski NA, Zepp RC. Simple steps for improving multiple-reader studies in radiology. AJR Am J Roentgenol. 1996;166:517-21.
25. Obuchowski NA. How many observers are needed in clinical studies of medical imaging? AJR Am J Roentgenol. 2004;182:867-9.#
26. Terslev L, Gutierrez M, Christensen R, 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.
27. Balint PV, Terslev L, Aegerter P, et al. Reliability of a consensus-based ultrasound definition and scoring for enthesitis in spondyloarthritis and psoriatic arthritis: an OMERACT US initiative. Ann Rheum Dis. 2018;77:1730-5.

Table 1. Systemic literature review

Study	Subjects	Sample size	Joint scanned	Scoring method used	Definition used for cartilage damage	Reliability	Correlation with other method
Iagnocco A. Scand J Rheumatol. 1992 (16)	Healthy, RA, OA	30/48/60	knee	Discrete measurement & binary grading	Blurring, irregularity	NA	NA
Grassi W. Scand J Rheumatol. 1993 (17)	Healthy/RA	20/20	MCP 2-3	Binary grading	Loss of definition, indistinctness	NA	NA
Möller B. Arthritis Rheum 2009 (7)	Healthy, RA, OA, unclassified arthritis	34/48/18/ 24	MCP 2-5, PIP 2-5	Discrete measure	Thinning of cartilage	ICC for cartilage thickness (bilateral sum score): 0.844 (interobserver) 0.928 (intraobserver)	Correlation between MSUS score and radiographic JSN for both hands: adjusted r^2: 0.513 ($p<0.001$) and JSW of the same finger joints: adjusted r^2: 0.635 ($p<0.001$)
Filippucci E. Ann Rheum Dis 2010 (8)	RA	20	MCP 2-3	Semiquantitative score 0-4, adapted from (23)	1=loss of the sharpness of the superficial margin of the hyaline cartilage; 2=partial thickness defect of the cartilage layer; 3=full thickness defect of the cartilage layer with a normal subchondral bone profile; 4=complete loss of the cartilage layer and subchondral bone involvement.	Weighted kappa: 0.672, 0.537 & 0.832 (interobserver reliability for dorsal, lateral and volar quadrant) Weighted kappa values for total additive scores per joint: 0.729 & 0.733, respectively, for detection and for semiquantitative assessment of cartilage damage	NA
Riente L. Clin Exp Rheumatol. 2010 (18)	RA	100	knee	Binary grading based on definition adopted from (24)	Thickening or thinning of cartilage layer, loss of definition of chondro-synovial margin and pitting of the articular surface	NA	NA
Yücesoy CY. Turk J Rheumatol	RA	30	knee	Discrete measurement & binary grading	Irregularity, loss of clarity	NA	Agreement between MRI & MSUS on cartilage

2011 (19)							morphology, kappa: 0.658, 0.851 (medial & lateral)
Di Geso L. Clin Exp Rheumatol. 2012 (20)	RA	52	hip	Binary grading based on definition adopted from (25)	Loss of sharpness of the cartilage margins, loss of homogeneity of the cartilage layer, cartilage thinning (focal or extending to the entire cartilaginous layer)	NA	NA
Mandl P. Ann Rheum Dis 2015 (9)	Healthy*/RA	3/5/35	MCP 2-5	Discrete measure	Thinning of cartilage	ICC for cartilage thickness (individual joints): ICC 0.78 (intraobserver), 0.80 (interobserver)	ICC between sonographic and anatomic cartilage thickness: 0.73
Onodera T. Foot Ankle Int 2015 (21)	RA#	15	MTP 2-5	Semiquantitative score 0-6, adapted from (26)	1=Blurred margin or partial lack of clarity, without thickness change 2=Blurred margin and partial lack of clarity, without thickness change 3=Blurred margin and complete lack of clarity 4=Difficult-to-define margin and completely opaque band; 5=Marked thickness change; 6=No visualized cartilage band	Correlation between in vivo & in vitro MSUS (before and after operation): r=0.74 (p<0.01)	Correlation between MSUS grading & histologic grading: r=0.67-0.83 (p<0.01)

*Including cadaver; # scans were performed pre-operatively as well as in vitro on the resected metatarsal heads; ICC: intraclass correlation coefficient, JSN: joint space narrowing, JSW: joint space width, MCP: metacarpophalangeal joint, MRI: magnetic resonance imaging, MSUS: musculoskeletal ultrasound, MTP: metatarsophalangeal joint, NA: not assessed, OA: osteoarthritis, RA: rheumatoid arthritis

Table 2. Statements and final agreement after the second round of Delphi exercise

Category	Statement	Round	Agreement
MSUS definition of normal hyaline cartilage	Normal hyaline cartilage has a homogeneous anechoic or hypoechic echostructure, parallel to the echogenic bony cortex, is delineated by a sharp subchondral margin, and possesses a sharp outer margin, when the cartilage is insonated orthogonally.	1	88%
Assessment of hyaline cartilage by MSUS	Hyaline cartilage must be assessed using orthogonal insonation (MSUS beam falling perpendicular to the hyaline cartilage surface). The joint should be positioned to expose the largest accessible area of hyaline cartilage. The entire area of cartilage accessible within the acoustic window should be scanned, in both longitudinal and transverse planes	1	96%
	The optimization of settings, including the position of the joint at the baseline examination as well as the maintenance of such settings for possible follow-up examination(s) is mandatory for the purpose of monitoring in clinical practice. Whenever possible, anatomical landmarks should be identified and utilized to ensure that follow-up assessments are conducted at the appropriate locations.	1	96%
	Semiquantitative grading system (0-2) to assess cartilage change in RA	1	80%
Elementary MSUS lesions of cartilage change	Blurring of the outer margin and/or the subchondral margin under orthogonal insonation	1	84%
	Focal or diffuse thinning of the hyaline cartilage layer	1	100%
	Incomplete or complete loss of homogeneity of the echostructure	1	88%
Grading MSUS cartilage changes	For the purpose of grading changes, hyaline cartilage should be assessed in both longitudinal and transverse planes	2	94%
Semiquantitative grading of cartilage	A 3-grade semiquantitative scoring system (i.e. grade 0, normal cartilage; grade 1, minimal change: focal thinning or incomplete loss of cartilage; grade 2, severe change: diffuse thinning or complete loss of cartilage can be used to grade hyaline cartilage change in rheumatoid arthritis	1	80%
Quantitative grading of cartilage	Hyaline cartilage thickness can be measured using the largest distance between the subchondral and outer margins, and if possible including the outer, but not the subchondral margin by the caliper tool. For monitoring purposes, the cartilage thickness measurement using the calipers is sufficient and does not need to be corrected, when correlating with anatomical/histological or other imaging measurement, the data obtained using the calipers should be corrected for the higher speed of sound in hyaline cartilage as compared to soft tissue.	2	77%

MSUS: musculoskeletal ultrasound, RA: rheumatoid arthritis

Table 3. Prevalence of semiquantitative grades for the patient-based exercise (mean prevalence for both rounds for all examiners)

MCP	Grade (0-2)	Observed prevalence (%)		PIP	Grade (0-2)	Observed prevalence (%)
		Dynamic	Standard			Dynamic
2	0	24.1	30.5	2	0	38.2
	1	26.4	22.6		1	41.3
	2	49.5	46.9		2	20.5
3	0	29.9	35.1	3	0	39.2
	1	38.0	33.7		1	32.3
	2	32.1	31.2		2	28.5
4	0	38.9	47.6	4	0	29.9
	1	45.1	36.1		1	42.4
	2	16.0	16.3		2	27.7
5	0	38.9	49.3	5	0	27.1
	1	38.8	28.8		1	47.2
	2	22.3	21.9		2	25.7

MCP: metacarpophalangeal joint, PIP: proximal interphalangeal joint

Table 4. Overall prevalence, intra- and interreader agreement and intra- and interreader reliability

Joint group and scanning method	Grade (0-2)	Prevalence range in %	Intrareader agreement in mean (range), %	Intrareader reliability kappa (%95CI)	Interreader agreement in mean (range) %	Interreader reliability kappa (%95CI)
MCP Standard	0	9.3-54.2	84.2 (64.9-100)	0.78 (0.74-0.82)	62.7 (28.1-79.1)	0.44 (0.38-0.51)
	1	12.5-55.2				
	2	14.6-43.7				
MCP Dynamic	0	18.7-53.1	76.2 (57.7-95.3)	0.83 (0.80-0.86)	64.3 (45.8-80.2)	0.48 (0.41-0.54)
	1	8.7-48.9				
	2	20.8-42.2				
PIP Dynamic	0	3.1-61.4	57.1 (23.7-92.9)	0.66 (0.60-0.71)	44.1 (18.7-69.8)	0.17 (0.13-0.21)
	1	23.9-67.7				
	2	8.3-56.2				

MCP: metacarpophalangeal joint, PIP: proximal interphalangeal joint

Table 5. Intra- and interreader reliability estimates for each joint

MCP	Scanning method	Intrareader reliability kappa (%95CI)	Interreader reliability kappa (%95CI)
2	Standard	0.86 (0.74-0.90)	0.47 (0.39-0.56)
	Dynamic	0.83 (0.77-0.88)	0.51 (0.41-0.59)
3	Standard	0.85 (0.77-0.90)	0.47 (0.34-0.59)
	Dynamic	0.86 (0.80-0.91)	0.57 (0.48-0.70)
4	Standard	0.78 (0.70-0.85)	0.33 (0.22-0.50)
	Dynamic	0.67 (0.53-0.75)	0.34 (0.22-0.52)
5	Standard	0.80 (0.73-0.86)	0.44 (0.31-0.58)
	Dynamic	0.69 (0.59-0.78)	0.41 (0.28-0.57)
PIP			
2	Dynamic	0.72 (0.62-0.79)	0.14 (0.09-0.19)
3	Dynamic	0.72 (0.61-0.81)	0.26 (0.17-0.36)
4	Dynamic	0.68 (0.57-0.77)	0.19 (0.13-0.26)
5	Dynamic	0.50 (0.31-0.64)	0.09 (0.04-0.16)

MCP: metacarpophalangeal joint, PIP: proximal interphalangeal joint

Figure 1. Semiquantitative (0-2) scoring system for cartilage change in rheumatoid arthritis; A) Grade 0: normal cartilage; B) Grade 1, minimal change: focal thinning or incomplete loss of cartilage; C) Grade 2, severe change: diffuse thinning or complete loss of cartilage.

Supplementary File 1. Flowchart for the systemic literature review

Supplementary File 2. Intra- and interreader reliability estimates for metacarpophalangeal-, interphalangeal joints as well as their combination for each patient

Acknowledgements

The authors thank the patients who participated in the patient reliability exercise. We are grateful to GE Austria for providing the ultrasound machines for the patient reliability exercise. List of study collaborators. OMERACT Ultrasound Cartilage Task Force Group: Peter Mandl, Emilio Filippucci, Artur Bachta, Marina Backhaus, David Bong, George AW Bruyn, Paz Collado, Nemanja Damjanov, Christian Dejaco, Andrea Delle-Sedie, Eugenio De Miguel, Christina Duftner, Marwin Gutierrez, Hilde Berner Hammer, Cristina Hernandez-Diaz, Annmaria Iagnocco, Kei Ikeda, David Kane, Helen Keen, Stephen Kelly, Ingrid Möller, Uffe Møller-Døhn, Esperanza Naredo E, Juan C Nieto, Carlos Pineda, Ana Rodriguez, Wolfgang A Schmidt, Marcin Szkudlarek, Lene Terslev, Ralf Thiele, Richard J Wakefield, Daniel Windschall, Maria-Antonietta D'Agostino, Peter V Balint

Contributorship

PM contributed to study design, managed the study, collected and interpreted clinical data, drafted the manuscript and approved the final version. PS, EF, AB, MB, DB, GAWB, PC, ND, CDe, ADS, EDM, CDu, IG, MG, HBH, CHD, AI, KI, DK, HK, SK, EK, IM, UMD, EN, JCN, CP, AP, AR, WAS, GS, MS, LT, RT, RJW, DW contributed to statistical analysis and interpretation of data, revised the manuscript and approved the final version. MADA, PVB contributed to study design, managed the study, collected and interpreted clinical data, drafted the manuscript and approved the final version

Role of the funding source

Unrestricted grant from UCB Pharma GmbH.

Competing interests

The authors declare no conflict of interest.

Patient consent

Patient consent was obtained.

Ethical approval information

The ethics committee of the Medical University of Vienna approved the study, which was conducted according to the guidelines of the Declaration of Helsinki.