

The OMERACT Ultrasound Task Force — Status and Perspectives

ESPERANZA NAREDO, RICHARD J. WAKEFIELD, ANNAMARIA IAGNOCCO, LENE TERSLEV, EMILIO FILIPPUCCI, FREDERIQUE GANDJBAKHCH, PHILIPPE AEGERTER, SIBEL AYDIN, MARINA BACKHAUS, PETER V. BALINT, GEORGE A.W. BRUYN, PAZ COLLADO, STEPHANIE FINZEL, JANE E. FREESTON, MARWIN GUTIERREZ, FREDERICK JOSHUA, SANDRINE JOUSSE-JOULIN, DAVID KANE, HELEN I. KEEN, INGRID MOLLER, PETER MANDL, SARAH OHRNDORF, CARLOS PINEDA, WOLFGANG A. SCHMIDT, MARCIN SZKUDLAREK, PHILIP G. CONAGHAN, and MARIA-ANTONIETTA D'AGOSTINO.

ABSTRACT. This article reports the most recent work of the Outcome Measures in Rheumatology (OMERACT) Ultrasound Task Force, and highlights the future research priorities discussed at the OMERACT 10 meeting. Results of the following studies were presented: (1) intra- and interobserver reliability of ultrasound detecting and scoring synovitis in different joints of patients with rheumatoid arthritis (RA); (2) systematic review of previous ultrasound scoring systems of synovitis in RA; (3) enthesitis systematic review and Delphi definition exercise in spondyloarthritis enthesitis; (4) enthesitis intra- and interobserver reliability exercise; and (5) Delphi definition exercise in hand osteoarthritis, and reliability exercises. Study conclusions were discussed, and a future research agenda was approved, notably further validation of an OMERACT ultrasound global synovitis score (GLOSS) in RA, emphasizing the importance of testing feasibility, predictive value, and added value over standard clinical variables. Future research areas will include validating scoring systems for enthesitis and osteoarthritis, and testing the metric qualities of ultrasound for evaluating tenosynovitis and structural damage in RA. (J Rheumatol 2011;38:2063–7; doi:10.3899/jrheum.110425)

Key Indexing Terms:

ULTRASOUND
ENTHESITIS

POWER DOPPLER
STANDARDIZATION

SYNOVITIS
OMERACT

Since 2004 an international collaborative group of ultrasound experts known as the OMERACT Ultrasound Task Force has worked to address the metric qualities of ultrasound in rheumatoid arthritis (RA) and other arthritides, according to criteria specified by the OMERACT filter. Despite its widespread use in daily practice, ultrasound has

been perceived as an imaging technique that lacks validity^{1,2,3}. In order to address such validity issues, the group has worked on standardizing ultrasound as applied to the most common rheumatic diseases. Over the last 6 years, the research work of the group has mainly focused on the development of a reliable standardized scoring system for syn-

From Hospital Universitario Severo Ochoa, Madrid, Spain; Section of Musculoskeletal Disease, University of Leeds and NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK; Sapienza University of Rome, Rome, Italy; University of Copenhagen, Hvidovre Hospital, Copenhagen, Denmark; University of Ancona, Ancona, Italy; Université Pierre et Marie Curie, La Pitié Salpêtrière Hospital, APHP, Paris, France; Université Paris Ouest-Versailles-Saint Quentin en Yvelines, Hôpital Ambroise Paré APHP, Boulogne-Billancourt, France; Humboldt University, Charité Hospital, Berlin, Germany; National Institute of Rheumatology and Physiotherapy, Budapest, Hungary; Medical Centre Leeuwarden, Leeuwarden, The Netherlands; University of Erlangen-Nuremberg, Department of Internal Medicine 3, Erlangen, Germany; University of New South Wales, Sydney, Australia; University of Brest, La Cavale Blanche Hospital, Brest, France; Adelaide and Meath Hospital, Dublin, Ireland; School of Medicine and Pharmacology, University of Western Australia, Perth, Australia; Instituto Poal, Barcelona, Spain; Instituto Nacional Rehabilitación, Mexico City, Mexico; and Berlin-Buch Hospital, Berlin, Germany.

E. Naredo, MD, Hospital Universitario Severo Ochoa; R.J. Wakefield, MD, Section of Musculoskeletal Disease, University of Leeds; A. Iagnocco, MD, Sapienza University of Rome; L. Terslev, MD, PhD, University of Copenhagen Hvidovre Hospital; E. Filippucci, MD, University of Ancona; F. Gandjbakhch, MD, Université Pierre et Marie Curie, Hôpital La Pitié Salpêtrière, APHP; P. Aegerter, MD, PhD,

Université Paris Ouest-Versailles-Saint Quentin en Yvelines, Hôpital Ambroise Paré, APHP; S. Aydin, MD, Academic Section of Musculoskeletal Disease, Leeds Institute of Molecular Medicine, University of Leeds; M. Backhaus, MD, Humboldt University; P.V. Balint, MD, PhD, National Institute of Rheumatology and Physiotherapy; G.A.W. Bruyn, MD, PhD, Medical Centre Leeuwarden; P. Collado, MD, Hospital Universitario Severo Ochoa; S. Finzel, MD, University of Erlangen-Nuremberg; J.E. Freeston, MD, Academic Section of Musculoskeletal Disease, Leeds Institute of Molecular Medicine, University of Leeds; M. Guitierrez, MD, University of Ancona; F. Joshua, MD, PhD, University of NSW; S. Jousse-Joulin, MD, University of Brest; D. Kane, MD, PhD, Adelaide and Meath Hospital; H.I. Keen, MD, PhD, School of Medicine and Pharmacology, University of Western Australia; I. Moller, MD, Instituto Poal; P. Mandl, MD, PhD, National Institute of Rheumatology and Physiotherapy; S. Ohrndorf, MD, Humboldt University; C. Pineda, MD, Instituto Nacional Rehabilitación; W.A. Schmidt, MD, Berlin-Buch, Berlin, Germany; M. Szkudlarek, MD, PhD, University of Copenhagen Hvidovre Hospital; P.G. Conaghan, MD, PhD, FRCP, Section of Musculoskeletal Disease, University of Leeds and NIHR Leeds Musculoskeletal Biomedical Research Unit; M.A. D'Agostino, MD, PhD, Université Paris Ouest-Versailles-Saint Quentin en Yvelines, Hôpital Ambroise Paré, APHP.

*Address correspondence to Prof. D'Agostino;
E-mail: maria-antonieta.dagostino@apr.ap-hop-paris.fr*

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2011. All rights reserved.

ovitis in RA, which combines gray-scale (GS) and power Doppler (PD) in a 0–3 scale that is applicable to all joints and is consistent between machines. We started at OMERACT 7 by presenting results of a general systematic literature, leading to our development of preliminary consensus definitions for ultrasound-defined inflammatory pathology⁴. The evaluation of the reliability of the technique was considered a priority and subsequently, a number of projects were undertaken to assess the reliability of ultrasound in inflammatory arthritis, but with a particular focus on RA^{5,6,7}.

Results of those exercises highlighted specific problems of standardization of the acquisition and interpretation of ultrasound images. The exercises followed were focused on metacarpophalangeal (MCP) joint synovitis in RA using the OMERACT definition and the same type of ultrasound machine. These iterative projects tested and retested interobserver and intraobserver reliability for both interpretation and acquisition of images of the MCP joint, including a newly introduced semiquantitative scoring at the joint level that combined PD, effusion, and synovial hypertrophy⁸. Results confirmed that the OMERACT ultrasound definitions and scoring system of synovitis combined with a standardized acquisition protocol provided good intra- and interobserver reliability. This work was presented at OMERACT 8⁹.

The validity of this scoring system at the MCP level was then tested by using different machines and multiple observers. The results varied according to the machine used, and confirmed the good reliability among experts concerning the definitions and the scoring system¹⁰. The definition and scoring system of synovitis were also evaluated on other joints commonly involved in RA, including static images of proximal interphalangeal (PIP), wrist, metatarsophalangeal (MTP), and knee joints. Interobserver reliability results ranged from good to excellent for both single pathologies and combined synovitis definitions¹¹. Two reliability studies of the shoulder were also conducted to compare ultrasound with magnetic resonance imaging for the detection of inflammatory and mechanical pathologies and to assess the intra- and interobserver reliability of ultrasound for detecting these lesions. The results were good, but underlined the difficulties for detecting and scoring mechanical-related abnormalities, such as rotator cuff lesions^{12,13}. For finalizing work on RA synovitis the group decided to develop recommendations for detecting and scoring synovitis including the production of an atlas with representative images for different joints (manuscript in preparation).

Finally, through group discussions and feedback sessions, the concept of developing an ultrasound-based global synovitis score in RA (RA GLOSS) was agreed on. This decision was based on the need for integrating the components of synovitis (i.e., synovial hypertrophy, effusion, and synovial Doppler signal) in a unique global score of joint inflammatory activity in RA. This ultrasound scoring sys-

tem at patient level would permit physicians to objectively follow patients under treatment in clinical and research practice, by using a feasible and economical tool that would be more representative of disease activity in RA patients than conventional clinical measures. In addition, the group decided to focus research on standardization of enthesitis in spondyloarthritis (SpA) and joint-related abnormalities in osteoarthritis (OA).

At OMERACT 10 in Kota Kinabalu, Malaysia, in May 2010, during a Special Interest Group session, the results of activities post-OMERACT 9 were presented and the future research agenda was discussed and endorsed by participants, as follows.

1. Intra- and interobserver reliability of ultrasound detection and scoring of synovitis in different RA joints

The group presented the intra- and interobserver reliability results of an exercise for scoring synovitis in non-MCP RA joints. After a consensus session on a synovitis scoring system at the joint level and training on images, 6 RA patients were scanned by 12 observers (2 rounds). The following joints were included: PIP, wrists, knees, tibiotalar, and MTP joints. Intra- and interobserver reliability for detecting and scoring effusion, synovial hypertrophy, PD signal, and PD ultrasound (PDUS) global synovitis (joint-level) were assessed according to OMERACT rules and usual ultrasound practice using standard κ -coefficient and weighted κ -coefficient with absolute weighting [$\kappa(w)$] for semiquantitative grades. According to Landis and Koch¹⁴, κ values < 0.40 were considered poor, 0.40–0.60 moderate, 0.60–0.80 good, and 0.80–1 excellent. The results showed a moderate to excellent intraobserver reliability (κ -coefficient range 0.43–0.98) and a moderate to good interobserver reliability (κ -coefficient range 0.49–0.76) for detection of synovitis components (i.e., effusion, synovial hypertrophy, and PD signal) and grading. In addition, intra- and interobserver reliability for both PDUS global synovitis according to OMERACT rules and usual ultrasound clinical practice were good to excellent (κ -coefficient range 0.58–0.86). This demonstrated that OMERACT rules for PDUS scoring of MCP synovitis were valuable for detecting and scoring synovitis in other joints, and they are now implemented in usual ultrasound practice.

2. Ultrasound-based global synovitis score in RA (RA GLOSS)

Moving from joint level to patient level, the group developed an OMERACT ultrasound global synovitis score. For this objective, further evaluation is needed of how many and what joints should be assessed and how to produce a cumulative score. The results of a systematic review in the field of previous ultrasound scoring systems of synovitis in RA were presented (e.g., joint studied, grading used, construct and predictive validity, discrimination, and feasibility) dur-

ing the Sharp symposium¹⁵. The results from 2 published multicenter longitudinal studies were discussed in more detail. The first, from Naredo and colleagues, compared an ultrasound 44-joint count with a number of reduced counts. It concluded that a 12-joint count was feasible and sensitive to change¹⁶. The second, from Backhaus and colleagues, proposed a 7-joint synovial recess count¹⁷. Both reduced joint counts demonstrated good responsiveness; however, correlation between ultrasound scores and clinical and laboratory findings varied according to the number and size of joints examined. The group agreed that both reduced joint counts required further testing in other multicenter longitudinal studies that also applied the recently devised OMER-ACT scoring system. For this purpose, a multicenter international study is under way in order to test, first, the sensitivity to change of the OMERACT synovitis scoring system at joint level, and second, to validate a reduced joint count (12 or 7 joints).

During the group discussions and feedback session, a need was expressed to develop the RA GLOSS systems by separating diagnostics from monitoring. For each system, it would be important to know what ultrasound findings and which joints should be assessed. For diagnosis, a large number of joints might need to be assessed, while for monitoring a smaller number is required.

3. Spondyloarthritis-associated enthesitis systematic review and Delphi definition exercise

A number of scoring methods for the peripheral enthesitis in SpA have been described^{18,19,20,21}, with limited data on their psychometric properties. Our group focused on standardization of peripheral enthesitis scores. Preliminary work on standardizing ultrasound enthesitis had already been performed by some members of the group^{22,23}, and this was used as a starting point. A systematic literature review of definition of enthesitis and its component lesions was presented during OMERACT 10. The results showed that most studies used variable definition of enthesitis and its components, making comparison of results difficult. Doppler assessment was included in only a few studies. Reliability, sensitivity, and specificity, and responsiveness were assessed in a minority of studies. A number of enthesitis scoring systems at the patient level have been used but all need consensus and validation. In order to produce a consensual definition of ultrasound enthesitis and elementary components, a Delphi exercise was conducted; the exercise focused on ultrasound elementary lesions and definitions among group members (26 ultrasonographers). The Delphi questionnaire included 4 main areas: (1) Definition of normal ultrasound enthesitis and other anatomical structures; (2) Which lesions in GS and PD to include; (3) Ultrasound definition of elementary lesions; and (4) Which lesions reflect inflammation and damage. There was high agreement (> 80%) for definitions of normal enthesitis and normal bursa

and for separating both. There was also high agreement for assessing the following ultrasound elementary lesions: hypoechogenicity and increased thickness of tendon insertion, enthesophytes, calcifications, erosions, and Doppler signal at enthesitis insertion.

The number of entheses to evaluate was discussed but it was suggested that we should evaluate 6 bilateral areas. These included quadriceps insertion, patellar tendon (proximal and distal), Achilles tendon, plantar fascia, and lateral epicondyle. In terms of scoring systems, agreement was obtained to develop one for diagnostic purposes and one for monitoring.

4. Preliminary results of an enthesitis intra- and inter-observer reliability exercise in SpA patients

The group presented the intra- and interobserver reliability results of an exercise for detecting elementary components of enthesitis in SpA patients. Prior to this exercise, a consensus session on which lesions and sites to include and a training session on images were conducted. Six SpA patients were scanned by 12 observers (2 rounds). The assessed elementary components were as follows: cortical erosions, enthesophytes, increased thickness, hypoechogenicity, PD at cortical insertion, PD outside insertion, bursitis, and calcifications. The following bilateral sites were assessed for the presence of the above components and global enthesitis: epicondyle, Achilles tendon, and patellar ligament (both insertions). Intra- and interobserver reliability for detecting the above components were assessed using standard κ -coefficient and weighted κ -coefficient, with absolute weighting [$\kappa(w)$] for semiquantitative grades¹⁴. The results showed a variable intra- and interobserver reliability for detecting the elementary components (better for PD than for GS). However, enthesitis as global definition was more reliable than single components (interobserver κ -coefficient, 0.67; intraobserver κ -coefficient range 0.45–0.80).

The results of an additional Web-based intra- and inter-reader reliability exercise on ultrasound detection of SpA enthesitis were also presented. We tested the elementary components and the definition of enthesitis in a large group of sonographers (18 participants, 120 ultrasound images). The results were variable but comparable to the reliability exercise in patients (interobserver κ -coefficient, 0.61; intraobserver κ -coefficient range 0.20–0.70).

5. Preliminary results of a Delphi definition exercise in hand OA and subsequent reliability exercise

The ultrasound group had previously decided to focus on standardization of joint-related pathologies in OA. After a systematic review of the literature²⁴, an OMERACT/OARSI US sub-task force focused on hand OA. A Delphi exercise focusing on ultrasound elementary lesions and definitions in OA was conducted among group members (21 sonographers). The Delphi process aimed at reaching consensus

on ultrasound elementary lesions, definitions, and ultrasound scanning technique. The results showed high to good agreement (> 80%) for evaluating cartilage, osteophytes, bone erosions, cortical irregularities, synovial membrane, and synovial fluid. It was suggested to differentiate ultrasound inflammation and structural damage.

The group undertook 2 reliability exercises in patients with early and late hand OA. There was a high variability in the detection of all elementary lesions in the former exercise, with moderate improvement in detection of some structural damage lesions in the second exercise. The results of these exercises highlighted specific problems of interpretation of cartilage lesions and cortical abnormalities.

FUTURE ACTIVITIES

Rheumatoid Arthritis; Synovitis

A multicenter international study is currently under way that aims to test the responsiveness of the composite RA OMERACT synovitis scoring system (GS + PD) as applied to different core sets of joints. In addition, group work has begun to focus on the use of ultrasound for evaluating tenosynovitis and structural damage (i.e., bone erosions and tendon damage) in RA. Testing the added value, predictive value in relation to disease-centered (e.g., erosions), and patient-centered (e.g., function) outcomes, as well as the diagnostic value of ultrasound in clinical trials, is a priority objective. Future research directions also include further validation of global ultrasound scoring of synovitis at the patient level for diagnostic purposes and for monitoring in RA.

Spondyloarthritis Enthesitis

After reaching consensus on preliminary selection and definition of the elementary lesions that should be evaluated in peripheral SpA enthesitis, the group is investigating scoring systems at joint level and at patient level with differentiation of elementary lesions useful for activity assessment and damage assessment. The group is also planning further research on the development of reliable and feasible scoring systems for early diagnosis and monitoring of SpA.

Osteoarthritis

Agreement on OA lesions using an image-based Web exercise is being pursued in order to improve ultrasound reliability. The group is also testing the reliability of ultrasound detection and grading of tenosynovitis in RA. The development of a global ultrasound scoring system to use in a longitudinal observational study in early hand and knee OA is also on the agenda.

REFERENCES

- Karim Z, Wakefield RJ, Conaghan PG, Lawson CA, Goh E, Quinn MA, et al. The impact of ultrasonography on diagnosis and management of patients with musculoskeletal conditions. *Arthritis Rheum* 2001;44:2932-3.
- D'Agostino MA, Ayral X, Baron G, Ravaud P, Breban M, Dougados M. Impact of ultrasound imaging on local corticosteroid injections of symptomatic ankle, hind- and mid-foot in chronic inflammatory diseases. *Arthritis Rheum* 2005;53:284-92.
- Joshua F, Lassere M, Bruyn GA, Szkudlarek M, Naredo E, Schmidt WA, et al. Summary findings of a systematic review of the ultrasound assessment of synovitis. *J Rheumatol* 2007;34:839-47.
- 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.
- Wakefield RJ, D'Agostino MA, Iagnocco A, Filippucci E, Backhaus M, Scheel AK, et al. The OMERACT Ultrasound Group: status of current activities and research directions. *J Rheumatol* 2007;34:848-51.
- Scheel AK, Schmidt WA, Hermann KG, Bruyn GA, D'Agostino MA, Grassi W, et al. Interobserver reliability of rheumatologists performing musculoskeletal ultrasonography: results from a EULAR "Train the Trainers" course. *Ann Rheum Dis* 2005;64:1043-9.
- Naredo E, Möller I, Moragues C, de Agustín JJ, Scheel AK, Grassi W, et al. Interobserver reliability in musculoskeletal ultrasonography: results from a "Teach the Teachers" rheumatologist course. *Ann Rheum Dis* 2006;65:14-9.
- D'Agostino MA, Wakefield R, Filippucci E, Backhaus M, Balint PV, Bouffard JA, et al. Intra- and inter-observer reliability of ultrasonography for detecting and scoring synovitis in rheumatoid arthritis: A report of a EULAR ESCISIT Task Force [abstract]. *Ann Rheum Dis* 2005;64:62.
- D'Agostino MA, Conaghan PG, Naredo E, Aegerter P, Iagnocco A, Freeston JE, et al. The OMERACT ultrasound task force — Advances and priorities. *J Rheumatol* 2009;36:1829-32. Erratum in: *J Rheumatol* 2009;36:2625.
- Wakefield R, D'Agostino MA, Backhaus M, Balint PV, Bruyn GAW, Filippucci E, et al. Combined evaluation of influence of sonographer and machine type on the reliability of power Doppler ultrasonography (PDUS) for detecting, scoring and scanning synovitis in rheumatoid arthritis (RA) patients: results of an intermachine reliability exercise of the EULAR/OMERACT US group [abstract]. *Ann Rheum Dis* 2008;67:421.
- D'Agostino MA, Wakefield R, Backhaus M, Balint PV, Bruyn GAW, Filippucci E, et al. Evaluation of the intra- and inter-reader reliability of power Doppler ultrasonography (PDUS) for detecting and scoring synovitis of several joints in rheumatoid arthritis (RA) by using a validated scoring system: a report of the EULAR/OMERACT US group [abstract]. *Ann Rheum Dis* 2008;67:422.
- Bruyn GA, Naredo E, Moller I, Moragues C, Garrido J, de Bock GH, et al. Reliability of ultrasonography in detecting shoulder disease in patients with rheumatoid arthritis. *Ann Rheum Dis* 2009;68:357-61.
- Bruyn GA, Pineda C, Diaz CH, Ventura-Rios L, Moya C, Garrido J, et al. Validity of ultrasonography and measures of adult shoulder function and reliability of ultrasonography in detecting shoulder synovitis in patients with rheumatoid arthritis using MRI as a gold standard. *Arthritis Care Res* 2010;62:1079-86.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- Mandl P, Naredo E, Wakefield RJ, Conaghan PG, D'Agostino MA. A systematic literature review analysis of ultrasound joint count and scoring systems to assess synovitis in rheumatoid arthritis according to the OMERACT filter. *J Rheumatol* 2011;38:2055-62.
- Naredo E, Rodríguez M, Campos C, Rodríguez-Heredía JM, Medina JA, Giner E, et al. Validity, reproducibility, and responsiveness of a twelve-joint simplified power Doppler

- ultrasonographic assessment of joint inflammation in rheumatoid arthritis. *Arthritis Rheum* 2008;59:515-22.
17. Backhaus M, Ohrndorf S, Kellner H, Strunk J, Backhaus TM, Hartung W, et al. Evaluation of a novel 7-joint ultrasound score in daily rheumatologic practice: A pilot project. *Arthritis Rheum* 2009;61:1194-201.
 18. Balint PV, Kane D, Wilson H, McInnes IB, Sturrock RD. Ultrasonography of enthesal insertions in the lower limb in spondyloarthropathy. *Ann Rheum Dis* 2002;61:905-10.
 19. D'Agostino MA, Said-Nahal R, Hacquard-Bouder C, Brasseur JL, Dougados M, Breban M. Assessment of peripheral enthesitis in the spondyloarthropathies by ultrasonography combined with power Doppler. *Arthritis Rheum* 2003;48:523-33.
 20. Alcalde M, Acebes JC, Cruz M, González-Hombrado L, Herrero-Beaumont G, Sánchez-Pernaute O. A sonographic enthesitis index (SEI) at lower limbs is a valuable tool in the assessment of ankylosing spondylitis. *Ann Rheum Dis* 2007;66:1015-9.
 21. De Miguel E, Cobo T, Muñoz-Fernández S, Naredo E, Usón J, Acebes JC, et al. Validity of enthesitis ultrasound assessment in spondyloarthropathy. *Ann Rheum Dis* 2009;68:169-74.
 22. D'Agostino MA, Aegerter P, Jousse-Joulin S, Chary-Valckenaere I, Lecoq B, Gaudin P, et al. How to evaluate and improve the reliability of power Doppler ultrasonography for assessing enthesitis in spondyloarthritis. *Arthritis Rheum* 2009;61:61-9.
 23. Filippucci E, Aydin SZ, Karadag O, Salaffi F, Gutierrez M, Direskeneli H, et al. Reliability of high-resolution ultrasonography in the assessment of Achilles tendon enthesopathy in seronegative spondyloarthropathies. *Ann Rheum Dis* 2009;68:1850-5.
 24. Keen HI, Wakefield RJ, Conaghan PG. A systematic review of ultrasonography in osteoarthritis. *Ann Rheum Dis* 2009;68:611-9.