The OMERACT Ultrasound Working Group 10 Years On: Update at OMERACT 12

George A. Bruyn, Esperanza Naredo, Annamaria Iagnocco, Peter V. Balint, Marina Backhaus, Frederique Gandjbakhch, Marwin Gutierrez, Andrew Filer, Stephanie Finzel, Kei Ikeda, Gurjit S. Kaeley, Silvia Magni Manzoni, Sarah Ohndorf, Carlos Pineda, Bethan Richards, Johannes Roth, Wolfgang A. Schmidt, Lene Terslev, and Maria Antonietta D’Agostino, on behalf of the OMERACT Ultrasound Task Force

**ABSTRACT.** Musculoskeletal ultrasound (US) now thrives as an established imaging modality for the investigation and management of chronic inflammatory arthritis. We summarize here results of the Outcome Measures in Rheumatology (OMERACT) US working group (WG) projects of the last 2 years. These results were reported at the OMERACT 12 meeting at the plenary session and discussed during breakout sessions. Topics included standardization of US use in rheumatic disease over the last decade and its contribution to understanding musculoskeletal diseases. This is the first update report of WG activities in validating US as an outcome measure in musculoskeletal inflammatory and degenerative diseases, including pediatric arthritis, since the OMERACT 11 meeting. (First Release March 15 2015; J Rheumatol 2015;42:2172–6; doi:10.3899/jrheum.141462)

**Key Indexing Terms:**

**Joint erosions**  **Outcome assessment**  **Ultrasoundography**

As of 2015, musculoskeletal ultrasound (US) can no longer be considered as controversial in rheumatology; on the contrary, US thrives as an established imaging modality for the investigation and management of chronic inflammatory arthritis. Last year marked the 10-year jubilee of the OMERACT US working group (WG). Members of the WG met in Budapest, Hungary, for the OMERACT 12 conference, where results of the last 2 years of ongoing projects were presented. The several milestones reached in standardizing the use of US in rheumatic disease over the last decade and the contribution of US to understanding musculoskeletal diseases were highlighted in the plenary session and discussed during breakout sessions. This report provides an update on the activities of the WG in validating US as an outcome measure in musculoskeletal inflammatory and degenerative diseases including pediatric arthritis, since the last report on WG activities at OMERACT 11.1
A Decade Put into Historical Perspective

At OMERACT 7 in 2004, a special interest group (SIG) dedicated to US was formed by a group of international rheumatologists with the aim of exploring the metric properties of musculoskeletal US. At this early stage, a systematic review of the musculoskeletal US literature in rheumatoid arthritis (RA) dissected the various gaps in existing knowledge, particularly underscoring the lack of US definitions of rheumatic pathology, instrument reliability, and instrument validity. Overall agreement was that because research resources of the SIG were limited, efforts had to be strictly prioritized. The very first publication of the group reported on a core set of practical US definitions for general rheumatic manifestations including synovitis, tenosynovitis, and erosions. In considering which strategy to use, iterative exercises on synovitis in patients with RA were carried out from 2004 to 2010. These exercises involved US assessment of synovitis at both the patient level and the joint level. It was not surprising that the intra- and interexaminer \( \kappa \) values for reading still images were better than for those of real-time image acquisition.

By 2008, the perspective of developing an US disease activity score based on synovitis at the patient level loomed as a logical next step, i.e., a global synovitis score (GLOSS). Development of a GLOSS was the result of an iterative, gradual, slow-moving process, implicating a step-by-step approach that included several issues, e.g., the optimal number of joints, how to scan these (dorsal, volar, and B-mode alone or in combination with power Doppler). On the basis of favorable results of the preceding exercises, an US-GLOSS, combining B-mode synovial hypertrophy and power Doppler in 1 score, was presented at OMERACT 10. An additional advantage is that the GLOSS can be performed à la carte, i.e., in various joint number configurations. Subsequently, responsiveness of the GLOSS was tested in an international multicenter open-label medication trial evaluating responsiveness of power Doppler US in patients with RA with incomplete clinical response to methotrexate and treated with abatacept. Preliminary results were reported at OMERACT 11. During the group discussions and feedback sessions, a need for separate development of diagnostic and monitoring RA GLOSS systems was expressed. Currently, questions need to be addressed on which US findings are strictly prioritized. The very first publication of the group reported on a core set of practical US definitions for general rheumatic manifestations including synovitis, tenosynovitis, and erosions. In considering which strategy to use, iterative exercises on synovitis in patients with RA were carried out from 2004 to 2010. These exercises involved US assessment of synovitis at both the patient level and the joint level. It was not surprising that the intra- and interexaminer \( \kappa \) values for reading still images were better than for those of real-time image acquisition.

As mentioned in the preceding report of OMERACT 11, testing the metric properties of US on tenosynovitis and tendon damage in patients with RA was another prioritized research area. From a clinical point of view, tendon damage may be an important endpoint in RCT; it would also be clinically relevant to understand which US findings at joint and tendon level are able to predict tendon damage. Results of the tendon damage study in patients with RA showed good to excellent \( \kappa \) values for intraobserver and interobserver reliability. Additionally, an atlas of US images on tenosynovitis and tendon damage in RA was published as online material.


During the workshop, the ongoing research agenda focused on additional data including the validation of US in RA erosions and in pediatric arthritis, as well as on new development of US as an outcome measure for other inflammatory rheumatic diseases, such as psoriatic arthritis (PsA) and gout. These topics were first presented in the plenary introduction and then discussed in the breakout sessions.

The first topic focused on the validation of US for detecting RA bone erosions. S. Finzel presented new findings on the prevalence of erosions versus normal cortical “breaks” in patients with RA and healthy controls, using high-resolution peripheral quantitative computed tomography as the gold standard. The rationale of these studies is to get a better idea of what a “true US erosion” represents. Subsequently, the intraobserver and interobserver reliability of US detecting these structures was tested in patients with RA and healthy controls by 12 rheumatologists expert in US (Table 1). Based on the outcome of this study, further studies are planned to define an US-detected RA erosion and the minimal size that can be accurately detected.

Next, a presentation by L. Terslev provided insights into how US can assess the 3 key domains in gout, i.e., inflammation, damage, and urate load. By using a previous systematic literature review, 4 elementary US components were identified, i.e., double contour sign, aggregates, tophi, and erosion. The US definitions of these 4 identified lesions were agreed upon by the group using a Delphi exercise. Subsequently, the metric properties of these components were

| Table 1. US testing intraobserver and interobserver reliability on small erosions in patients with RA and healthy controls. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | Intraobserver   | Interobserver   |
|                | Normal Break    | Abnormal Break  | Normal Break    | Abnormal Break  |
| Palmar long    | 0.1–0.8         | 0.6–0.9         | 0.8             | 0.9             |
| Dorsal long    | 0.4–0.6         | –0.1–0.5        | 0.4             | 0.7             |
| Palmar transv  | 0.2–0.7         | 0.5–0.7         | 0.7             | 0.5             |
| Dorsal transv  | 0.3–0.9         | 0.4–0.8         | 0.3             | 0.4             |

US: ultrasound; RA: rheumatoid arthritis.
assessed in a patient reliability study conducted in Berlin, December 2013. Preliminary results were presented, showing acceptable intraobserver reliability for detecting and acquiring images of double contour, tophi, and erosions, but not for aggregates. Interobserver κ values were even lower16. On the basis of the reliability results, overall agreement was that further validation was needed for double contour sign and aggregates.

A. Iagnocco presented work conducted in hand osteoarthritis (OA). Results of a reliability study focusing on cartilage damage showed intrarater and interrater κ of 0.52 and 0.80 using dichotomous scoring19. A second reliability exercise was aimed at evaluating the possibility to grade together structural damage in hand OA, by using a semiquantitative grading of both cartilage and osteophyte lesions. This study showed good results for osteophyte scoring, but moderate for cartilage20. Overall agreement was that an US core domain set to be used in hand OA structural lesions should include cartilage scoring in a dichotomous way and osteophyte scoring on a semiquantitative scale (0–3).

J. Roth presented the latest concepts of how US can be used as an instrument for assessment of pediatric pathology. A core domain set for pediatric pathology has yet to be determined. The US definitions of joints of healthy children have recently been published21. The next step is to define synovitis in children with juvenile idiopathic arthritis (JIA), which shall be done by consensus through consecutive Delphi rounds. The main objective of the pediatric Delphi process is to obtain consensus on the B-mode and Doppler US elementary components to include in the definition of synovitis in children. The secondary objective is to obtain consensus on the type of scoring system that will be developed. Both the synovitis definition and the scoring system will subsequently be tested in future US exercises in children with JIA.

The last topic was dactylitis, presented by G. Kaeley. He explained that dactylitis was identified as part of a domain core set for PsA. US candidate elementary components have been identified through a literature review22. A Delphi process is under way to reach consensus on the initial set of elements that warrant study. Based on the results of the first round, the candidate elements were prioritized (Figure 1). A second round of the Delphi process is being conducted to plan a reliability exercise looking at evaluating the identified elementary lesions.

Following these presentations, each subgroup was divided into smaller discussion groups (about 15 participants each, including 2 patient partners), who were then asked to consider a set of 4 draft questions based on endorsement of the work done and the future research agenda of the group by OMERACT participants (Table 2). Draft questions pertained to construct validity of hand OA, a core domain set of US to be used in gout patients, a core domain set to be used in patients with PsA, and lastly, future research in RA erosions. Each discussion group then reported its main points to all participants at the end of the breakout sessions.

Following this report, the questions were voted on for

![Figure 1. Categorization of candidate elements in dactylitis into domains after first round of Delphi process. Nail plate Tri Loss / Irrg: loss of nail plate trilaminar structure and/or irregularity of nail plate; Intracap calcification: intracapsular calcification; Lig: ligament; Tend.: tendon; Enth.: enthesitis.](www.jrheum.org)
potential endorsement by all conference participants at the final plenary session on the last day of the conference. The topics proposed in the formulated questions were endorsed by a strong majority of attendees.

Below, the main points of discussion are reported.

Regarding the US detection of erosions, there was widespread recognition of the importance of developing an US validated measure of erosions, since this tool is widely introduced in the evaluation of RA synovitis. Participants agreed that the evaluation of erosions by US would provide valuable support for early detection of erosive disease. In addition, the higher sensitivity of US for detecting erosions compared to radiography, owing to its better resolution and to the tomographic nature of the technique, is considered an added value. The detection of erosions in early inflammatory disease was felt to be a priority research area and an objective to be tested in future clinical trials. However, additional validation was required before proposing US as a standard outcome measurement of structural damage. For example, more data on the discriminative capability of US for distinguishing between normal cortical breaks and small erosions is needed. One breakout group pointed out the need for additional RCT supporting the responsiveness of inflammatory findings, such as synovitis, before moving to structural damage. Nevertheless, general agreement was expressed on the potential interest of this tool in evaluating erosions.

There was also strong agreement that US is a valuable tool for evaluation of patients with gout. Based on discussions in the breakout groups, several key points were raised by participants, especially as related to the role of US in gout. The importance of US in evaluating urate load was underscored. Participants agreed on the valuable role of US in distinguishing and measuring acute and chronic gout and in identifying core domains for both stages of disease (tophi, synovial inflammation, aggregates, and urate deposits). However, there remains a lack of clear definitions of elementary lesions detected by US. Therefore, discussions were mostly related to which lesions should be assessed by US and which definitions should be used. The discriminative ability of US gout lesions in comparison to other arthropathies has been suggested as a priority for validation.

The third question, based on the development of US in PsA, also received agreement from the majority of participants. In each breakout group in which this topic was discussed, unanimous concordance on the need to pursue standardization of US for management of PsA was reached. The value of US in the evaluation of PsA synovitis was recognized and supported, as well as the potential value of US in the evaluation of dactylitis. The development of US as a responsive tool for following this clinical manifestation was unanimously supported. Finally, the potential development of a structural US score in hand OA was discussed. On the basis of the work already performed by the WG in terms of inflammatory abnormalities, agreement was obtained that the future research agenda should focus on correlations between structural and inflammatory lesions and clinical outcomes in symptomatic hand OA.

The objectives of this workshop were to present both the existing knowledge on the use of US in areas that have been explored over the last decade and to decide priorities for future research. US is a unique outcome measure that reveals both the past and present status of various rheumatic diseases. Considerable progress has been reported in different areas, including synovitis and structural damage in RA, tenosynovitis in RA, and structural damage in hand OA. At this stage it is not possible to predict the influence of the workshop's success in these areas, but the effects may be far-reaching, both for daily practice and clinical research. Examples of the aspect of daily practice may be other treatment expectations or less use of radiographic radiation; an example of the clinical research aspect may be novel insights into pathogenetic mechanisms, e.g., in OA.

Here follows the research agenda drafted to address existing gaps in our knowledge regarding work to be done in hand OA, gout, PsA, and erosions (Table 3):

- To investigate the construct validity of US assessment of hand OA as compared to clinical manifestations of the disease
- To assess the metric properties of US in other OA joints (e.g., the knee)
- To further define the basic abnormalities evaluable with US in gout and to test the reliability, responsiveness, and discriminant capacity of these lesions
- To further identify and define the basic US abnormalities that can be included in the US assessment of PsA and to test their metric properties
- To further address the concurrent validity and sensitivity to change of US-detected early bone erosions
- To develop definitions for joint inflammatory pathology in childhood

Other areas of future research include systemic vasculitis, synovial biopsy, and knee OA. Over the next 2 years, fresh data will be reported on the different topics of the research agenda.
Table 3. Future research agenda and time line of the OMERACT US working group.

<table>
<thead>
<tr>
<th>US Research Field</th>
<th>Outcome</th>
<th>Research Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection of minimal erosions in RA</td>
<td>Minimal erosion</td>
<td>Validity studies</td>
</tr>
<tr>
<td>Definition of an US core domain set in gout</td>
<td>Inflammation, damage, urate load</td>
<td>Delphi study on definitions/reliability</td>
</tr>
<tr>
<td>Definition of an US core domain set in hand OA</td>
<td>Structural lesions</td>
<td>Delphi study/reliability</td>
</tr>
<tr>
<td>Definition of an US core domain set in PsA</td>
<td>Dactylitis</td>
<td>Delphi study</td>
</tr>
<tr>
<td>Definition of an US core domain set in pediatric arthritis</td>
<td>Synovitis</td>
<td>Delphi study on normal sonoanatomy and synovitis</td>
</tr>
</tbody>
</table>

US: ultrasonography; RA: rheumatoid arthritis; OA: osteoarthritis; PsA: psoriatic arthritis.

APPENDIX


REFERENCES