A Systematic Literature Review Analysis of Ultrasound Joint Count and Scoring Systems to Assess Synovitis in Rheumatoid Arthritis According to the OMERACT Filter

PETER MANDL, ESPERANZA NAREDO, RICHARD J. WAKEFIELD, PHILIP G. CONAGHAN, and MARIA ANTONIETTA D'AGOSTINO, on behalf of the OMERACT Ultrasound Task Force

ABSTRACT. ***Objective.*** The OMERACT Ultrasound Task Force is currently developing a global synovitis score (GLOSS) with the objective of feasibly measuring global disease activity in patients with rheumatoid arthritis (RA). In order to determine the minimal number of joints to be included in such a scoring system, and to analyze the metric properties of proposed global (i.e., patient level) ultrasound (US) scoring systems of synovitis in RA, a systematic analysis of the literature was performed.

Methods. A systematic literature search of Pubmed and Embase was performed (January 1, 1984, to March 31, 2010). Original research reports written in English including RA, ultrasound, Doppler, and scoring systems were included. The design, subjects, methods, imaging protocols, and performance characteristics studied were analyzed, as well as the ultrasound definition of synovitis.

Results. Of 3004 reports identified, 14 articles were included in the review. We found a lack of clear definition of synovitis as well as varying validity data with respect to the proposed scores. Scoring systems included a wide range and number of joints. All analyzed studies assessed construct validity and responsiveness by using clinical examination, laboratory findings, and other imaging modalities as comparators. Both construct validity and responsiveness varied according to the number and size of joints examined and according to the component of synovitis measured [i.e., gray-scale (GS) or power Doppler (PD) alone or in combination]. With regard to feasibility, time of evaluation varied from 15 to 60 min and increased with the number of joints involved in the examination.

Conclusions. Ultrasound can be regarded as a valuable tool for globally examining the extent of synovitis in RA. However, it is presently difficult to determine a minimal number of joints to be included in a global ultrasound score. Further validation of proposed scores is needed. (J Rheumatol 2011; 38:2055–62; doi:10.3899/jrheum.110424)

Key Indexing Terms:
SYSTEMATIC LITERATURE REVIEW SCORING SYSTEM ULTRASOUND
POWER DOPPLER RHEUMATOID ARTHRITIS SYNOVITIS

Musculoskeletal ultrasound is primarily used by rheumatologists for detecting and assessing inflammation of joints and joint damage in rheumatoid arthritis (RA)[1]. Specifically, ultrasound is capable of evaluating the 2 elementary findings associated with synovitis: synovial hypertrophy (SH) and synovial fluid/effusion (SF)[2]. SF is visualized as an anechoic area within the joint capsule, while SH is visualized as hypoechoic material within and involving the joint capsule. Ultrasound has been shown to be superior to clinical examination in detecting and evaluating these 2 crucial components in a range of studies performed on various joints[3,4]. Both SF and SH are evaluated primarily on gray-scale (GS) ultrasound, while Color Doppler (CD) and Power Doppler (PD) are utilized to demonstrate activity related to SH. By visualizing the intravascular movement of blood cells, CD and PD detect microvascular blood flow in synovial and entheseal inflammation[5,6].

At the single-joint level, synovitis and effusion in GS were initially evaluated by binary grading (presence/absence)[7]. This was followed by the creation of several semiquantitative scoring systems[8,9], which rated individual synovial hyperthrophy and effusion as well as the combination thereof. Additional studies utilized quantitative measurements for evaluating synovitis in GS, based on the volume/depth of synovial tissue[8,10].

From the National Institute of Rheumatology and Physiotherapy, Budapest, Hungary; Hospital Universitario Severo Ochoa, Madrid, Spain; Section of Musculoskeletal Disease, University of Leeds and NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK; and Université Paris Île de France Ouest-Versailles-Saint Quentin en Yvelines, Hôpital Ambroise Paré, APHP, Boulogne-Billancourt, France.

P. Mandl, MD, PhD, National Institute of Rheumatology and Physiotherapy; E. Naredo, MD, Hospital Universitario Severo Ochoa; R.J. Wakefield, MD, Section of Musculoskeletal Disease, University of Leeds; P.G. Conaghan, MBBS, PhD, FRACP, FRCP, Section of Musculoskeletal Disease, University of Leeds and NIHR Leeds Musculoskeletal Biomedical Research Unit; M.A. D'Agostino, MD, PhD, Université Paris Île de France Ouest-Versailles-Saint Quentin en Yvelines, Hôpital Ambroise Paré, APHP.

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

Binary grading[11] as well as a semiquantitative scoring system[9] were also developed for evaluating vascularization of synovial hypertrophy by CD/PD. Quantitative approaches are also utilized, including the determination of the number of Doppler pixels by dedicated pixel-counting software[12,13,14,15], measurement of pulsatility and resistive indices, and microbubble contrast material.

However, several factors are known to influence the sensitivity of detecting synovitis by GS and PD ultrasound. Machine characteristics and resolution, as well as varying parameter settings and the use of different transducers and presets may have significant effects on sensitivity[16]. Differentiating between normal and abnormal joints is further complicated by the fact that certain normal joints (e.g., the knee) may contain small amounts of SF and SH[17]. Doppler activity in a joint that is otherwise considered normal is less common, and is mostly due to greater intermachine variation in sensitivity with respect to Doppler[13,18].

In addition to machine-dependent effects, operator-dependent factors, including factors affecting both acquisition and interpretation, have to be considered. In order to improve acquisition and interpretation, the EULAR imaging working group has established a consensual acquisition protocol[19] for individual joints. Subsequently, preliminary consensus definitions for ultrasound for common pathological lesions seen in patients with inflammatory arthritis, including synovitis and intraarticular effusion, have been established by the OMERACT Ultrasound Task Force[20].

Our Task Force has worked towards the development of a reliable standardized scoring system for synovitis in RA that is applicable to all joints and is consistent between machines, and which combines GS and PD in a semiquantitative 0–3 scale[21,22]. Results confirmed that a consensus scoring system of synovitis based on consensus definitions, combined with a standardized acquisition protocol, provided good intra- and inter-reliability[21,22]. In order to be able to make assumptions on global disease activity, it is necessary to move from the level of single joints to the level of the patient as a global entity. The objective of the group is therefore to propose a global ultrasound scoring system of synovitis in RA at the patient level.

However, at the moment, there is a lack of consensus regarding the optimal number of joints to evaluate and the appropriate components and scoring to use at joint level. In order to determine the minimal number of joints and the appropriate scoring system to include to correctly assess RA patients by ultrasound, we analyzed the validity of proposed global (i.e. patient level) ultrasound scoring systems according to the OMERACT filter within the framework of a systematic review of the literature.

MATERIALS AND METHODS

To extract data on ultrasound scoring systems of synovitis in RA, a systematic search of the literature was performed using a 4-step strategy: (1) Definition of the objective of the review, (2) definition of criteria of selection, (3) selection of articles, and (4) data extraction.

Selection criteria consisted of original articles involving humans, published in English between January 1984 and April 2010, and referring to binary grading, as well as to semiquantitative and quantitative scoring systems and ultrasonography/ultrasound.

Search strategy and study selection. The search of articles was performed in Pubmed and Embase. In Pubmed, article search was performed using the following key words: (Ultrasound OR Ultrasonography OR Power Doppler OR Color Doppler OR Doppler OR Musculoskeletal Ultrasound) AND (Rheumatoid Arthritis OR Inflammatory Arthritis OR Synovitis) AND (Joint count) with limits (language = English, humans only, from January 1st 1984 to March 31st 2010). In Embase, the search was performed using the following key words: (Ultrasound OR Ultrasonography OR Power Doppler OR Color Doppler OR Doppler OR Musculoskeletal Ultrasound) AND (Rheumatoid Arthritis OR Inflammatory Arthritis OR Synovitis) AND (Joint count), with limits (language = English, humans only, from January 1st 1984 to March 31st 2010). For both searches, key words referred to MeSH Terms, or if not available, to key words present in the title/abstract. Titles, abstracts and full reports of articles identified were systematically screened by one author (PM) with regard to inclusion and exclusion criteria. The final search was verified by a second author (MADA). Articles were not included if they were not in English, or studied healthy subjects, or concerned cadavers. In addition, abstracts of scientific congresses and reviews were also excluded (i.e., exclusion criteria). Further, a manual search of secondary sources including article references, reviews, and metaanalysis without limitation of date of publication was also performed.

Data extraction. During data extraction, special attention was given to the "Patients and Methods" and "Results" sections of each article. All data were extracted using a standardized template that was specifically designed for the review. Afterwards, data were collected on an Excel sheet. All selected articles were rated in order to determine the number and choice of evaluated joints in the scoring system, the characteristics of the sytem, and to evaluate the quality of the studies according to the OMERACT filter[23].

Each article was analyzed and assessed in order to determine whether it fulfilled some aspect of validity. We evaluated in particular: face and content validity, construct validity, criterion validity, discriminant validity (i.e., reliability and responsiveness), and feasibility. Each of these criteria was independently evaluated in every article, including whether the methods for assessing it and their measurement were available or not. Moreover, the following characteristics, related to the acquisition and detection of synovitis, were searched for: ultrasound technique in GS and Doppler, ultrasound definition of synovitis if present, modality and components of grading method (SH, SF, or both combined, GS, PD, or both combined), grading system (binary/qualitatively, semiquantitative, quantitative).

A standardized tool for assessment of quality of the analyzed studies based on a set of 6 predefined criteria was developed and assessed in a binary mode (yes/no). These criteria were based on concepts from reviews of quality assessment tools used in systematic reviews of observational studies[24]. The predefined criteria were the following: (1) Was the recruitment of patients well-defined in the methods section? (2) Was the choice of number of joints to include in the scoring mentioned and justified? (3) Was there a description of the ultrasound scanning technique? (4) Was there a description of attempted blinding of observers? (5) Was there a description of synovitis scoring; Which source was this scoring based on? (6) Was the choice of comparator adequately explained and results completely given? Quality was reported on a scale of 0–6, with higher results indicating higher quality. Selected articles scored less than 1 on a scale of 0–6 were excluded from the final analysis.

Statistical analysis. Descriptive statistics were used to report data. Frequencies and percentages were used for categorical variables.

RESULTS

Figure 1 shows a flowchart of the systematic review process.

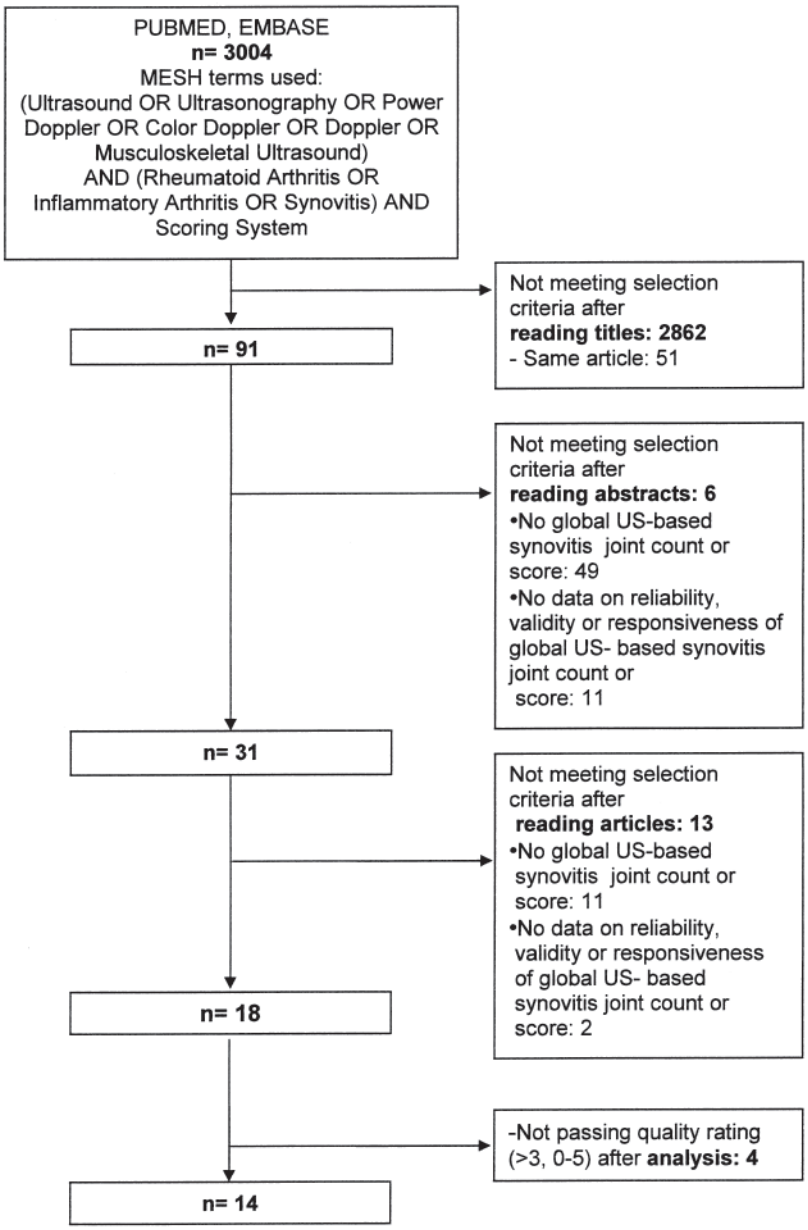

Figure 1. The systematic review process.

Of 3004 studies identified initially, 14 articles were selected to be included in the review[8,9,14,25,26,27,28,29,30,31,32,33,34,35]. Table 1 shows the study design characteristics of the studies. The overwhelming majority featured a blinded design, and sample size ranged between 24 and 278 patients. Only 28% (4 of 14) of articles included control patients. All studies included clinical examination as a comparator for assessing construct validity, with all except one study including laboratory values as well. Imaging modalities (radiography, ultrasound, and MRI) were used in 43% (6 of 14) of the studies as a comparator. In 43% (6 of 14) of articles an arbitrary number of joints are chosen, while 57% (8 out of 14) based their ultrasound evaluation on available clinical indices or frequency of involvement of joints in the disease, according to literature or clinical practice.

Table 2 shows characteristics related to ultrasound examination and scoring. Definition and detection of synovitis at joint level was variable within the studies. One article included GS evaluation of global synovitis without differentiation between SH and SF, without PD evaluation, while another article evaluated only PD activity. The majority of articles evaluated both GS and PD, with GS either evaluated globally or in separate components (SH and SF) in addition to PD activity, which was evaluated separately. No study assessed a

Table 1. Design characteristics of included studies.

Year	Study	Blinded Design	Sample Size	No. Control Patients	Comparator*	No. Joints Assessed by Ultrasound**	Arbitrary Choice of Joints***
2003	Szudlarek[9]	Yes	50	0	Clinical	5	Yes
2004	Taylor[13]	Yes	24	12	Laboratory, x-ray	10	Yes
2005	Naredo[27]	Yes	94	0	Clinical, laboratory	60	No
2005	Scheel[8]	No	46	10	Clinical, laboratory, MRI	8	Yes
2005	Naredo[28]	Yes	94	0	Clinical, laboratory, US	60	No
2005	Naredo[29]	Yes	42	0	Clinical, laboratory, x-ray	28	No
2008	Hameed[30]	Yes	50	25	Clinical, laboratory	20	Yes
2008	Ozgocmen[31]	Yes	54	0	Clinical, laboratory, MRI	14	Yes
2008	Naredo[26]	Yes	160	0	Clinical, laboratory	44	No
2008	Naredo[32]	Yes	278	0	Clinical, laboratory, x-ray	28	Yes
2009	Scire[33]	Yes	106	0	Clinical, laboratory	44	No
2009	Backhaus[25]	Yes	120	0	Clinical, laboratory	7	Yes
2009	Dougados[34]	Yes	76	0	Clinical, laboratory	38	Yes
2010	Balsa[35]	Yes	113	16	Clinical, laboratory	42	No

* Comparator for evaluation of responsiveness or construct validity; ** maximum number of joints assessed by ultrasound within the study; *** arbitrary inclusion of joints in global ultrasound scoring system. US: ultrasound.

Table 2. Characteristics related to ultrasound and scoring.

Year	Study	Ultrasound Definition*	Component Studies**	Binary Grade†	Semiquant. Grade†	Quantitative Grade†	Cumulative Score††	No. US Assessments
2003	Szudlarek[9]	Yes	GS + PD	No	Yes	No	Yes	1
2004	Taylor[13]	No	GS + PD	No	Yes	No	Yes	2
2005	Naredo[27]	Yes	GS + PD	No	Yes	No	Yes	1
2005	Scheel[8]	Yes	GS	Yes	Yes	Yes	Yes	1
2005	Naredo[28]	Yes	GS + PD	Yes	Yes	No	Yes	1
2005	Naredo[29]	Yes	GS + PD	Yes	Yes	No	Yes	1
2008	Hameed[30]	No	GS + PD	Yes	Yes	No	Yes	2
2008	Ozgocmen[31]	Yes	GS + PD	Yes	Yes	No	Yes	1
2008	Naredo[26]	Yes	GS + PD	Yes	Yes	No	Yes	2
2008	Naredo[32]	Yes	PD	No	Yes	Yes	Yes	5
2009	Scire[33]	Yes	GS + PD	No	Yes	No	Yes	5
2009	Backhaus[25]	Yes	GS + PD	Yes	Yes	No	Yes	3
2009	Dougados[34]	Yes	GS + PD	Yes	Yes	No	Yes	3
2010	Balsa[35]	No	GS + PD	No	Yes	No	Yes	1

* Ultrasound definition of synovitis in gray-scale (i.e., synovial fluid, synovial hypertrophy, or both) and in power Doppler; ** Ultrasound component evaluated in scoring system; † Grading method used in the study; †† Global score calculated as sum of individual joint scores; GS: gray-scale, PD: power Doppler, US: ultrasound.

composite synovitis scoring system consisting of a combination of GS and PD. Quantification of activity also varied with all studies featuring semiquantitative scores (i.e., 0–3, 0–4, 0–5). An additional number of studies included binary or quantitative measures (e.g., thickness, resistive index, region of interest) as well. Several studies also included tenosynovitis and bursitis, but without clear definitions of lesion.

Table 3 shows the metric qualities (reliability, validity, discrimination, and feasibility) studied in the articles as part of the OMERACT filter. Regarding construct validity, correlation with clinical and laboratory findings varied according to the number and size of joints examined. Responsiveness was found to be variable according to the component tested (GS and PD) and the size of the joint. Two weeks seems to be the minimal time for visualizing minimal response (PD) and 24 weeks was found to be the best cutoff. With regard to feasibility, time of evaluation was variable (15–60 min) and increased with the number of joints involved in the examination.

The number of joints assessed by ultrasound varied between 5 and 60 joints among articles. Two joints, the second and third metacarpophalangeal (MCP) joints, were included in the synovitis scoring system of each article. Additionally, the second proximal interphalangeal (PIP) joint and the third and fourth MCP joints were assessed in 86% (12 of 14) of articles. Interphalangeal joints of the feet and the subtalar and midtarsal joints were the least commonly assessed joints, evaluated in only 14% (2 of 14) of articles. In previous studies, propositions on the number and composition of reduced joint count were

Table 3. Metric properties of the studies.

Year	Study	Validity				Responsiveness	Reliability	Feasibility
		Face	Content	Construct	Criterion			
2003	Szudlarek[9]	Yes	Yes	Yes	NA	NA	Yes	Yes
2004	Taylor[13]	Yes	Yes	Yes	NA	Yes	NA	Yes
2005	Naredo[27]	Yes	Yes	Yes	NA	NA	Yes	Bo
2005	Scheel[8]	Yes	Yes	Yes	NA	NA	NA	Yes
2005	Naredo[28]	Yes	Yes	Yes	NA	NA	NA	Yes
2005	Naredo[29]	Yes	Yes	Yes	NA	NA	Unclear	Yes
2008	Hameed[30]	Yes	Yes	Yes	NA	NA	NA	Yes
2008	Ozgocmen[31]	Yes	Yes	Yes	NA	NA	No	Yes
2008	Naredo[26]	Yes	Yes	Yes	NA	Yes	Yes	Yes/No*
2008	Naredo[32]	Yes	Yes	Yes	NA	Yes	Yes	Yes
2009	Scire[33]	Yes	Yes	Yes	NA	Yes	Yes	Yes
2009	Backhaus[25]	Yes	Yes	Yes	NA	Yes	Unclear	Yes
2009	Dougados[34]	Yes	Yes	Yes	NA	Yes	Yes	Yes/No*
2010	Balsa[35]	Yes	Yes	Yes	NA	Unclear	Yes	No

* Several scoring systems were evaluated in the study, feasibility varied with number of joints assessed. NA: not assessed in study.

based on suggested frequency of involvement[25], feasibility[8], or representative value of target joints[9], or were developed in a logistic regression model[26]. Two reduced joint counts seemed to present good validity issues: the 12-joint count proposed by Naredo, *et al*[26] and the 7-joint count by Backhaus, *et al*[25]. Joints selected in the proposed 12-joint count are wrist, MCP-2, MCP-3, knee, ankle, and elbow evaluated bilaterally[26]. Examining the other proposed joint counts, we found that some of them included a minimal number of 7 joints in their scoring systems and, in particular, joints featured in the 7-joint count[25] combination [i.e., wrist, MCP-2, MCP-3, PIP-2, PIP-3, metatarsophalangeal (MTP)-2, and MTP-5] were included in the global synovitis scores of 50% (7 of 14) of the articles. In order to evaluate the applicability of the 7-joint count in another dataset, we analyzed data from Naredo and colleagues by using the proposed selection of 7 joints from Backhaus[25]. Comparative results on responsiveness by using the 2 joint counts are presented in Table 4. The use of the 7-joint count in the new dataset also showed good responsiveness; however, the application of this joint count bilaterally (14 instead of 7 joints) was characterized by higher sensitivity to change, which was closer to that observed with the evaluation of all initially evaluated joints (i.e., 44 joints).

DISCUSSION

The prospect of developing a global ultrasound joint score is attractive, in that it might potentially be able to more objectively reflect the "real" level of synovitis, and hence disease activity of patients with RA, compared with conventional clinical measures, i.e., disease activity indices. In order to be accepted as an objective tool, ultrasound must demonstrate reliability and sensitivity to change, and the evaluation of several joints must also appear feasible. This review has demonstrated that ultrasound is a worthwhile tool for assessing global joint inflammation in RA.

This review has highlighted that discrepancies were present among studies, relating to the definition and detection of synovitis, and in the composition of joints included in the global evaluation of disease activity. The variability of definition and detection of synovitis at joint level within the studies

Table 4. Evaluation of responsiveness of the 7-joint score developed by Backhaus, *et al*[25] in an additional dataset obtained from Naredo, *et al*[26]. Data for this analysis is courtesy of Marina Backhaus and Esperanza Naredo.

Joint Count	Gray-scale Synovitis		Power Doppler Activity	
	Mean decrease‡ (95% CI)	SRM	Mean decrease‡ (95% CI)	SRM
Wrist, MCP2, MCP3, knee ankle, elbow (bilateral)*	2.5 (2.0–2.9)	0.925	2.0 (1.6–2.4)	0.797
Simplified 12-joint PDUS model (12 joints 24 recesses)#	2.3 (1.9–2.8)	0.881	1.8 (1.3–2.2)	0.717
Wrist, MCP2, MCP3, PIP2, PIP3, MTP2, MTP5 (unilateral right)**	1.5 (1.2–1.7)	0.842	1.0 (0.8–1.2)	0.678
Wrist, MCP2, MCP3, PIP2, PIP3, MTP2, MTP5 (unilateral left)**	1.3 (1.0–1.6)	0.732	0.7 (0.5–1.0)	0.581
Wrist, MCP2, MCP3, PIP2, PIP3, MTP2, MTP5 (bilateral)†	2.8 (2.3–3.4)	0.890	1.8 (1.3–2.2)	0.721
44 joints††	7.4 (6.1–8.8)	0.934	4.6 (3.5–5.6)	0.732

* Joints selected in the 12-count from Naredo, et al[26]; # for detailed list of recesses see Naredo, et al [26], ** 7-joint count 25, see Backhaus, et al for details on recesses and scoring modalities; † 7-joint count evaluated bilaterally; †† 44 joints include bilateral shoulder, elbow, wrist, MCP1-5, PIP-5, hip, knee, ankle, tarsal, MTP1-5; ‡ sensitivity to change; SRM: standardized response mean; MCP: metacarpophalangeal joint; PDUS: power Doppler ultrasound.

was the most important weakness raised by this review. This was most apparent in articles published before 2005, where GS definitions of synovitis, including its elementary components (SH, SF), were found to be lacking. After 2005 and the publication of the preliminary OMERACT definitions[19], synovitis was defined in all articles. Moreover, the OMERACT ultrasound definitions for synovitis and elementary components were used in most articles. Regarding the evaluation of synovial vascularization, less variability was found. All articles evaluated PD, rather than CD activity. The definition of PD activity proposed by Szkudlarek, *et al* was adopted by almost all articles[9]. The quantification of synovial activity at the single-joint level was also found to be variable. Some authors focused on the quantification of GS only, whereas others quantified PD only. Generally, authors evaluated both components separately. The semiquantitative method was the most frequently used method of quantification for both GS and PD, although the scales varied. For PD, the semiquantitative scale most commonly used was that proposed by Szkudlarek, *et al*[9] (i.e., grade 0: no flow in the synovium; grade 1: single vessel signals; grade 2: confluent vessel signals in less than half the area of the synovium; grade 3: vessel signals in more than half the area of the synovium). This high variability in the evaluation of joint activity made the comparison of studies, as well as the correct evaluation of validity, difficult.

The number of joints evaluated for creating a measure of global activity of RA and the explanation for the inclusion of joints was also found to be highly variable. In addition to providing a valid and reliable measurement, feasibility is of paramount importance with respect to ultrasound-based indices, as the examination of a large number of joints takes a considerable amount of time. Therefore, the number of joints that need to be assessed, and thus incorporated into any global ultrasound scoring system, is an important issue. Propositions on the number and composition of reduced joint counts were based either on suggested frequency of involvement[14,25,30] or representative value of target joints[9,31,36]; or they were developed in a logistic regression model[26]. Validation of proposed joint scores was quite often omitted, and only 2 papers examined the metric properties of the proposed reduced joint score[25,26]. Independent of the metric properties of a proposed joint score, validation is still necessary, and the choice and number of joints included remains a crucial issue. Candidate target joints to be included in a global ultrasound score may also be derived from clinical disease activity indices (i.e., Disease Activity Score 28), or based on other predictive studies, for example MRI studies (i.e., wrist)[37] and prediction of structural damage, or clinical prediction of severity. Considering the composition of available reduced global ultrasound scores, we found that the second and third MCP joints and the wrist were always included, regardless of how the joint score was developed. Almost all analyzed papers included the evaluation of at least 7 joints, similarly to the German ultrasound 7 score[25]; in addition the joint scores used in 50% of the articles included the 7 joints present in the German ultrasound 7 score. In order to test the external validity of this choice, in particular the number and the type of joints, data from the Naredo study[26] were reanalyzed using the German ultrasound 7 score (Table 4). This analysis revealed that the 7 joints included were good candidates for evaluating disease activity and responsiveness, even if sensitivity to change was inferior to the 12-joint score used in that database. Responsiveness is increased if the evaluation of joints is done bilaterally. As these are the joints most frequently involved in RA, it is probable that these joints would be included in the GLOSS as well; however, in order to guarantee C-reactive protein or erythrocyte sedimentation rate responsiveness, a certain number of large joints likely need to be included, as well.

Regarding the quantification of global synovitis, a number of important questions remain unanswered. Minimal activity at the single-joint level still needs to be determined. There is currently no agreement on what constitutes a "normal" level of GS and PD findings. What appears clear is that only joint activity (i.e., inflammation of the synovial membrane by GS and PD) should be included in a future score. Indeed, the inclusion of structural damage (i.e., erosions) did not demonstrate responsiveness, as it appears in the analysis of the article by Backhaus and colleagues[25]. This is probably due to the duration of followup (i.e., 6 months), which is probably too short for evaluating the effect on damage. A longer followup would probably have shown better sensitivity to change, as was demonstrated by Loeuille and colleagues[38]. Another factor could be the choice and number of joints evaluated for erosions (MCP 2, 3, PIP 2, 3, and MTP 2, 5 unilaterally), or the disease duration of the patients (mean 8.3 yrs). Clearly, additional studies on the responsiveness of erosions should be performed for demonstrating effect on damage, and therefore sensitivity to change of erosions detected by ultrasound. The use of GS evaluation alone would probably carry the same lack of sensitivity. In fact, it is sometimes difficult to differentiate between active synovitis (i.e., hypoechoic SH) and inactive or fibrous synovial thickening (i.e., echoic, hyperechoic SH), based only on the evaluation of echogenicity in GS, as such assessment is subjective and extremely dependent on the experience of the operator. This can also explain the greater sensitivity to change of PD signal (easier to detect), even if it is dependent on the quality of equipment. Based on this systematic review of the literature, it is difficult to suggest a minimal number of joints to score and which scoring system to use at joint level. The mathematical formulation (e.g., add all semiquantitative or quantitative scores up to produce a cumulative score) of the scoring system must also be determined. The validity of this simplified assessment, and that of others, remains to be tested and confirmed.

The OMERACT Ultrasound group is currently using the developed synovitis scoring system at joint level in an on-

going multicenter European study, in order to propose a standardized and reliable ultrasound synovitis GLOSS (Global OMERACT Scoring System). At the same time, data from this ongoing study will also be tested for responsiveness by using a different number of joints, including the 7- and 12-joint counts. This will probably permit the validation of the proposed joint count. Ultimately, the overall "usefulness and truthfulness" of GLOSS will be determined by its composition. We might well, however, need different indices for diagnosis and therapeutic monitoring.

APPENDIX

List of study collaborators: OMERACT Ultrasound Task Force: Philippe Aegerter, Sibel Aydin, Marina Backhaus, Peter V. Balint, David Bong, George A.W. Bruyn, Isabelle Chary-Valckenaere, Paz Collado, Eugenio De Miguel, Emilio Filippucci, Jane E. Freeston, Frederique Gandjbakhch, Walter Grassi, Marwin Gutierrez, Annamaria Iagnocco, Frederick Joshua, Sandrine Jousse-Joulin, David Kane, Helen I. Keen, Damien Loeuille, Ingrid Moller, Peter Mandl, Carlos Pineda, Lene Terslev, Wolfgang A. Schmidt, Marcin Szkudlarek, and Hans-Rudolf Ziswiler.

REFERENCES

1. Grassi W, Salaffi F, Filippucci E. Ultrasound in rheumatology. Best Pract Res Clin Rheumatol 2005;19:467-85.
2. Kane D, Grassi W, Sturrock R, Balint PV. Musculoskeletal ultrasound — a state of the art review in rheumatology. Part 2: Clinical indications for musculoskeletal ultrasound in rheumatology. Rheumatology 2004;43:829-38.
3. Szkudlarek M, Narvestad E, Klarlund M, Court-Payen M, Thomen HS, Østergaard M. Ultrasonography of the metatarsophalangeal joints in rheumatoid arthritis: comparison with magnetic resonance imaging, conventional radiography, and clinical examination. Arthritis Rheum 2004;50:2103-12.
4. Backhaus M, Burmester GR, Sandrock D, Loreck D, Hess D, Scholz A, et al. Prospective two year follow up study comparing novel and conventional imaging procedures in patients with arthritic finger joints. Ann Rheum Dis 2002;61:895-904.
5. Wakefield RJ, Brown AK, O'Connor PJ, Emery P. Power Doppler sonography: improving disease activity assessment in inflammatory musculoskeletal disease. Arthritis Rheum 2003;48:285-8.
6. D'Agostino MA, Said-Nahal R, Hacquard-Bouder C, Brasseur JL, Dougados M, Breban M. Assessment of peripheral enthesitis in the spondylarthropathies by ultrasonography combined with power Doppler: a cross-sectional study. Arthritis Rheum 2003;48:523-33.
7. Backhaus M, Kamradt T, Sandrock D, Loreck D, Fritz J, Wolf KJ, et al. Arthritis of the finger joints: a comprehensive approach comparing conventional radiography, scintigraphy, ultrasound, and contrast-enhanced magnetic resonance imaging. Arthritis Rheum 1999;42:1232-45.
8. Scheel AK, Hermann KG, Kahler E, Passewaldt D, Fritz J, Hamm B, et al. A novel ultrasonographic synovitis scoring system suitable for analyzing finger joint inflammation in rheumatoid arthritis. Arthritis Rheum 2005;52:733-43.
9. Szkudlarek M, Court-Payen M, Jacobsen S, Klarlund M, Thomen HS, Østergaard M. Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis. Arthritis Rheum 2003;48:955-62.
10. Ribbens C, André B, Marcelis S, Kaye O, Mathy L, Bonnet V, et al. Rheumatoid hand joint synovitis: gray-scale and power Doppler US quantifications following anti-tumor necrosis factor-alpha treatment: pilot study. Radiology 2003;229:562-9.
11. Szkudlarek M, Court-Payen M, Strandberg C, Klarlund M, Klausen T, Østergaard M. Power Doppler ultrasonography for assessment of synovitis in the metacarpophalangel joints with dynamic magnetic resonance imaging. Arthritis Rheum 2001;44:2018-23.
12. Terslev L, Torp-Pedersen S, Savnik A, von der Recke P, Qvistgaard E, Danneskiold-Samsøe B, et al. Doppler ultrasound and magnetic resonance imaging of synovial inflammation of the hand in rheumatoid arthritis: a comparative study. Arthritis Rheum 2003;48:2434-41.
13. Taylor PC, Steuer A, Gruber J, Cosgrove DO, Blomley MJ, Marsters PA, et al. Comparison of ultrasonographic assessment of synovitis and joint vascularity with radiographic evaluation in a randomized, placebo-controlled study of infliximab therapy in early rheumatoid arthritis. Arthritis Rheum 2004;50:1107-16.
14. Qvistgaard E, Røgind H, Torp-Pedersen S, Terslev L, Danneskiold-Samsøe B, Bliddal H. Quantitative ultrasonography in rheumatoid arthritis: evaluation of inflammation by Doppler technique. Ann Rheum Dis 2001;60:690-3.
15. Teh J, Stevens K, Williamson L, Leung J, McNally EG. Power Doppler ultrasound of rheumatoid synovitis: quantification of therapeutic response. Br J Radiol 2003;76:875-9.
16. Lee V, Zayat A, Wakefield RJ. The effect of joint position on Doppler flow in finger synovitis. Ann Rheum Dis 2009;68:603-4.
17. Schmidt WA, Schmidt H, Schicke B, Gromnica-Ihle E. Standard reference values for musculoskeletal ultrasonography. Ann Rheum Dis 2004;63:988-94.
18. Koski JM, Saarakkala S, Helle M, Hakulinen U, Heikkinen JO, Hermunen H. Power Doppler ultrasonography and synovitis: correlating ultrasound imaging with histopathological findings and evaluating the performance of ultrasound equipments. Ann Rheum Dis 2006;65:1590-5.
19. Backhaus M, Burmester GR, Gerber T, Grassi W, Machold KP, Swen WA, et al. Guidelines for musculoskeletal ultrasound in rheumatology. Ann Rheum Dis 2001;60:641-9.
20. 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.
21. 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.
22. 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.
23. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. J Rheumatol 1998;25:198-9.
24. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomised studies for health care interventions. J Epidemiol Community Health 1998;52:377-84.
25. 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.
26. Naredo E, Rodríguez M, Campos C, Rodríguez-Heredia 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.
27. Naredo E, Bonilla G, Gamero F, Uson J, Carmona L, Laffon A. Assessment of inflammatory activity in rheumatoid arthritis: a comparative study of clinical evaluation with grey scale and power Doppler ultrasonography. Ann Rheum Dis 2005;64:375-81.
28. Naredo E, Gamero F, Bonilla G, Uson J, Carmona L, Laffon A.

Ultrasonographic assessment of inflammatory activity in rheumatoid arthritis: comparison of extended versus reduced joint evaluation. Clin Exp Rheumatol 2005;23:881-4.
29. Naredo E, Collado P, Cruz A, Palop MJ, Cabero F, Richi P, et al. Longitudinal power Doppler ultrasonographic assessment of joint inflammatory activity in early rheumatoid arthritis: predictive value in disease activity and radiologic progression. Arthritis Rheum 2007;57:116-24.
30. Hameed B, Pilcher J, Heron C, Kiely PD. The relation between composite ultrasound measures and the DAS28 score, its components and acute phase markers in adult RA. Rheumatology 2008;47:476-80.
31. Ozgocmen S, Ozdemir H, Kiris A, Bozgeyik Z, Ardicoglu O. Clinical evaluation and power Doppler sonography in rheumatoid arthritis: evidence for ongoing synovial inflammation in clinical remission. South Med J 2008;101:240-5.
32. Naredo E, Möller I, Cruz A, Carmona L, Garrido J. Power Doppler ultrasonographic monitoring of response to anti-tumor necrosis factor therapy in patients with rheumatoid arthritis. Arthritis Rheum 2008;58:2248-56.
33. Scirè CA, Montecucco C, Codullo V, Epis O, Todoerti M, Caporali R. Ultrasonographic evaluation of joint involvement in early rheumatoid arthritis in clinical remission: power Doppler signal predicts short-term relapse. Rheumatology 2009;48:1092-7.
34. Dougados M, Jousse-Joulin S, Mistretta F, d'Agostino MA, Backhaus M, Bentin J, et al. Evaluation of several ultrasonography scoring systems of synovitis and comparison to clinical examination: Results from a prospective multi-center study of rheumatoid arthritis. Ann Rheum Dis 2000;69:828-33.
35. Balsa A, de Miguel E, Castillo C, Peiteado D, Martín-Mola E. Superiority of SDAI over DAS-28 in assessment of remission in rheumatoid arthritis patients using power Doppler ultrasonography as a gold standard. Rheumatology 2010;49:683-90.
36. Sheane BJ, Beddy P, O'Connor M, Miller S, Cunnane G. Targeted ultrasound of the fifth metatarsophalangeal joint in an early inflammatory arthritis cohort. Arthritis Rheum 2009;61:1004-8.
37. McQueen FM, Benton N, Perry D, Crabbe J, Robinson E, Yeoman S, et al. Bone edema scored on magnetic resonance imaging scans of the dominant carpus at presentation predicts radiographic joint damage of the hands and feet six years later in patients with rheumatoid arthritis. Arthritis Rheum 2003;48:1814-27.
38. Gill G, Chary-Valckenaere I, Sommier JP, Rat AC, Blum A, Loeuille D. Is ultrasonography a useful, reproducible and relevant tool to assess erosion progression in rheumatoid arthritis? [abstract]. Arthritis Rheum 2009;60 Suppl:1455.