

Reproducibility of Joint Swelling Assessments in Long-lasting Rheumatoid Arthritis: Influence on Disease Activity Score-28 Values (SEA-Repro Study Part I)

THIERRY MARHADOUR, SANDRINE JOUSSE-JOULIN, GÉRARD CHALÈS, LAURENT GRANGE, CÉCILE HACQUARD, DAMIEN LOEUILLE, JÉRÉMIE SELLAM, JEAN-DAVID ALBERT, JACQUES BENTIN, ISABELLE CHARY VALCKENAERE, MARIA-ANTONIETTA d'AGOSTINO, FABIEN ETCHEPARE, PHILIPPE GAUDIN, CHRISTOPHE HUDRY, MAXIME DOUGADOS, and ALAIN SARAUX

ABSTRACT. Objectives. To evaluate the reproducibility of clinical synovitis assessments in rheumatoid arthritis and the effect of variability on the Disease Activity Score-28 (DAS28).

Methods. Seven healthcare professionals from different cities examined the same patients with active non-early rheumatoid arthritis (RA; duration > 4 yrs), for whom a treatment change was being considered. There was no training session and the examination was to be performed as quickly as possible. The healthcare professionals assessed the 28 joints of the DAS28 in 7 patients (196 joints), then reexamined the same 28 joints in 4 of these 7 patients (112 joints), who had been rendered unrecognizable. Then 7 sonographers examined each of the 7 patients twice, using B-mode and power Doppler ultrasound (PD). The reference standards were presence of synovitis according to at least 50% of clinical examiners and 50% of sonographers. Agreement was assessed by Cohen's kappa statistic.

Results. Intraobserver reliability ranged from 0.31 (least experienced research technician) to 0.77 (most experienced physician). Interobserver reliability ranged from 0.18 to 0.62. The largest difference between the lowest and the highest swollen joint counts in the same patient was 15, and the greatest variation in the DAS28 score was 0.92. Agreement between clinical and sonographic reference standards was 0.46, 0.37, and 0.36 for B-mode, PD, and both, respectively.

Conclusion. Clinical inter- and intraobserver reliability is highly dependent on the examiner. Consequences on the DAS28 score can be substantial. Agreement with sonography is poor when both B-mode and PD are used but seems better, although low, when B-mode is used alone. (First Release April 1 2010; J Rheumatol 2010;37:932-7; doi:10.3899/jrheum.090879)

Key Indexing Terms:

RHEUMATOID ARTHRITIS
JOINT SWELLING

REPRODUCIBILITY

SYNOVITIS
DISEASE ACTIVITY SCORE-28

Synovitis, or inflammatory hypertrophy of the synovial membrane, is a hallmark of rheumatoid arthritis (RA) that can cause joint swelling. The number of joints with synovi-

tis is related to disease activity. Studies have shown that counts on a limited number of joints are valid for assessing disease activity^{1,2}. In everyday practice, counts are usually

From the Rheumatology Unit, CHU, and University Hospital La Cavale Blanche, Brest; Rheumatology Unit, CHU, and University Hospital, Hôpital Sud, Rennes; Rheumatology Unit, GREPI TIMC, Imag UMR CNRS 5525, University Hospital, Hôpital Sud A. Michallon, Grenoble; Rheumatology Unit, CHU, and University Hospital Ambroise Paré, Boulogne Billancourt; Rheumatology Unit, CHU, and University Hospital, Nancy-Brabois, Vandoeuvre-Lès-Nancy; Rheumatology Unit, CHU, and University Hospital Bicêtre, Le Kremlin-Bicêtre, France; Rheumatology Unit, University Hospital, Brugmann, Brussels, Belgium; Rheumatology Unit, CHU, and Les Charmilles Hospital, Arpajon; and Rheumatology Unit, CHU and Cochin University Hospital, Paris, France. Supported by Abbott France, Paris, France.

T. Marhadour, MD; S. Jousse-Joulin, MD, Rheumatology Unit, CHU, and University Hospital La Cavale Blanche; G. Chalès, MD, Rheumatology Unit, CHU, and University Hospital, Hôpital Sud; L. Grange, MD, Rheumatology Unit, GREPI TIMC, Imag UMR CNRS 5525, University Hospital, Hôpital Sud A. Michallon; C. Hacquard, MD, Rheumatology Unit, CHU, and University Hospital Ambroise Paré; D. Loeuille, MD,

PhD, Rheumatology Unit, CHU, and University Hospital, Nancy-Brabois; J. Sellam, MD, Rheumatology Unit, CHU, and University Hospital Bicêtre; J.-D. Albert, MD, Rheumatology Unit, CHU, and University Hospital, Hôpital Sud; J. Bentin, MD, Rheumatology Unit, University Hospital, Brugmann, Brussels; I. Chary Valckenaere, MD, PhD, Rheumatology Unit, CHU, and University Hospital, Nancy-Brabois; M.-A. d'Agostino, MD, PhD, Rheumatology Unit, CHU, and University Hospital Ambroise Paré; F. Etchepare, MD, Rheumatology Unit, CHU, and Les Charmilles Hospital; P. Gaudin, MD, PhD, Rheumatology Unit, GREPI TIMC, Imag UMR CNRS 5525, University Hospital, Hôpital Sud A. Michallon; C. Hudry, MD; M. Dougados, MD, Rheumatology Unit, CHU and Cochin University Hospital; A. Saraux, MD, PhD, Rheumatology Unit, CHU, and University Hospital La Cavale Blanche.

Address correspondence to Prof. A. Saraux, Rheumatology Unit, La Cavale Blanche Hospital, Brest Teaching Hospital, BP 814, F29609 Brest Cedex, France. E-mail: Alain.Saraux@univ-brest.fr

Accepted for publication December 29, 2009.

done on the following 28 joints: metacarpophalangeal joints, proximal interphalangeal joints, wrists, elbows, shoulders, and knees.

The Disease Activity Score-28 (DAS28)³ is a composite index of RA activity that is useful for making treatment decisions. It is calculated from 4 variables: the tender joint count (TJC) and swollen joint count (SJC) on the 28 joints above, the erythrocyte sedimentation rate (ESR; mm/h), and general health (GH) assessed by the patient on a 100-mm visual analog scale (VAS). The formula is as follows: $DAS28 = 0.56 \times \sqrt{TJC} + 0.28 \times \sqrt{SJC} + 0.70 \times ESR + 0.014 \times GH$. Scores < 2.6 indicate complete remission, scores 2.6–3.2 low disease activity, scores 3.2–5.1 moderate disease activity, and scores > 5.1 high disease activity.

Joint swelling is the only presumably objective clinical variable in the DAS28 index. However, the difference between a normal joint and a swollen joint is not clearly defined, thus SJC may be affected by intra- and interobserver variability. These variations in the SJC may influence the DAS28 value and therefore the assessment of disease activity.

Sonography is rapidly becoming a major method for assessment of inflammatory joint disease. Joint swelling can be detected by sonography. Sonography is a widely available and inexpensive investigation that is increasingly used as a complement to the clinical examination^{4,5}. However, poor agreement has been reported between clinical and sonographic SJC⁶.

We have conducted the SEA-Repro (Sonographic Evaluation of Arthritis) study, of which the objectives were to evaluate the intra- and interobserver reproducibility of clinically determined SJC in 7 patients with RA, to evaluate the influence of intra- and interobserver variability on the DAS28 score, and to evaluate agreement between clinical and sonographic SJC.

MATERIALS AND METHODS

Patients. Seven patients with active RA recruited at the Rheumatology Department of Brest Teaching Hospital were included (Table 1). They met the 1987 revised American Rheumatism Association (American College of

Rheumatology)⁷ criteria for RA. There were 5 women and 2 men, with a mean age of 57.1 years (SD 6.8) and a mean disease duration of 22.1 years (SD 13.6). Three patients had rheumatoid nodules. All patients were receiving corticosteroids (mean dosage 8.1 mg/day, SD 3.6) and disease-modifying antirheumatic drugs (methotrexate in 6 patients). Two patients were receiving tumor necrosis factor (TNF) antagonist therapy. Because of high disease activity, a change in treatment (introduction of a TNF antagonist or switch to another TNF antagonist) was being considered for all 7 patients. Mean TJC was 9 (SD 6.9), mean ESR was 23.8 mm (SD 11.2), and mean pain intensity on 100-mm VAS was 56.8 (SD 21.9).

Conduct of the study. The patients and examiners attended a meeting for the assessments. All patients were examined by 7 clinical healthcare professionals, and 4 of the 7 patients were examined twice. Then 7 experienced sonographers examined the 28 joints in each of the 7 patients on 2 separate occasions. The assessment technique was standardized during a consensus meeting held just before the assessment session. During this consensus meeting, sonographers received training on the joint assessment and information on the study methodology.

The 7 clinical examiners were recruited from different cities in France and Belgium. There were 5 physicians, 1 clinical research technician, and 1 occupational therapist. They were defined as senior if they had at least 5 years of experience and otherwise as junior. Only 5 minutes were allowed for the joint examination in each patient. No instructions were given before the examinations, and there was no training session⁸. The 28 joints used for the DAS28 were assessed for swelling (total of 196 joints), and the findings were scored using a semiquantitative scale (0, no synovitis; 1, synovitis unlikely; 2, synovitis probable; and 3, synovitis present). Joints with scores of 2 or 3 were counted as swollen. Four patients then donned masks and gowns to make them unrecognizable and were reexamined by the 7 clinical healthcare professionals (112 joints).

Then, 7 sonographers from different cities examined the 28 joints in each of the 7 patients and repeated the investigation in the patients wearing masks and gowns. They used an Esaote Technos MPX apparatus with a 12.5-MHz transducer (Esaote Biomedica, Genoa, Italy). The joints were assessed using the OMERACT preliminary definition of synovitis and a grading system based on Szkudlarek's semiquantitative method⁹⁻¹¹ for both B-mode and power Doppler ultrasound (Table 2). As reported¹², synovitis was defined as a grade ≥ 1 by both B-mode ultrasound (at least synovial thickening bulging over the line linking the tops of the periarticular bones but without extension along the bone diaphysis) and power Doppler (up to 3 discrete spots or 1 confluent spot plus up to 2 discrete spots).

Thus, the clinical examiners performed their assessments in conditions that replicated routine practice (no training session, no instructions, and only 5 minutes for the joint examination of each patient). Sonography, in contrast, was performed under optimal conditions (training session and instructions) to evaluate the concordance between clinical examination and sonography for detecting clinically relevant synovitis.

Table 1. Patient characteristics.

| Patient | Age, yrs | Disease Duration, yrs | VAS | Swollen Joint Count* | Tender Joint Count | ESR | DMARD | Biologics |
|---------|----------|-----------------------|-----|----------------------|--------------------|-----|----------|-----------|
| 1 | 55 | 29 | 52 | 9 | 2 | 15 | MTX | TNF |
| 2 | 55 | 4 | 72 | 10 | 7 | 19 | SLZ, MTX | 0 |
| 3 | 62 | 20 | 83 | 11 | 13 | 19 | MTX | 0 |
| 4 | 67 | 37 | 24 | 6 | 5 | 46 | LFN, HCQ | 0 |
| 5 | 57 | 5 | 65 | 12 | 22 | 14 | MTX | TNF |
| 6 | 45 | 23 | 32 | 8 | 4 | 22 | MTX | 0 |
| 7 | 59 | 37 | 70 | 16 | 11 | 22 | MTX, HCQ | 0 |

* SJC according to at least 50% (4/7) of the clinical healthcare professionals. VAS: 100-mm visual analog scale; ESR: erythrocyte sedimentation rate; DMARD: disease-modifying antirheumatic drug; MTX: methotrexate; TNF: tumor necrosis factor; SLZ: sulfasalazine; LFN: leflunomide; HCQ: hydroxychloroquine.

Table 2. Sonographic evaluation: scanning technique and scoring of synovitis (synovitis was defined as at least grade 1 by both B-mode and power Doppler).

| Scanning Technique | |
|--------------------|--|
| Glenohumeral joint | Presence or absence of fluid and/or hypoechoic synovial hypertrophy in the bicipital groove, transducer transversal and longitudinal to the anterior part of the humerus, shoulder in neutral position. Axillary recess, longitudinal scan |
| Elbow | Transverse anterior scan and posterior scan with elbow flexed at 90°: presence or absence of fluid and/or hypoechoic synovial hypertrophy from the humeral capitellum or the coronoid fossa to the joint capsule |
| Wrist | Longitudinal and transversal, from the dorsal aspect with the joint in neutral position, probe located on the 3rd finger: presence or absence of fluid and/or hypoechoic synovial hypertrophy from the bones to the joint capsule |
| 5 MCP and 5 PIP | Longitudinal scanning of the dorsal MCP and PIP joints in extension: presence or absence of fluid and/or hypoechoic synovial hypertrophy from the articular bony margin to the joint capsule |
| Knee | Longitudinal and transversal, from the suprapatellar recess, in a supine position, with the joint in 30° of flexion: presence or absence of fluid and/or hypoechoic synovial hypertrophy in the anteroposterior diameter of the suprapatellar recess |
| Scoring System | |
| B-mode | Grade 0 or normal: Normal joint (no synovial hypertrophy, no joint effusion) |
| | Grade 1 or minimal: Minimal synovitis (minimal synovial hypertrophy, with or without minimal joint effusion) |
| | Grade 2 or moderate: Moderate synovitis (moderate synovial hypertrophy, with or without minimal or moderate joint effusion) |
| | Grade 3 or severe: Severe synovitis (severe synovial hypertrophy, with or without severe joint effusion) |
| Power Doppler | Grade 0: no vessel in the synovium |
| | Grade 1: up to 3 discrete spots or 1 confluent spot plus up to 2 discrete spots |
| | Grade 2: vessel signals in less than 50% of the surface area of the synovium |
| | Grade 3: vessel signals in more than 50% of the surface area of the synovium |

MCP: metacarpophalangeal; PIP: proximal interphalangeal.

Reference standards. We used the clinical reference standard, namely, synovitis found by at least 50% (4/7) of the clinical healthcare professionals. The sonographic reference standard was synovitis found by at least 50% (4/7) of the sonographers.

Statistical analysis. Statistical analysis was performed using SPSS 15.0 for Windows (SPSS Inc., Chicago, IL, USA). Reproducibility was assessed based on Cohen's kappa coefficient, as follows: excellent, 0.80; good, 0.60–0.79; fair, 0.40–0.59; and poor, < 0.40). Intraobserver reproducibility was assessed by comparing the results of the first and second examinations of the same 4 patients. Interobserver reproducibility was assessed by comparing the results of the first examination to the 2 clinical reference standards.

To evaluate the effects of intra- and interobserver variability on the DAS28 values, we computed the DAS28 in each patient using the lowest and highest SJC for that patient.

Finally, to evaluate agreement between the clinical and sonographic examinations, we compared the joints that had synovitis according to at least 50% of the clinical healthcare professionals to those that had synovitis according to at least 50% of the sonographers.

RESULTS

Intraobserver reproducibility. Cohen's kappa values indicated that intraobserver reproducibility ranged from poor to good (0.31 to 0.77; Table 3). Reproducibility was best for

the rheumatologist with the most experience and worst for the occupational therapist. None of the examiners had kappa values in the excellent range. Values indicated fair reproducibility for 3 of the 7 examiners.

For the sonographers, intraobserver reproducibility was between 0.37 and 0.75 in B-mode ultrasound and between 0.25 and 0.77 in power Doppler mode.

Interobserver reproducibility. Considerable interobserver variability was found (Table 3). With the 50% reference standard, kappa values ranged across examiners from 0.40 to 0.62.

For the sonographers, interobserver reproducibility was between 0.43 and 0.63 in B-mode and between 0.27 and 0.56 in power Doppler mode.

Agreement between clinical and sonographic SJC. Agreement was poor overall (Table 4). The kappa coefficient was 0.36 for the comparison of the 50% sonographic reference standard (B-mode and power Doppler) to the 50% clinical reference standard. However, agreement was better when the 50% sonographic reference standard was com-

Table 3. Intraobserver and interobserver reproducibility of the experience of the clinical examiners.

| Observer | Experience > 5 Years | Intraobserver Reproducibility (κ) | Interobserver Reproducibility (κ) with 50% Reference Standard |
|----------|----------------------|--|--|
| 1 | Yes | 0.77 | 0.52 |
| 2 | No | 0.31 | 0.45 |
| 3 | Yes | 0.65 | 0.62 |
| 4 | No | 0.53 | 0.40 |
| 5 | Yes | 0.42 | 0.55 |
| 6 | Yes | 0.38 | 0.60 |
| 7 | No | 0.48 | 0.61 |

Table 4. Interobserver reproducibility between the 50% clinical reference standard and the 50% sonography reference standard [B-mode: at least minimal synovial thickening; and power Doppler (PD) mode: grade 1].

| Clinical Synovitis | Sonographic Synovitis | | | | | |
|--------------------|-----------------------|-----|---------|-----|---------------|-----|
| | B-mode | | PD Mode | | B-mode and PD | |
| | No | Yes | No | Yes | No | Yes |
| No | 75 | 49 | 116 | 7 | 116 | 7 |
| Yes | 6 | 66 | 44 | 28 | 45 | 27 |
| Kappa | 0.46 | | 0.37 | | 0.36 | |

pared to the 50% clinical reference standard using B-mode only (kappa = 0.46). Agreement was better at the interphalangeal and metacarpophalangeal joints than at the wrist, elbow, and knee (data not shown; as described¹²).

Agreement was poor between the 50% sonographic reference standard and the 50% clinical reference standard when power Doppler was used alone (kappa = 0.37): the proportion of patients who were grade 1 or above by both B-mode and power Doppler was similar to the proportion who were grade 1 or above by power Doppler only.

Effect on DAS28 value. Using B-mode plus power Doppler sonography (with either grade 1 or grade 2), all patients had a lower 28-joint count by sonography than by clinical examination. Using B-mode only and grade 1, all patients had higher 28-joint counts by sonography than by clinical examination. Using B-mode only and grade 2, the differences with the clinical examination were smaller, although they remained substantial. The mean difference between the highest and lowest SJC in a given patient was 12, and the maximum difference was 15 (2 vs 17; Table 5). The mean DAS28 variation due to the SJC differences was 0.59, and the maximum variation was 0.92 (Table 6).

DISCUSSION

A measurement tool must be valid, reproducible, and sensitive to change¹³. The SJC is considered a valid indicator of

Table 5. Swollen joint counts (SJC) according to clinicians and sonographers.

| Patient | Clinicians* SJC (range) | Sonographers** | | | |
|---------|----------------------------|------------------|--------|------------------|--------|
| | | At Least Grade 1 | | At Least Grade 2 | |
| | | B | B + PD | B | B + PD |
| 1 | 9 (4–15) | 18 | 7 | 9 | 6 |
| 2 | 10 (7–17) | 12 | 1 | 9 | 1 |
| 3 | 11 (7–17) | 18 | 8 | 11 | 9 |
| 4 | 6 (3–15) | 17 | 1 | 6 | 1 |
| 5 | 12 (11–22) | 15 | 4 | 11 | 3 |
| 6 | 8 (2–17) | 13 | 2 | 6 | 1 |
| 7 | 16 (9–24) | 22 | 10 | 14 | 7 |

* According to at least 50% of the clinical healthcare professionals.

** According to at least 50% of the sonographers, for both B-mode and power Doppler. B: B-mode sonography; PD: power Doppler.

RA activity, based on a small number of published studies. The SJC has both criterion validity (e.g., correlation with bone erosions) and construct validity (e.g., correlation with acute-phase proteins)^{14–17}. The SJC is included in the core set of disease activity measures developed by the European League Against Rheumatism¹⁸.

Studies of the SJC have shown good or excellent intraobserver reproducibility^{19,20}, usually with lower interobserver reproducibility^{18,21}. However, we found that both intra- and interobserver reproducibility ranged from poor to good. This discrepancy may be ascribable to the large number of examiners in our study and to the considerable differences in their levels of experience and background training.

In France, in routine practice, rheumatologists evaluate the SJC. Nurses may determine the SJC with similar intraobserver reproducibility. Generally, for both routine practice and research studies, sonographers undergo specific training, whereas clinicians do not. In a preliminary evaluation of the duration of the clinical and sonographic examinations, we found that the clinicians never spent more than 5 minutes determining the SJC, whereas the sonographers sometimes wanted more than 15 minutes. For this study, we therefore chose to evaluate the reproducibility of SJC evaluations and we deliberately invited clinicians who came from various cities and who differed in their experience. Training the clinicians, or not training the sonographers, might have modified our results. Thus, our study did not compare clinical and sonographic examination in routine practice. Instead, we compared reliability across clinicians in routine practice, reliability across sonographers after standardization of the scanning technique and synovitis grading system, and reliability between clinicians in routine practice and sonographers after standardization of the sonographic synovitis evaluation. We did not evaluate routine practice of sonographers, because no clear ultrasound definition of synovitis has been published (for example, in routine practice, some sonographers may define synovitis as B-mode grade 2 without power Doppler abnormalities, whereas others may use different definitions).

Another limitation of the study may be the small number of patients. However, studying a larger number would be challenging, as each patient underwent 28 examinations on the same day (14 sonograms and 14 clinical examinations) and each clinician and sonographer performed 14 evaluations.

We did not assess sensitivity to change, because the study was conducted on a single day. SJC varied widely across clinical examiners in the same patients. One possible explanation is that the examiners may have used various examination methods, as there was no training session before the study. Further, our results emphasize the effects of experience and skill on the assessment of synovitis. The occupational therapist was more used to examining the hands than the other joints. The rheumatologist who had the most years of experience was also best at grading doubtful synovitis.

Table 6. DAS28 values using the maximum and minimum clinician swollen joint counts (SJC) obtained by the 7 clinicians.

| Patient | DAS28 with Maximum SJC | DAS28 with Minimum SJC | Difference | DAS28 Using SJC of $\geq 50\%$ of the Clinicians | DAS28 Using SJC of $\geq 50\%$ of the Sonographers using Grade 1 | |
|---------|------------------------|------------------------|------------|--|--|--------|
| | | | | | B-Mode | B + PD |
| 1 | 4.5 | 3.98 | 0.52 | 4.26 | 4.26 | 4.10 |
| 2 | 5.71 | 5.29 | 0.42 | 5.44 | 5.39 | 4.83 |
| 3 | 6.4 | 5.98 | 0.42 | 6.17 | 6.17 | 6.03 |
| 4 | 5.35 | 4.75 | 0.60 | 4.95 | 4.95 | 4.55 |
| 5 | 6.70 | 5.78 | 0.92 | 6.35 | 6.31 | 5.87 |
| 6 | 4.89 | 4.13 | 0.76 | 4.52 | 4.42 | 4.01 |
| 7 | 6.37 | 5.84 | 0.53 | 6.12 | 6.05 | 5.74 |

B: B-mode sonography; PD: power Doppler.

The SJC variations led to variations in DAS28 score values. The greatest difference was 0.92 points. These DAS28 variations may affect treatment decisions. Nevertheless, the SJC contributes only half as much to the DAS28 value compared to the TJC. As a result, the DAS28 differences across examiners would have affected the treatment decision (to start a TNF antagonist or switch TNF antagonists) in a single patient.

Most studies comparing clinical and sonographic joint assessments have shown fair agreement²²⁻²⁵, with greater sensitivity of sonography for detecting synovitis²⁶. Our study confirms these results. Nevertheless, agreement was good when the SJC determined by a good clinical examiner was compared to the SJC determined by a good sonographer using B-mode only. B-mode imaging assesses only the degree of synovial membrane hypertrophy and the presence of joint effusions; whereas power Doppler investigates the blood supply, which reflects the degree of inflammation. Clinical examination of the joints assesses synovial membrane hypertrophy and joint effusion, similar to B-mode sonography.

We recently evaluated the clinimetric properties of various sonography scoring systems, and found that sonography was at least as good as the clinical scores (intraobserver reliability range 0.61–0.97 vs 0.53–0.82; construct validity range 0.76–0.89 vs 0.76–0.88; correlation with C-reactive protein range 0.28–0.34 vs 0.28–0.35; and sensitivity to change range 0.60–1.21 vs 0.96–1.36 for sonography vs clinical scoring systems, respectively)²⁷. These results suggest that sonographic evaluation of synovitis may be at least as relevant as the clinical examination. Further studies are now required to develop optimal scoring systems for monitoring RA patients based on either clinical or sonographic evaluation of the SJC.

Our results were obtained in patients with long-lasting RA, as we excluded patients with early RA. Our feeling was that, in early RA, power Doppler sonography probably detects more joints with synovitis than do clinicians. Similar studies in early RA are needed.

We did not evaluate the clinimetric properties of the VAS score, TJC, or ESR. The reproducibility of these 3 variables may also modify the DAS28 score.

Although the SJC seems less subjective than the TJC or VAS score, it varies across examiners. Our results indicated that intraobserver reproducibility was variable, but can be good if the clinical examiner is highly skilled. Interobserver reproducibility was fair. Intra- and interobserver variability may influence DAS28 score values and therefore affect treatment decisions. Finally, although agreement between clinical examination and sonography was poor, close agreement can occur between a good clinical examiner and a good sonographer using B-mode ultrasound only, without power Doppler. Thus power Doppler sonography clearly gives additional information for the clinician and its significance for the evaluation of disease activity and treatment decisions requires further evaluation.

REFERENCES

1. Prevoo MLL, van Riel PLCM, van 't Hof MA, van Rijswijk MH, van Leeuwen MA, Kuper HH, et al. Validity and reliability of joint indices: a longitudinal study in patients with recent onset rheumatoid arthritis. *Br J Rheumatol* 1993;32:589-94.
2. Fuchs HA, Pincus T. Reduced joint counts in controlled clinical trials in rheumatoid arthritis. *Arthritis Rheum* 1994;37:470-5.
3. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;38:44-8.
4. Balint P, Sturrock RD. Musculoskeletal ultrasound imaging: a new diagnostic tool for the rheumatologist? *Br J Rheumatol* 1997;36:1141-2.
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. 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.
7. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315-24.

8. D'Agostino MA, Maillefert JF, Said-Nahal R, Breban M, Ravaud P, Dougados M. Detection of small joint synovitis by sonography: the learning curve of rheumatologists. *Ann Rheum Dis* 2004;63:1284-7.
9. Wakefield RJ, D'Agostino MA, Iagnocco A, Filippucci E, Backhaus M, Scheel AK, et al, OMERACT Ultrasound Group. The OMERACT ultrasound group: status of current activities and research directions. *J Rheumatol* 2007;34:848-51.
10. Wakefield RJ, Balint PV, Szkudlarek M, Filippucci E, Backhaus M, D'Agostino MA, et al, OMERACT 7 Special Interest Group. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. *J Rheumatol* 2005;32:2485-7.
11. Szkudlarek M, Court-Payen M, Jacobsen S, Klarlund M, Thomsen HS, Østergaard M. Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis. *Arthritis Rheum* 2003;48:955-62.
12. Jousse-Joulin S, D'Agostino MA, Marhadour T, Albert JD, Bentin J, Chary Valckenaere I, et al. Reproducibility of joint swelling assessment by sonography in patients with rheumatoid arthritis (SEA-Repro Study Part II). *J Rheumatol* 2010;37: [In press].
13. Boers M, Van Riel P, Felson D, Tugwell P. Assessing the activity of rheumatoid arthritis. *Bailliere's Clin Rheumatol* 1995;9:305-17.
14. Thompson PW, Silman AJ, Kirwan JR, Currey HL. Articular indices of joint inflammation in rheumatoid arthritis. Correlation with the acute-phase response. *Arthritis Rheum* 1987;30:618-23.
15. van der Heijde DM, van 't Hof MA, van Riel PL, van Leeuwen MA, van Rijswijk MH, van de Putte LB. Validity of single variables and composite indices for measuring disease activity in rheumatoid arthritis. *Ann Rheum Dis* 1992;51:177-81.
16. Prevo ML, van Riel PL, van 't Hof MA, van Rijswijk MH, van Leeuwen MA, Kuper HH, et al. Validity and reliability of joint indices. A longitudinal study in patients with recent onset rheumatoid arthritis. *Br J Rheumatol* 1993;32:589-94.
17. Fuchs HA, Pincus T. Reduced joint counts in controlled clinical trials in rheumatoid arthritis. *Arthritis Rheum* 1994;37:470-5.
18. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum* 1993;36:729-40.
19. Marks JS, Palmer MK, Burke MJ, Smith P. Observer variation in examination of knee joints. *Ann Rheum Dis* 1978;37:376-7.
20. Goossens PH, Heemskerk B, Van Tongeren J, Zwinderman AH, Vliet Vlieland TPM, Huizinga WJ. Reliability and sensitivity to change of various measures of hand function in relation to treatment of synovitis of the metacarpophalangeal joint in rheumatoid arthritis. *Rheumatology* 2000;39:909-13.
21. Rees JD, Pilcher J, Heron C, Kiely PD. A comparison of clinical vs ultrasound determined synovitis in rheumatoid arthritis utilizing gray-scale, power Doppler and the intravenous microbubble contrast agent 'Sono-Vue'. *Rheumatology* 2007;46:454-9.
22. Szkudlarek M, Court-Payen M, Strandberg C, Klarlund M, Klausen T, Østergaard M. Power Doppler ultrasonography for assessment of synovitis in the metacarpophalangeal joints of patients with rheumatoid arthritis: a comparison with dynamic magnetic resonance imaging. *Arthritis Rheum* 2001;44:2018-23.
23. Szkudlarek M, Narvestad E, Klarlund M, Court-Payen M, Thomsen 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.
24. Luukkainen R, Sanila MT, Saltyshev M, Huhtala H, Koski JM. Relationship between clinically detected joint swelling and effusion diagnosed by ultrasonography in elbow joints in patients with rheumatoid arthritis. *Clin Rheumatol* 2005;24:228-31.
25. Luukkainen R, Sanila MT, Luukkainen P. Poor relationship between joint swelling detected on physical examination and effusion diagnosed by ultrasonography in glenohumeral joints in patients with rheumatoid arthritis. *Clin Rheumatol* 2007;26:865-7.
26. 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.
27. 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* 2009 Sep 9. [Epub ahead of print]