Comparison of discrimination and prognostic value of two US Doppler scoring systems in rheumatoid arthritis patients: a prospective cohort study

K. Ellegaard[1], L. Terslev[1,2], R. Christensen[1], M. Szkudlarek[3], W.A. Schmidt[4], P.S. Jensen[1], H. Bliddal[1], S. Torp-Pedersen[1]

[1]The Parker Institute, Department of Rheumatology, Copenhagen University Hospital, Frederiksberg, Copenhagen, Denmark; [2]Centre for Rheumatology and Spinal Diseases, Copenhagen University Hospital Glostrup, Copenhagen, Denmark; [3]Department of Rheumatology, Copenhagen University Hospital Koege, Koege, Denmark; [4]Immanuel Krankenhaus Berlin, Medical Centre for Rheumatology, Berlin-Buch, Berlin, Germany.

Abstract

Objective

The aim of this paper is to investigate sensitivity to change (SRM), predictive validity and discriminative ability of a quantitative (QS) and a semi-quantitative (SQS) Doppler ultrasound scoring systems in patients with rheumatoid arthritis (RA) treated with anti-TNF-α therapy.

Methods

RA patients with wrist joint affection treated with TNF-α inhibitor were followed for one year. The wrist was examined with Doppler before initiating therapy and after one year. DAS28 was determined at both visits. One person trained in the SQS system and one in the QS system evaluated the anonymised images. The SRM, predictive validity and discriminative ability for both systems were calculated using DAS28 as the measure of disease improvement.

Results

Fourty-six patients with RA (80% females) were included. The mean Doppler activity at baseline was QS:24.4% (SD=17.7%) and SQS:2.0 (SD=0.6). A decrease in Doppler activity was seen for both systems after anti-TNF-α therapy. Sensitivity to change was seen, SRM=-0.52 (95%CI; -0.83 to -0.21; QS) and -0.24 (-0.53 to -0.05; SQS). Predictive value was poor (QS r_s=-0.24; SQS r_s=-0.05). Construct validity was; QS: r_s=0.29, SQS: r_s=0.23.

Conclusion

Both systems were to some extent sensitive to change. Predictive validity and discriminate capacity of both systems showed only a weak association to DAS 28 in the study population. The QS was a little superior to the SQS.
The results do not necessarily reflect Doppler evaluation as being ineffective, but may be caused by DAS28 not being a perfect marker of inflammation.

Key words
US Doppler, scoring systems, predictive validity

Karen Ellegaard, PT, PhD
Lene Terslev, MD, PhD
Robin Christensen, MSc, PhD
Marcin Szkudlarek, MD, PhD
W.A. Schmidt, MD, Prof.
Peter S. Jensen, MD
Henning Bliddal, MD, Prof.
Søren Torp-Pedersen, MD

Please address correspondence to:
Karen Ellegaard, PhD,
The Parker Institute,
Department of Rheumatology,
Ndr. Fasanvej 57,
2000 Frederiksberg,
Copenhagen, Denmark.
E-mail: parker@frh.regionh.dk

Received on September 12, 2013; accepted in revised form on January 23, 2014.

Funding: this study was supported by the Oak Foundation.

Competing interests: none declared.

Introduction

Valid measurements for monitoring patients with rheumatoid arthritis (RA) are relevant, as it has been demonstrated that early sufficient treatment is crucial for the course of the disease (1). Both power and colour Doppler ultrasound (US) are used to monitor treatment strategy in patients with RA and has been shown to be able to detect treatment response or lack thereof (1-3); as a consequence it has been of interest in rheumatology practice and in research to score the changes over time using US Doppler measurements.

For the purpose of monitoring pathological changes over time, different types of US scoring systems have been proposed where those of interest are the semi-quantitative scoring(SQS) (4) and the quantitative scoring (QS) (5) systems. They all grade the amount of colour pixels in the region of interest (ROI) including estimates of the synovial hypertrophy. QS systems calculate the amount of colour pixels in the synovium (1, 5-6) whereas the SQS systems measure the amount of Doppler on an ordinal scale, mostly with four response categories from 0–3 (4, 7-8). Several different SQS systems have been presented, while none has been chosen by consensus (9).

According to the OMERACT filter, it is imperative to investigate the discriminative ability and/or truth to assess the prognostic value of measurements to verify the usefulness in both clinical trials and clinical practice (10). The discriminative ability refers to a measurement's reliability and its sensitivity to changes (to measure changes over time) and the truth according to the prognostic value is of importance for the patients, as it reflects a measurement's ability to forecast future events – such as, for instance, worsening or remission (11-12).

Though these US measures (SQS and QS) are used extensively, it is not known which of the two scoring systems is the best to (i) discriminate between effective and ineffective therapies, and (ii) prognostically predict a good clinical response to treatment with, for example, a biological drug.

The aim of this prospective cohort study, in patients with RA, was with a head-to-head comparison to investigate the discriminative ability, sensitivity to change and predictive validity of two different types of scoring systems in patients with RA treated with a TNF-alpha inhibitor over a period of 1 year.

Methods

Patients and clinical outcome measures

A cohort of patients fulfilling the ACR 1987 criteria for RA (13) was enrolled at baseline and re-evaluated for clinical response after 1 year on therapy with an anti-TNF drug (adalimumab, etanercept or infliximab). Included in this prospective study were only patients who remained on the same therapy and had disease activity in the wrist joint with presence of both synovial hypertrophy and colour Doppler activity at baseline (14).

For each patient one image, taken centrally in the most active wrist joint, whether dominant or not, was obtained at both visits and these scans were anonymised and distributed to the two investigators with extensive experience in scoring with either the quantitative or semi-quantitative system. The same wrist joint was scanned at both baseline and 1 year follow-up using Doppler US. Clinical information was obtained from the records with Disease Activity Score based on a 28 joint count (DAS-28) as determined for all patients using the C-reactive protein (CRP) at both baseline and 1-year follow-up.

Ultrasound assessment

The US evaluations were performed with a Siemens Acuson Sequoia™ (Mountainview, CA, USA) using a linear array transducer with 14 MHz centre frequency. Colour Doppler was used for evaluation of hyperemia, as it was more sensitive than power Doppler on this machine. The same colour Doppler pre-set was used for all examinations and no re-adjustment of Doppler parameters was performed: the gain setting for the colour Doppler was just below noise level and the system was adjusted to highest sensitivity for slow flow (Nyquist limit 0.014 m/s [=PRF: 256 Hz]), lowest wall filter and 7MHz

Doppler frequency) (15). During the examination, the patient was seated opposite the investigator with the hand in prone position.

The dorsal wrist was scanned longitudinally in the central position. Care was taken not to exert pressure with the transducer during the examination. The central position was chosen for evaluation as the probe position here is found to be the most reproducible position with little variation in acquisition in our experience. In the central position both the radiocarpal joint and the intercarpal joint were included and evaluated as one joint. The scan plane with the most colour Doppler activity was identified, the transducer was held in this position for a couple of heart cycles whereupon the image was frozen. The image with most colour Doppler activity was then selected from the cine-loop and stored. This image acquisition technique has previously shown an intra-class correlation coefficient (ICC) value of 0.77 when assessing the test-retest reliability (16), suggesting an excellent (ICC>0.75) reliability (17).

Four persons trained in ultrasound performed the US examinations during a 4-year period.

Image evaluation

One person (MS) trained in the SQS system performed the scoring according to the Szkudlarek *et al*. scoring system (4), and one person (KE) trained in the QS system performed the scoring according to Qvistgaard *et al*. (5). The images were sent by mail to the two persons on a CD. In all images the identity and visit number of the patient were blinded and the images were displayed in random order.

Image scoring

In both scoring systems the region of interest (ROI) was defined as the synovial tissue. In the SQS the Doppler activity was graded as followed: grade 0=no Doppler activity; grade 1=up to 2 single Doppler spots; grade 2=more than grade 1 and up to 50% Doppler activity in the ROI; grade 3=more than 50% Doppler activity in the ROI. For quantitative estimation of the vascularisation in the synovial tissue, the digitally stored CD image in DICOM format was transferred to a processing programme (ImagePro®). The synovial tissue inside the colour box was traced. The amount of colour pixels was then expressed in relation to the total amount of pixels in the marked ROI=the colour fraction (CF). Figure 1 shows two US images of the central position of the wrist. Image B shows the trace to estimate the CF. The time spent on scoring one image when received on a CD is approximately 1 minute for the QS system and 20 seconds for the SQS system.

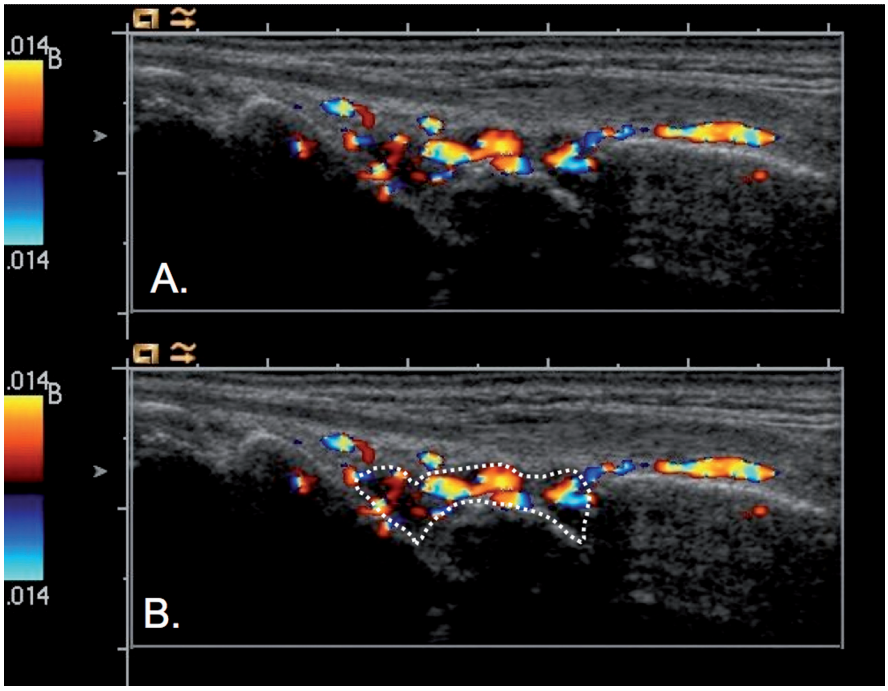

Fig. 1. US image of the central position of the wrist in. In image B the trace for the quantitative score is shown. The score of the two systems were: quantitative system (Colour Fraction) = 39.75%; semi-quantitative system = Score 2.

Statistical analysis and power considerations

All statistical analyses were based on the cohort of patients (14) that completed 1 year on therapy with thes same TNF alpha blocker agent; this sampling methodology reflects a per protocol analysis. All descriptive statistics were reported with a point estimate and a measure for dispersion (*e.g*. means and standard deviations or medians and interquartile ranges) depending on the data distribution. To assess the bivariate associations such as "Baseline ultrasound assessment as a predictor for change in disease activity" and the "Association between change in ultrasound score and change in disease activity" Spearman's rank-order correlation was applied. For a test of a correlation between two normal variables using Fisher's z statistic with a two-sided nominal significance level of 0.05, a sample size of 40 has a power of 81% to detect a correlation of at least 0.43 with a statistical significance level of 0.05.

To evaluate the different measures' sensitivity to change a unit-less measure was used, derived from the paired sample statistics (mean change from baseline) and divided with the corresponding standard deviation of the change. This measure is traditionally referred to as a standardised response mean (18) interpretation of a mutual set of SRMs, generated from the same longitudinal cohort is that the larger the SRM (in absolute terms) the better the sensitivity to change *per se*.

For a paired test comparing SRMs, with a two-sided significance level of 0.05, assuming a common SD of 1 and correlation between the measures compared 0.75, a sample size of 46 pairs has a power of 0.804 to detect a mean difference of 0.3 SRM points.

Table I. Baseline characteristics for the per protocol patients continuing therapy (>1 year).

Mean	n	Mean	SD	Median	Q_1	Q_3	Minimum	Maximum
Age, years	46	57.2	14.2	57	50	66	27	84
RA disease duration, years	46	9.7	8.7	6.5	3.5	13.0	0.1	34.2
Height, cm	45	168.5	8.2	168	163	173	148	191
Weight, kg	45	72.8	17.4	70.8	61.0	80.0	44.0	130.0
BMI, kg/m^2	45	25.6	5.4	24.8	22.6	27.9	16.9	45.0
HAQ20, score: 0–3	46	1.307	0.626	1.250	0.875	1.750	0.000	2.625
CF, ratio: 0–1	46	0.244	0.177	0.223	0.099	0.395	0.000	0.605
Semi-quantitative scoring, score: 0, 1, 2, 3	46	2.0	0.6	2	2	2	0	3
DAS28 (CRP)	46	5.1	1.5	5.3	4.0	6.0	1.4	7.8
TJC, count: 0–28	46	10.8	8.0	11	4	18	0	27
SJC, count: 0–28	46	7.4	5.6	7.5	2	11	0	21
Patient global, VAS: 0–100 mm	46	64.4	24.0	70	45	79	6	100
CRP, mg/l	46	21.1	24.4	12.5	2	30	1	90

Out of the 46 participating RA patients on TNFi therapy for a year, 9 (20%) were males.
RA: rheumatoid arthritis; BMI: body-mass index; HAQ: health assessment questionnaire; CF: colour fraction; TJC: tender joint count; SJC: swollen joint count; CRP: C-reactive protein; Q_3-Q_1: interquartile range.

Table II. Change from baseline analysed using analysis of covariance (ANCOVA).

Variable	Mean change	(95% CI)	*p*-value	appr. SD	SRM	(95% CI)
ΔQS	-0.08	(-0.13 ; -0.04)	0.001	0.16	-0.519	(-0.827 ; -0.211)
ΔSQS	-0.22	(-0.48 ; 0.04)	0.1013	0.91	-0.240	(-0.533 ; 0.053)
ΔDAS28	-2.21	(-2.55 ; -1.86)	<.0001	1.18	-1.865	(-2.343 ; -1.386)
ΔTJC	-7.50	(-9.03 ; -5.97)	<.0001	5.30	-1.415	(-1.824 ; -1.006)
ΔSJC	-5.54	(-6.36 ; -4.73)	<.0001	2.81	-1.974	(-2.470 ; -1.477)
ΔPatient global VAS	-32.96	(-39.22 ; -26.69)	<.0001	21.69	-1.520	(-1.944 ; -1.095)
ΔCRP	-16.38	(-17.79 ; -14.97)	<.0001	4.87	-3.366	(-4.112 ; -2.620)

QS: quantitative scoring; SQS: semi-quantitative scoring; DAS28: disease activity score based on a 28 joint count; TJC: tender joint count; SJC: swollen joint count; Patient global VAS: visual analogue scale; CRP: C-reactive protein; SRM: standardised response mean.

Results

Patient characteristics (baseline)

As presented in Table I, 46 patients (80% females) with RA who all completed one year on therapy with the same TNF alpha blocker agent were included in the study. The mean age was 57.2 (SD:14.2) years with a median disease duration of 6.5 years (IQR:3.5; 13.0) years at inclusion. The mean number of tender and swollen joints was 10.8 (SD:8.0) and 7.4 (SD:5.6), respectively. Patient global VAS was 64.4 (SD:24.0) mm and the median CRP level was 12.5 (IQR:2; 30) mg/L. Mean HAQ was 1.3 (SD:0.6) and the median Doppler activity measured in QS was 0.223 (IQR:0.099; 0.395), whereas the median value of the SQS was 2.0 (IQR:2; 2).

Clinical outcome (1 year from baseline)

After one year of treatment, 22 of the 46 patients were in clinical remission (DAS28<2.6) (mean 1.8; range 1.3–2.6), six had low disease activity (DAS28<3.2) (mean 2.8; range 2.6–3.1), and 18 were assessed to still have high disease activity (DAS28 >5.1) (mean 4.1; range 3.3–6.3).

Ultrasound results

In accordance with all other measurements that are assessing disease activity in RA, the mean of both QS and SQS decreased in the first year of treatment with an anti-TNF-α drug (Table II).

Of the 22 patients in clinical remission seven were Doppler negative. In the low disease activity group one of six was Doppler negative, and in the high disease activity group all 18 patients had Doppler activity.

The ability to register changes in Doppler activity was better when using QS system than the SQS system. The correlation between the change from baseline to 1 year in DAS28 and the two Doppler scoring systems was only modest. The correlation with QS was slightly better than the correlation seen with SQS (Fig. 2). The discriminative ability (ability to predict treatment success measured as decrease in DAS28) was poor for both scoring systems (data not shown).

Discussion

In this study of 46 patients with RA, changes in Doppler US were related to the clinical outcome after the first year of treatment with an anti-TNF drug. In general, Doppler activity decreased according to both scoring systems with little difference in sensitivity to change, predictive validity and discriminative ability. The sensitivity to change was only modest, however, better for the QS that the SQS. A small predictive value at baseline was found for the QS r_s=-0.24; whereas there was no evidence support-

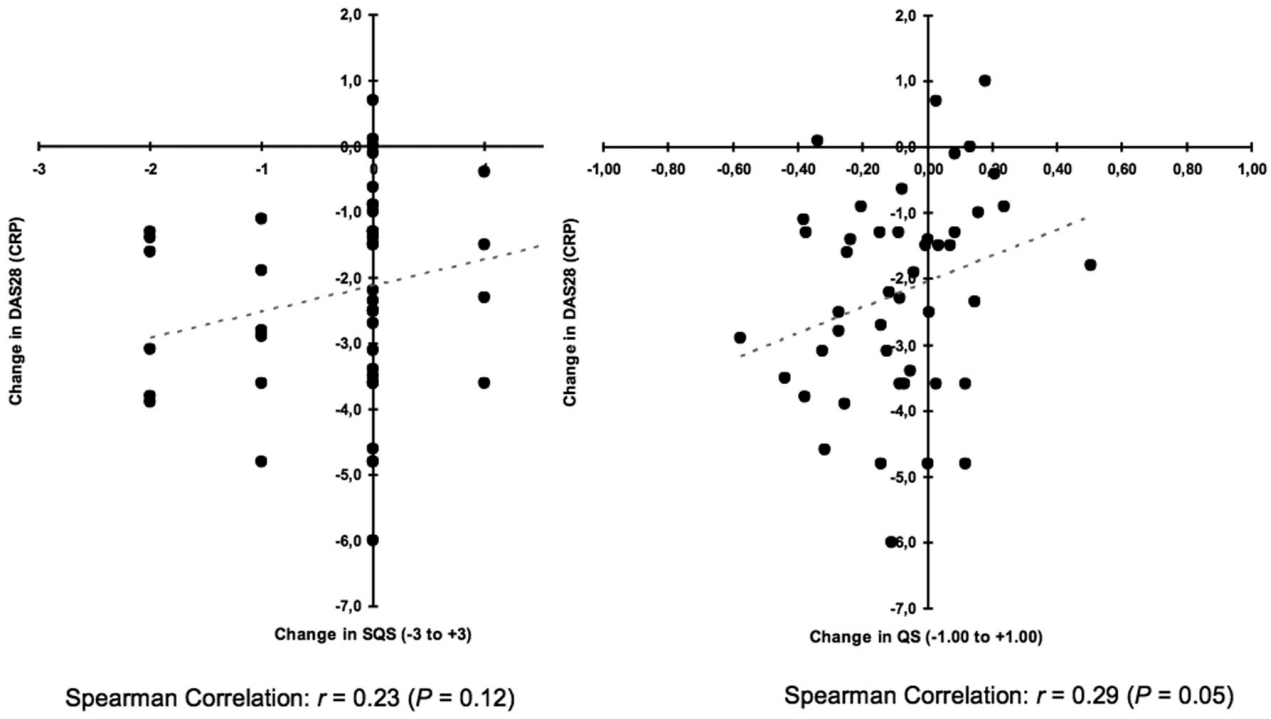

Fig. 2. Association between change in ultrasound (US) score and change in disease activity (DAS28).
A: X-axis: change in SQS (semi-quantitative score) and Y-axis: change in DAS28, both after one year of treatment with a TNF-alpha inhibitor.
B: X-axis: change in QS (quantitative score) and Y-axis: change in DAS28, both after one year of treatment with a TNF-alpha inhibitor.

ing SQS predictive ability with r_s=-0.05. The construct validity assessing changes in Doppler and changes in DAS28 was modest for both measures.

In a previous study, using the baseline data of this cohort, we investigated the inter-and intra-observer reliability and the correlation of the two Doppler scoring systems (14). In this study, we found an excellent intra-reader reliability of both scoring systems, with the ICC for the QS a little prior to SQS (QS; ICC=0.94 and SQS; ICC=0.82). The inter-reader agreement showed 95% limits of agreement for QS between -7.7% and +6.7% and for SQS between -0.8 and +0.8. Substantial relationship between the two scoring systems was seen in this study with r=0.73 (Spearman correlation coefficient). No correlation with DAS28 was found for any of the scoring systems (14).

The modest correlation between baseline Doppler and DAS28 and the insignificant association to the change over time assessed in relation to DAS28 might not be due Doppler being ineffective as marker of disease activity. It may also be caused by the opposite, *i.e*. DAS28 not being a perfect marker of inflammation. In our present cohort, some degree of persistent Doppler activity was observed in 15 of 22 patients in clinical remission (DAS28<2.6). The fact that DAS28 is not a perfect marker for clinical disease activity is supported by other studies that have found persistent Doppler activity in patients in clinical remission, but with progression of erosions (19-21). The development of erosions in patients with Doppler activity assessed as being in clinical remission and the superior predictive validity of the Doppler measurements may indicate that in some way Doppler US is a more sensitive measure of inflammation than DAS28 (21, 22). Recent work evaluating remission criteria in RA patients has underlined the crucial role of US Doppler examination when evaluating the disease activity and treatment strategy (23) and joints with US Doppler activity are often falsely assed as unaffected in clinical examination (24). These results all together underline the crucial role of correct US Doppler examination in patients with RA.

In contrast to the results in the present study, another study by Taylor *et al*. has displayed a predictive value of Doppler examination, as baseline Doppler assessment and progression in erosions after one year has been demonstrated in patients treated for RA (1). In the study by Taylor *et al*., all MCP joints were scanned, thus it could be argued that they could display a predictive value due to more joints being evaluated. However, in a previous study in RA patients we evaluated only one target joint (wrist) in accordance with the present study and found the Doppler measurements to be superior to all other disease markers, including DAS28 for their ability to predict which patients were still on the same biological agent after one year of treatment (22). Continuance on therapy is believed to be a good proxy for overall effectiveness and safety (25-26).

Whether to use QS or SQS systems must depend on the setting. In the daily clinical praxis, the SQS may be the most suitable system because the time spent on the scoring is a little shorter and no post-processing is needed. However, the post-processing in the QS scoring will soon be made easier by the introduction of software for calculating a CF directly on the ultrasound machines. This will make QS more

useful in clinical praxis. The fact that QS is a little superior to SQS in all the assessed statistical parameters suggests utilisation of the QS scoring system for research purposes.

In conclusion, both the quantitative and the semi-quantitative scoring systems were to some extent sensitive to change. Predictive validity and discriminant capacity of both systems showed only a modest association in patients completing one year of treatment with a TNF-α inhibitor.

The results do not need to reflect Doppler evaluation being ineffective, but may very well be caused by DAS28 not being a perfect marker of inflammation. The quantitative system (QS) is a little superior to the semi-quantitative (SQS) in all the assessed variables. The QS is more time consuming and the choice of scoring system must depends on individual circumstances, *e.g.* equipment and education.

Acknowledgments

Lene Terslev wishes to thank the Danish Rheumatism Association for support.

References

1. TAYLOR PC, STEUER A, GRUBER J *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.
2. TERSLEV L, TORP-PEDERSEN S, QVISTGAARD E *et al.*: Effects of treatment with etanercept (Enbrel, TNRF:Fc) on rheumatoid arthritis evaluated by Doppler ultrasonography. *Ann Rheum Dis* 2003; 62: 178-81.
3. 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.
4. SZKUDLAREK M, COURT-PAYEN, 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.
5. 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.
6. HAU M, KNEITZ C, TONY HP, KEBERLE M, JAHNS R, JENETT M: High resolution ultrasound detects a decrease in pannus vascularisation of small finger joints in patients with rheumatoid arthritis receiving treatment with soluble tumour necrosis factor alpha receptor (etanercept). *Ann Rheum Dis* 2002; 61: 55-8.
7. STONE M, BERGIN D, WHELAN B, MAHER M, MURRAY J, MCCARTHY C: Power Doppler ultrasound assessment of rheumatoid hand synovitis. *J Rheumatol* 2001; 28: 1979-82.
8. FILIPPUCCI E, IAGNOCCO A, SALAFFI F, CERIONI A, VALESINI G, GRASSI W: Power Doppler sonography monitoring of synovial perfusion at the wrist joints in patients with rheumatoid arthritis treated with adalimumab. *Ann Rheum Dis* 2006; 65: 1433-7.
9. 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.
10. BOERS M, BROOKS P, STRAND CV, TUGWELL P: The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol* 1998; 25: 198-9.
11. JOHNSTON MV, KEITH RA, HINDERER SR: Measurement standards for interdisciplinary medical rehabilitation. *Arch Phys Med Rehabil* 1992; 73: S3-23.
12. ØSTERGAARD M, WIELL C: Ultrasonography in rheumatoid arthritis: a very promising method still needing more validation. *Curr Opin Rheumatol* 2004; 16: 223-30.
13. ARNETT FC, EDWORTHY SM, BLOCH DA *et al.*: The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988; 31: 315-24.
14. TERSLEV L, ELLEGAARD K, CHRISTENSEN R *et al.*: Head-to-head comparison of quantitative and semi-quantitative ultrasound scoring systems for rheumatoid arthritis: reliability, agreement and construct validity. *Rheumatology* (Oxford) 2012; 51: 2034-8.
15. TORP-PEDERSEN ST, TERSLEV L: Settings and artefacts relevant in colour/power Doppler ultrasound in rheumatology. *Ann Rheum Dis* 2008; 67: 143-9.
16. ELLEGAARD K, TORP-PEDERSEN S, LUND H *et al.*: Quantification of colour Doppler activity in the wrist in patients with rheumatoid arthritis - the reliability of different methods for image selection and evaluation. *Ultraschall Med* 2008; 29: 393-8.
17. FLEISS JL: The design and analysis of clinical experiments. New York, Wiley, 1986.
18. BINGHAM CO, III, ALTEN R, BARTLETT SJ *et al.*: Identifying preliminary domains to detect and measure rheumatoid arthritis flares: report of the OMERACT 10 RA Flare Workshop. *J Rheumatol* 2011; 38: 1751-8.
19. 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.
20. WAKEFIELD RJ, FREESTON JE, HENSOR EM, BRYER D, QUINN MA, EMERY P: Delay in imaging versus clinical response: A rationale for prolonged treatment with anti-tumor necrosis factor medication in early rheumatoid arthritis. *Arthritis Rheum* 2007; 57: 1564-7.
21. BROWN AK, CONAGHAN PG, KARIM Z *et al.*: An explanation for the apparent dissociation between clinical remission and continued structural deterioration in rheumatoid arthritis. *Arthritis Rheum* 2008; 58: 2958-67.
22. ELLEGAARD K, CHRISTENSEN R, TORP-PEDERSEN S *et al.*: Ultrasound Doppler measurements predict success of treatment with anti-TNF-α drug in patients with rheumatoid arthritis: a prospective cohort study. *Rheumatology* (Oxford) 2011; 50: 506-12.
23. NEMOTO T, OGASAWARA M, MATSUKI Y *et al.*: Can routine clinical measures predict ultrasound-determined synovitis and remission in rheumatoid arthritis patients? *Clin Exp Rheumatol* 2014; 32: 54-60.
24. MURAYAMA G, OGASAWARA M, NEMOTO T *et al.*: Clinical miscount of involved joints denotes the need for ultrasound complementation in usual practice for patients with rheumatoid arthritis. *Clin Exp Rheumatol* 2013; 31: 506-14.
25. GALINDO-RODRIGUEZ G, VINA-ZUBIETA JA, RUSSELL AS, SUAREZ-ALMAZOR ME: Disappointing longterm results with disease modifying antirheumatic drugs. A practice based study. *J Rheumatol* 1999; 26: 2337-43.
26. CHOY EH, SMITH C, DORE CJ, SCOTT DL: A meta-analysis of the efficacy and toxicity of combining disease-modifying anti-rheumatic drugs in rheumatoid arthritis based on patient withdrawal. *Rheumatology* (Oxford) 2005; 44: 1414-21.