Acta Radiologica

Draft Manuscript for Review

Diagnostic accuracy of simplified ultrasound hand examination protocols for detection of inflammation and disease burden in patients with rheumatoid arthritis.

Journal:	*Acta Radiologica*
Manuscript ID	SRAD-2017-0419.R2
Manuscript Type:	Original Article
Keyword:	Ultrasound-Power Doppler < Modalities/Techniques, Hand < Structures, musculoskeletal ultrasound

Diagnostic accuracy of simplified ultrasound hand examination protocols for detection of inflammation and disease burden in patients with rheumatoid arthritis

Abstract

Background: There is no consensus regarding the minimum of joints that should be included in an ultrasound (US) scoring system to reliably assess for disease activity in rheumatoid arthritis (RA).

Purpose: To assess whether simplified US protocols for hand examination are as informative as the examination of 22 joints in patients with RA, and to correlate the US parameters with disease activity (DAS-28).

Material and Methods: This is a cross-sectional study of 224 RA patients stratified based on their DAS-28 scores and assessed using eight preselected US examination protocols, including 22, 18, 16, 14, 10, 8 and two different combinations of 4 joints, respectively.

Results: We found a significant difference between different US hand scores regarding their ability to detect active inflammation and erosions. DAS-28 scores correlated very well with the Power Doppler (PD) scores generated by all eight US examination protocols (r=0.89-1,

P<0.05), irrespective of patients' disease activity. Simplified US scores missed information on presence of PD in 20.6 - 40.2% patients (P<0.05), and misdiagnosed non-erosive hand RA in 12 - 38.4% patients (P<0.05), depending on the number of joints excluded from US hand examination.

Conclusion: Preselected simplified US scores are less reliable in appreciating the disease burden when compared with an extended protocol for 22 joint US examination, raising clinicians' awareness regarding the need to comprehensively assess multiple hand joints to reliably rule out subclinical inflammation.

Keywords: hand, ultrasound, Power Doppler ultrasound.

Introduction

RA is a chronic inflammatory condition associated with well-recognised inflammatory joint features, which are amenable to US examination. The use of US facilitated a significant progress in the early diagnosis of RA, enabling a better assessment of the disease activity, prognosis and response to different therapeutic interventions. The implementation of US scoring systems in addition to clinical examination could help standardising the way RA is monitored; however, based on local availability of US and sonographer expertise, different scoring systems have been used in clinical practice. Despite significant research progress in supporting the role of US in RA, no consensus was reached with regard to what scoring system is the most useful. The OMERACT US Task Force defined the US pathology associated with RA (1), which combines tendon, joint and bone abnormalities (1, 2). The presence of Power Doppler (PD) is recognized as a reliable objective measure of active joint inflammation (3). Different semi-quantitative scoring systems are currently used for

assessing synovial hypertrophy (SH), joint effusion, tendon abnormalities and erosions (4), and protocols for hand and feet US examination are well-established (5).

A recent systematic review of the scoring systems used to evaluate synovitis in RA found difficult to determine the least number of joints that needed to be assessed for a global US score (1). The purpose of our study was to investigate how much we can simplify the US examination of hands in RA, without compromising the ability of a certain US scoring system to evaluate the disease activity and damage associated with hand RA. The authors focused on the US examination of hands as this is the most commonly used in routine clinical practice.

Material and Methods

This is a real-life, cross-sectional study, which evaluated patients referred to our US rheumatology outpatient clinics, presenting with inflammatory sounding hand joint pains. The patients were referred based on clinician indication to have an US scan to help with identifying joint inflammation that was not confidently assessed clinically. We examined 604 patients between Jan 2012 and August 2015. For each patient, a set of demographic, clinical and laboratory data were recorded at the time of the scan. Of 604 patients referred to our clinic, 224 patients with RA were included in the study analysis based on their final diagnosis made using the 2010 ACR/EULAR classification criteria, following complete investigations and revision of the clinical notes. Fig. 1 details the patient selection and stratification based on DAS-28 scores.

This study evaluated the same set of reported outcomes and clinical and laboratory parameters for all the patients, to ensure homogeneity of the collected data. The following information was analyzed: disease duration (in months), hand tender joint count (TJC) and

swollen joint count (SJC), as well as a patient reported global disease assessment score (GVAS).

Additional data about the high sensitivity C-reactive protein (hsCRP), erythrocyte sedimentation rate (ESR), presence of rheumatoid factor (RF), anti citrullinated cyclic peptides antibodies (ACPA) and anti-nuclear antibodies (ANA) were collected at the time of the scan (needed to exclude associated pathology).

For each patient, a detailed record was compiled of their medication at the time of the US scan, including paracetamol and NSAIDs, disease-modifying drugs (DMARDs), biologic therapies and glucocorticoids, either oral or intramuscular depot injection.

The US protocol examination used included the extensor tendons and 22 joint assessments (dorsal longitudinal and transverse views of wrists, including extensor tendons, metacarpo-phalangeal – MCP joints, and proximal interphalangeal – PIP joints), as per our local clinic protocol. The same US examination protocol was used for each patient, irrespective of their hand symptoms. The US findings were scored according to the OMERACT scoring system (1). The hand US examination was performed by two clinicians (CC and LA) in the same session, and for each patient a consensus was obtained.

We used a Logiq S8 US machine (GE Healthcare, Wauwatosa, Wisconsin, WI, USA), equipped with a multi-frequency linear matrix array transducer (6-15 MHz). B-mode and PD machine setting were optimized and standardized for all our patients' US examinations. The settings used were: B-mode frequency 11-15 MHz depending on the depth of the anatomical area, Doppler frequency 7.5-15, depending on the depth of anatomical area; Doppler gain 18-20 dB, low wall filters and pulse repetition frequency around 800 Hz. In this study, we only used Power Doppler (PD) mode.

The information collected comprised the following US parameters: SH grade (graded 1-3), erosions (present/absent), PD signal (graded 1-3), joint effusion (present/absent), osteophytes (present/absent), and tendon abnormalities (PD signal present/absent) using the US definition of joint pathology as defined by the OMERACT group (2) (Fig. 2 exemplifies two MCP joints with different SH and PD grades). Well controlled disease was defined as PD score zero (including joints and tendons).

To address our research question and assess how many joints would require scanning, and which joints are most likely to provide the answer as to whether or not there is active disease, we tested and compared the following scoring systems (bilateral examination):

- 22 joints (MCPs, PIPs, wrists)
- 18 joints (wrists, MCP 2-5 and PIP 2-5)
- 16 joints (MCP 2-5 and PIP 2-5)
- 14 joints (wrists, MCP 2-4 and PIP 2-4)
- 10 joints (wrists, MCP 2-3 and PIP 2-3)
- 8 joints (MCP 2-3 and PIP 2-3)
- 4 joints (wrists +MCP5)
- 4 joints (MCP 2-3)

The above joint combination score was selected based on our experience of performing US examination of hands in more than 1000 patients, which identified that the most affected joints in RA were the wrists, MCP 2,3 and 5, and PIP 2 and 3 (unpublished observation).

The SH grade 1 score was calculated as the total number of the joints with SH grade 1, the SH grade 2 score as the total number of the joints with SH grade 2, and the SH grade 3 score as the total number of the joints with SH grade 3 per patient. The total PD score was the sum

of all individual PD scores per patient, and the erosion score was calculated as the total number of erosions per patient.

Data about active inflammation affecting tendons overlying the above mentioned joints were also collected and reported separately. The total grey scale scores and PD scores for joints were calculated as a sum of the individual scores for all the joints included in the US examination protocol the score refers to. The duration of the US examination was approximately 25 minutes/patient. This 22 joint protocol is used routinely in our US clinics, which have 30 minute slots for clinical and US examination of patients with RA.

Descriptive statistics were used to characterize the RA population, and Student T test, Mann-Whitney U and Kuskal-Wallis tests were implemented for the assessment of different parameters and US scoring systems (IBM SPSS Statistics 22, IBM Corporation, 1 New Orchard Road, Armonk, New York 10504-1722, US). A P-value of <0.05 was considered a statistically significant result. Spearman's correlation coefficients were used to correlate permutations of pairs of US scores and the total PD scores with the disease activity, as assessed by the disease activity score assessing 28 joints (DAS-28).

The data were collected as standard of practice. The study analyzed cross-sectionally the results of the US examinations of patients seen in our US clinics over a defined period of time. No ethical approval or patient's consent were required as no patient information was used for teaching or new intervention research. The results of our study analysis had no impact on the clinical management of patients and their confidentiality was maintained.

Results

To characterize in detail our RA cohort, we stratified patients based on DAS-28 (ESR) assessment of disease activity (Table 1). Demographic parameters were similar among different disease activity groups. As expected, patients with higher disease activity scores had significantly higher TJC, SJC, ESR and GVAS, while the CRP levels were similar between different groups. Both objective and subjective parameters included in the DAS-28 composite score were significantly increased in patients with active disease compared to moderate or low disease activity groups and with the group in remission.

There were no significant differences in the total US scores including the majority of US parameters, or in the disease duration or type of medication used (for both conventional and biologic DMARDs). The only significant difference was between the proportion of patients with SH grade 2 at the US examination of their hands, which was higher in patients with moderately-active and strongly-active RA ($P<0.05$) (Table 1). The SH grade 2 total score also correlated with the SJC ($r=0.89$, $P<0.05$).

The comparative analyse of the above-mentioned US scores showed no significant differences between the ability of the pre-selected US scores to capture information regarding SH grade 2 and 3, and the total PD scores per patient; however, the proportion of patients with no active disease at the US examination differed significantly based on the number of joints included in the examination protocol ($P<0.05$) (Table 2). Similarly, different US scores varied significantly in their ability to assess the total erosion score per patient and the proportion of patients with erosions ($P<0.05$). By simplifying the US examination of the hand in RA patients, active RA was underdiagnosed in a proportion of 20.6 to 40.2% of patients; similarly, the erosive burden was underappreciated in 12 - 28.4% RA patients (Table 2).

Strong correlations were found between the PD score generated by the 22 joint examination and all of the other US score combinations ($r = 0.68 - 0.74$, $P<0.05$). The scores that

correlated very strongly were those assessing 8, 10 and 14 joints (r = 0.92-0.96, P<0.05). The weakest correlation was found between the 8 and the 4 joint score (wrist and MCP 5 bilaterally) (r=0.28, P<0.05) (Suppl. Table 1).

The permutation comparisons between pairs of US scores related to their ability to detect the presence active joint inflammation found no significant differences between the total PD scores assessed by 8, 10, 12 and 16 joint US scores and 10, 12, 16 and 18 joint scores, respectively (Suppl. Table 2). Similarly, the total grey scale score (combining the total scores for SH grade 2 and 3) identified no significant differences between the permutation comparisons between the scores assessing 8, 10, 14, 16 and 18 joints (Suppl. Table 3).

The analysis was also focused on correlating the total PD scores derived from all the pre-set US examination protocols with the DAS-28 scores in patients stratified based on their disease activity, to identify if certain US hand examination protocols can be used differentially in patients with active disease compared to patients in remission. All the total PD scores derived from the eight US examination protocols correlated very strongly with DAS-28 assessment, irrespective of how well the disease was controlled (r = 0.88-1, P<0.005).

Discussion

This is the first large cross-sectional study correlating different US examination protocols (derived from a 22-hand joint comprehensive score) with DAS-28 score in patients in RA, stratified based on their disease activity.

Quantitative and semi-quantitative US scores have been previously compared in RA (3), and US examination have been found to be sensitive to therapeutic interventions (4-8). A comprehensive study comparing several US score systems in RA found that all were sensitive

to change when assessing the response of RA patients to adalimumab (9, 10). In addition, simplified US scores (including 6 or 12 joints) have previously been compared with extensive US protocol examinations (assessing 12, and 44 joints respectively), and showed good sensitivity to change in three separate studies (11-13). However, none of these studies stratified patients based on their disease activity scores or included RA patients based on the clinical indication to have an US scan, as it is the case with our study. The need to use a comprehensive US scoring system, capturing both active and chronic inflammatory changes for assessment of RA disease activity, is supported by the good correlation between US and MRI findings (8, 14). The presence of SH and PD signal was found to be associated with structural damage in RA (15), even in patients in clinical remission (16), and was associated with risk of flares (17, 18).

The role and reliability of US in the disease activity assessment in patients with RA is supported by several studies (19-21).

Previous studies reported good correlation between hand US scores and DAS-28 assessment using three different US scores (22, 23), result that was also replicated by our study, which included a larger number of joint combinations, and also assessed US parameters stratified based on the DAS-28 scores.

Our comparative analysis of several US scoring systems showed that there is significant difference in terms of the equivalence of several US hand-scoring systems. Our study found that age, duration of symptoms, duration of disease, type of medication and total PD score generated by US examination of hands were not able to inform about the inclusion of patients in one specific disease activity group, as patients stratified based on DAS 28 scores had similar parameters.

In addition to previous studies, we have been interested in exploring the amount and significance of missed information related to the use of simplified US hand examination protocols. A significant proportion of patients have been diagnosed as having well-controlled or non-erosive hand RA by using US protocols limiting the number of joints examined (10.6 - 40.2 % and 12-34.8%, respectively). Our study found that the assessment of our preselected 8, 10 and 14 joints captured comparable amounts of information regarding disease activity in RA (still misdiagnosing around 40% of patients as having well controlled disease, equivalent to PD score zero), while the two 4 joint scores missed significant information when compared to the others (around 60% patient were diagnosed in remission despite having active disease at least in one joint). The scores including 20 and 22 joints captured more information than the 8, 10 and 14 joint scores, even if all the eight US scores we explored correlated very well with the DAS-28 assessment. This is particularly relevant for our patient group, characterized by a small number of active joints and clinical indication to have an US scan to establish if their disease was well controlled or not. In this context, underdiagnosing active disease would have erroneously led to classifying our patients as being in remission. The clinical consensus is that we cannot predict which joints are the most likely to flare in patients with RA patients; therefore examining only the joints that previously flared using a patient-tailored US protocol is not justified.

Even if a comprehensive hand joint score is time-consuming, it can provide significant additional information compared to a simplified score, as our study showed. As expected, all the scores correlated very well with each other, because they are derived from a comprehensive US hand score, while missing significant information proportional to the number of joints excluded from US examination. All the pre-selected US hand scores correlated with the disease activity scores, despite the fact that the patient groups stratified based on disease activity had similar median total PD scores. This showed that subclinical in

inflammation can be find in similar proportion in RA patients, irrespective of their degree of chronic joint changes that are likely to influence their DAS-28 score.

Limitations: Our study did not have strict inclusion criteria: the patients were included based on clinical indication to exclude subclinical synovitis. Therefore, there is a significant selection bias, as the study did not capture patients with obvious active synovitis detected by clinical examination. In this particular clinical context, detection of active disease in at least one joint is clinically relevant, as US examination triggered treatment optimization to minimize joint damage (e.g. guided steroid injection targeting the active joints or escalation of therapy). In conclusion, even if simplified US scores for hand assessment of RA disease activity can be useful in practice, by examining additional joints, clinicians are able to detect subclinical inflammation, which is not captured by the simplified US scores. If previously studies re-assured clinicians that various US examination protocols correlated well with the DAS-28 assessment or were sensitive to change following therapy, our study showed that a significant proportion of patients can be misclassified as having well-controlled or non-erosive disease as a result of simplified US protocols. Further studies, including large longitudinal cohorts, are needed to establish the smaller number of joints needed to be examined to minimize the risk of under detecting subclinical inflammation in patients with hand RA.

Conflict of interest:

The authors declared no conflicts of interest.

References:

1. Mandl P, Naredo E, Wakefield RJ, et al. 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;38:2055-2062.
2. Wakefield RJ, Balint PV, Szkudlarek M, et al. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. J Rheumatol 2005;32:2485-2487.
3. Ellegaard K, Terslev L, Christensen R, et al. Comparison of discrimination and prognostic value of two US Doppler scoring systems in rheumatoid arthritis patients: a prospective cohort study. Clin Exp Rheumatol 2014;32:495-500.
4. Epis O, Scioscia C, Locaputo A, et al. Use of ultrasound in treatment decisions for patients with rheumatoid arthritis: an observational study in Italy. Clin Rheumatol 2016;35:1923-1929.
5. Hau M, Kneitz C, Tony HP, et al. 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;61:55-58.
6. Terslev L, Torp-Pedersen S, Qvistgaard E, et al. Estimation of inflammation by Doppler ultrasound: quantitative changes after intra-articular treatment in rheumatoid arthritis. Ann Rheum Dis 2003;62:1049-1053.
7. Epis O, Filippucci E, Delle Sedie A, et al. Clinical and ultrasound evaluation of the response to tocilizumab treatment in patients with rheumatoid arthritis: a case series. Rheumatol Int 2014;34:737-742.

8. Rowbotham EL, Grainger AJ**.** Rheumatoid arthritis: ultrasound versus MRI. Am J Roentgenol 2011;197:541-546.
9. Hammer HB, Kvien TK**.** Comparisons of 7- to 78-joint ultrasonography scores: all different joint combinations show equal response to adalimumab treatment in patients with rheumatoid arthritis. Arthritis Res Ther 2011;13:R78.
10. Filippucci E, Iagnocco A, Salaffi F, et al**.** Power Doppler sonography monitoring of synovial perfusion at the wrist joints in patients with rheumatoid arthritis treated with adalimumab. Ann Rheum Dis 2006;65:1433-1437.
11. Perricone C, Ceccarelli F, Modesti M, et al. The 6-joint ultrasonographic assessment: a valid, sensitive-to-change and feasible method for evaluating joint inflammation in RA. Rheumatology (Oxford) 2012;51:866-73.
12. Naredo E, Valor L, De la Torre I, et al. Ultrasound joint inflammation in rheumatoid arthritis in clinical remission: how many and which joints should be assessed? Arthritis Care Res (Hoboken) 2013;65:512-7.
13. Naredo E, Rodriguez M, Campos C, 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.
14. Schmidt WA, Schicke B, Ostendorf B, et al. Low-field MRI versus ultrasound: which is more sensitive in detecting inflammation and bone damage in MCP and MTP joints in mild or moderate rheumatoid arthritis? Clin Exp Rheumatol 2013;31:91-96.
15. Dougados M, Devauchelle-Pensec V, Ferlet JF, et al. The ability of synovitis to predict structural damage in rheumatoid arthritis: a comparative study between clinical examination and ultrasound. Ann Rheum Dis 2013;72:665-671.
16. Brown AK, Quinn MA, Karim Z, et al. Presence of significant synovitis in rheumatoid arthritis patients with disease-modifying antirheumatic drug-induced clinical remission:

evidence from an imaging study may explain structural progression. Arthritis Rheum 2006;54:3761-3773.

17. Nguyen H, Ruyssen-Witrand A, Gandjbakhch F, et al**.** Prevalence of ultrasound-detected residual synovitis and risk of relapse and structural progression in rheumatoid arthritis patients in clinical remission: a systematic review and meta-analysis. Rheumatology (Oxford) 2014;53:2110-2118.
18. Bellis E, Scire CA, Carrara G, et al. Ultrasound-detected tenosynovitis independently associates with patient-reported flare in patients with rheumatoid arthritis in clinical remission: results from the observational study STARTER of the Italian Society for Rheumatology. Rheumatology (Oxford) 2016;55:1826-1836.
19. Naredo E, Bonilla G, Gamero F, et al**.** Assessment of inflammatory activity in rheumatoid arthritis: a comparative study of clinical evaluation with grey scale and power Doppler ultrasonography. Ann Rheum Dis 2005;64:375-381.
20. Mandl P, Balint PV, Brault Y, et al. Metrologic properties of ultrasound versus clinical evaluation of synovitis in rheumatoid arthritis: results of a multicenter, randomized study. Arthritis Rheum 2012;64:1272-1282.
21. Szkudlarek M, Terslev L, Wakefield RJ, et al. Summary Findings of a Systematic Literature Review of the Ultrasound Assessment of Bone Erosions in Rheumatoid Arthritis. J Rheumatol 2016;43:12-21.
22. Dougados M, Jousse-Joulin S, Mistretta F, et al. Evaluation of several ultrasonography scoring systems for synovitis and comparison to clinical examination: results from a prospective multicentre study of rheumatoid arthritis. Ann Rheum Dis 2010;69:828-833.
23. Rosa J, Ruta S, Saucedo C, et al. Does a Simplified 6-Joint Ultrasound Index Correlate Well Enough With the 28-Joint Disease Activity Score to Be Used in Clinical Practice? J Clin Rheumatol 2016;22:179-183.

Figure legends:

Figure 1. Flowchart of the study population.

Figure 2: Examples of MCP joint grading: SH grade 3 and PD grade 2 (above), and SH grade 2 and PD grade 3 (below).

Table 1- Comparison between RA patient groups stratified based on their DAS 28 scores using the 22 joint US scoring system as detailed above (Kruskal-Wallis test, $p<0.05$ shows a significant difference between the patient groups).

RA patients stratified based on disease activity	DAS28 >5.1	DAS 28 3.2-5.1	DAS 28 2.6-3.2	DAS 28 <2.6	P value
Age Mean ± SD	55.6 ± 13.8	53.2 ± 16.	54.2 ± 15.5	50.3 ± 15.3	P=0.44
% Female	80.0	89.5	71.4	75.6	P=0.11
Disease duration (months): Mean ± SD	120.3 ± 107	111.7± 135	70.5 ± 58.4	95.4 ± 169.9	P=0.51
% of patients on steroids at the time of the scan (all patients were on ≤ 10 mg daily)	38	43.3	29.0	36.8	P=0.36
% of patients on conventional DMARDs at the time of the scan	74	54.2	64.5	65.8	P=0.11

% of patients on biologic treatment at the time of the scan	24	19.3	29.0	26.3	P=0.67
CRP Mean ± SD	8.2 ± 10.8	6.1 ± 14.8	4.3 ± 7.9	4.3 ± 9.6	P=0.42
ESR Mean ± SD	31.1 ± 26.7	16.4 ± 15.1	11.5 ± 15.5	6.8 ± 6.7	**P<0.05**
SJC Mean ± SD	5.3 ± 4.6	2.4 ± 2.5	1.5 ± 1.9	0.5 ± 0.8	**P<0.05**
TJC Mean ± SD	17.1 ± 7.5	6.7 ± 5.2	4.5 ± 4.6	1.3 ± 2.0	**P<0.05**
GVAS Mean ± SD	74.6 ± 19.8	48.4 ± 26.4	34.4 ± 23.3	25 ± 26.2	**P<0.05**
Mean DAS 28 score ± SD	5.9 ± 0.75	3.9 ± 0.5	2.9 ± 0.16	1.7 ± 0.6	**P<0.05**
Total number of joints with SH grade 1 / patient Mean ± SD	2.4 ± 3.1	2.6 ± 3.7	1.7 ± 2.7	1.2 ± 1.8	P=0.52
Percentage of patients with joints with SH grade 1:	58.0	66.3	51.6	47.4	P=0.23
Total number of joints with SH grade 2 / patient Mean ± SD	2.4 ± 3.4	2.2 ± 3.2	1.6 ± 4.4	1.3 ± 2.5	P=0.36
Percentage of patients with joints with SH grade 2:	58.0	53.0	48.4	29.0	**P<0.05**
Total number of joints with SH grade 3 / patient Mean ± SD	2.1 ± 2.8	1.2 ± 1.7	1.1 ± 2.1	0.7 ± 1.3	P=0.49
Percentage of patients with joints with SH grade 3:	52.0	44.6	35.5	31.6	P=0.23
PD score	1.9 ± 2.9	1.2 ± 2.4	1.1 ± 1.5	0.7 ± 1.3	P=0.58

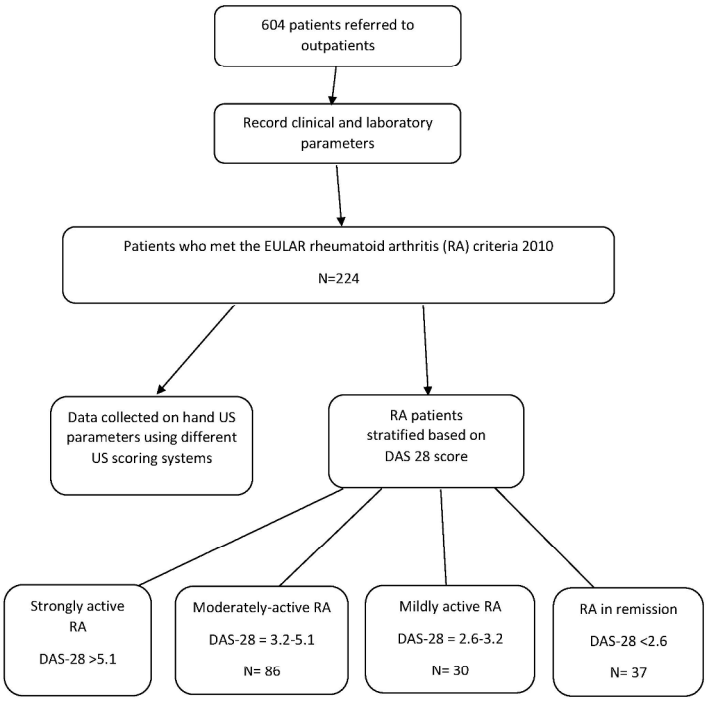
Mean ± SD					
Percentage of patients with PD signal	58.0	42.2	45.2	36.8	P=0.19
Total number of joints with erosions / patient Mean ± SD	6.5 ± 7.3	5.0 ± 5.5	4.3 ± 3.8	2.7 ± 4.1	P=0.17
Percentage of patients with erosions:	64.0	49.4	64.5	39.5	P=0.09
Percentage of patients with tendon abnormalities (GS score ≥ 2)	8.16	10.4	13.3	10.81	P=0.42
Percentage of patients with active tenosynovitis (PD score ≥ 1)	6.12	5.81	8.13	10.81	P=0.13

Table 2: Comparison between 8 different US scores (P<0.05 was considered significant).

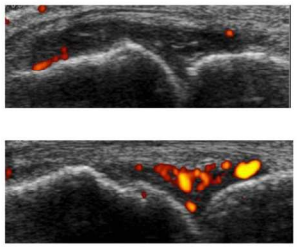
US findings/ scores	22 joints (wrists, MCPs, PIPs)	18 joints (wrists, MCP 2-5, PIP 2-5)	16 joints (MCP 2-5, PIP 2-5)	14 joints (wrists, MCP 2-4, PIP 2-4)	10 joints (wrists, MCP2-3, PIP 2-3)	8 joints (MCP2-3, PIP 2-3)	4 joints (wrists, MCP5)	4 joints (MCP 2-3 bilaterally)	P value
SH grade 2 score /patient	Median: 0 IQR: 3	Median: 0 IQR:1	Median: 0 IQR:1	Median: 0 IQR: 1	Median: 0 IQR: 1	Median: 0 IQR: 0	Median: 0 IQR: 0	Median: 0 IQR:1	-
Mean SH grade 2 score ± SD:	2 ± 3.21	1.19 ± 2.32	0.9 ± 1.83	0.71 ± 1.4	0.7 ± 1.39	0.15 ± 0.4	0.43 ±0.92	1.19 ±2.32	0.43
Percentage of patients with no evidence of SH grade 2:	51.8	64.3	64.3	66.5	69.6	69.6	89.3	76.8	0.15
SH grade 3 score /patient	Median: 0 IQR: 2	Median: 0 IQR:1	Median: 0 IQR:1	Median: 0 IQR:1	Median: 0 IQR: 1	Median: 0 IQR: 1	Median: 0 IQR: 0	Median: 0 IQR: 0	-
Mean SH grade 3 score ± SD:	1.3 ± 2.08	0.71 ± 1.41	0.7 ± 1.40	0.58 ±1.17	0.5 ± 0.97	0.5 ± 0.95	0.09 ±0.34	0.35±0.74	0.15
Percentage of patients with no evidence of SH grade 3:	57.1	69.2	69.2	71.4	74.1	74.1	93.3	77.7	0.15
PD score/patient	Median: 0 IQR: 2	Median: 0 IQR:1	Median: 0 IQR:1	Median: 0 IQR: 0	Median: 0 IQR: 0	Median: 0 IQR: 0	Median: 0 IQR: 0	Median: 0 IQR: 0	-

Mean PD score ± SD:	1.28 ± 2.31	0.69 ± 1.68	0.66 ± 1.66	0.61 ± 1.39	0.48 ±1.1	0.45 ± 1.07	0.09 ±0.36	0.32 ± 0.76	0.15
Percentage of patients with well controlled hand RA (PD score = 0)	54.0	74.6	75.9	75.4	77.7	79.0	94.2	81.3	**< 0.05**
Percentage of patients misdiagnosed with well controlled disease by the simplified US scores	N/A	20.6	21.9	21.4	23.7	25	40.2	27.3	**<0.05**
Erosion score/patient	Median: 2 IQR: 7.75	Median: 1 IQR: 4	Median: 1 IQR: 4	Median: 1 IQR:4	Median: 1 IQR: 3	Median IQR:3	Median: 1 IQR:1	Median: 1 IQR:2	**< 0.05**
Percentage of patients with non-erosive hand RA:	30.4	42.4	42.4	46.9	46.9	46.9	68.8	55.4	**< 0.05**
Percentage of patients misdiagnosed with non-erosive hand RA by the	N/A	12	12	16.5	16.5	16.5	38.4	25	**< 0.05**

simplified US scores									



210x297mm (300 x 300 DPI)



297x420mm (300 x 300 DPI)

Suppl. Table 1: Permutation correlations between pairs of US scoring systems assessing the total PD score (Spearman's correlation rank test, P<0.05 shows significant correlation).

Joints with PD	22 joints	18 joints (wrists + MCP 2-5, PIP 2-5)	16 joints (MCP 2-5, PIP 2-5)	14 joints (wrists + MCP 2-4, PIP 2-4)	10 joints (wrists MCP 2-3, PIP 2-3)	8 joints (MCP 2-3, PIP 2-3)	4 joints (wrists +MCP5)	4 joints (MCP 2-3 bilaterally)
4 joints (MCP 2-3)	**r = 0.64 P <0.05**	**r =0. 85 P <0.05**	**r = 0.87 P ≤ 0.05**	**r = 0.87 P ≤ 0.05**	**r = 0.91 P ≤ 0.05**	**r = 0.94 P ≤ 0.05**	**r =0.29 P ≤ 0.05**	-
4 joints (wrists, MCP5)	**r = 0.37 P ≤ 0.05**	**r = 0.48 P ≤ 0.05**	**r = 0.36 P ≤ 0.05**	**r = 0.44 P ≤ 0.05**	**r = 0.41 P ≤ 0.05**	**r = 0.28 P ≤ 0.05**	-	**r =0.29 P ≤ 0.05**
8 joints (MCP2-3, PIP 2-3)	**r = 0.68 P ≤ 0.05**	**r = 0.90 P ≤ 0.05**	**r = 0.93 P ≤ 0.05**	**r = 0.92 P ≤ 0.05**	**r = 0.96 P ≤ 0.05**	-	**r = 0.28 P ≤ 0.05**	**r = 0.94 P ≤ 0.05**
10 joints (wrists, MCP 2-3, PIP 2-3)	**r = 0.69 P ≤ 0.05**	**r = 0.93 P=<0.05**	**r = 0.90 P ≤ 0.05**	**r = 0.96 P ≤ 0.05**	-	**r = 0.96 P ≤ 0.05**	**r = 0.41 P ≤ 0.05**	**r = 0.91 P ≤ 0.05**
14 joints (wrists, + MCP 2-4, PIP 2-4)	**r = 0.72 P ≤ 0.05**	**r = 0.98 P ≤ 0.05**	**r = 0.95 P ≤ 0.05**	-	**r = 0.96 P ≤ 0.05**	**r = 0.92 P ≤ 0.05**	**r = 0.44 P ≤ 0.05**	**r = 0.87 P ≤ 0.05**
16 joints (MCP 2-5, PIP 2-5)	**r = 0.73 P≤ 0.05**	**r =0.97 P ≤ 0.05**	-	**r = 0.95 P ≤ 0.05**	**r = 0.90 P ≤ 0.05**	**r = 0.93 P ≤ 0.05**	**r = 0.36 P ≤ 0.05**	**r = 0.87 P ≤ 0.05**
18 joints (wrists + MCP 2-5, PIP 2-5)	**r = 0.74 P ≤ 0.05**	-	**r =0.97 P ≤ 0.05**	**r = 0.98 P ≤ 0.05**	**r = 0.93 P=<0.05**	**r = 0.90 P ≤ 0.05**	**r = 0.48 P ≤ 0.05**	**r = 0. 85 P ≤ 0.05**
22 joints	-	**r = 0.74 P ≤ 0.05**	**r = 0.73 P≤ 0.05**	**r = 0.72 P ≤ 0.05**	**r = 0.69 P ≤ 0.05**	**r = 0.68 P≤ 0.05**	**r = 0.37 P ≤ 0.05**	**r = 0.64 P ≤ 0.05**

Suppl. Table 2: Permutation comparisons between pairs of US scoring related to their ability to detect the presence of PD signal (P<0.05 shows significant difference between scores).

Joints with PD	22 joints	18 joints (wrists, MCP 2-5, PIP 2-5)	16 joints (MCP 2-5, PIP 2-5)	14 joints (wrists, MCP 2-4, PIP 2-4)	10 joints (wrists, MCP 2-3, PIP 2-3)	8 joints (MCP 2-3, PIP 2-3	4 joints (wrists, MCP5)	4 joints (MCP 2-3 bilaterally)
4 joints (MCP 2-3)	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	-
4 joints (wrists + MCP5)	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	-	**P<0.05**
8 joints (MCP2-3, PIP 2-3)	**P<0.05**	**P<0.05**	0.056	0.09	0.38	-	**P<0.05**	**P<0.05**
10 joints (wrists + MCP 2-3, PIP 2-3)	**P<0.05**	0.059	0.09	0.14	-	0.38	**P<0.05**	**P<0.05**
14 joints (wrists + MCP 2-4, PIP 2-4)	**P<0.05**	0.28	0.35	-	0.14	0.09	**P<0.05**	**P<0.05**
16 joints (MCP 2-5, PIP 2-5)	**P<0.05**	0.42	-	0.35	0.09	0.056	**P<0.05**	**P<0.05**
18 joints (wrists + MCP 2-5, PIP 2-5)	**P<0.05**	-	0.421	0.28	0.059	**P<0.05**	**P<0.05**	**P<0.05**
22 joints	-	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**

Suppl. Table 3: Permutation comparisons between pairs of US scoring related to their ability to detect moderate-severe SH (P<0.05 shows significant correlations).

Assessment of moderate-severe SH	22 joints	18 joints (wrists, MCP 2-5, PIP 2-5)	16 joints (MCP 2-5, PIP 2-5)	14 joints (wrists, MCP 2-4, PIP 2-4)	10 joints (wrists, MCP 2-3, PIP 2-3)	8 joints (MCP2-3, PIP 2-3)	4 joints (wrists, MCP5)	4 joints (MCP 2-3 bilaterally)
4 joints (MCP 2-3 bilaterally)	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	-
4 joints (wrists + MCP5)	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	-	**P<0.05**
8 joints (MCP2-3, PIP 2-3)	**P<0.05**	**P<0.05**	**P<0.05**	P=0.077	P=0.91	-	**P<0.05**	**P<0.05**
10 joints (wrists + MCP 2-3, PIP 2-3)	**P<0.05**	P=0.077	**P<0.05**	P=0.096	-	P=0.91	**P<0.05**	**P<0.05**
14 joints (wrists + MCP 2-4, PIP 2-4)	**P<0.05**	0.09	0.105	-	0.0967	0.077	**P<0.05**	**P<0.05**
16 joints (MCP 2-5, PIP 2-5)	**P<0.05**	0.945	-	0.105	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**
18 joints (wrists + MCP 2-5, PIP 2-5)	**P<0.05**	-	0.945	0.0902	0.077	**P<0.05**	**P<0.05**	**P<0.05**
22 joints	-	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**	**P<0.05**

Formatted: Line spacing: Double

Diagnostic accuracy of simplified ultrasound hand examination protocols for detection of inflammation and disease burden in patients with rheumatoid arthritis

Formatted: Font: 16 pt

Abstract:

Formatted: Font: 14 pt

Background: There is no consensus regarding the minimum of joints that should be included in an ultrasound (US) scoring system to reliably assess for disease activity in rheumatoid arthritis (RA).

Purpose: To assess whether simplified US protocols for hand examination are as informative as the examination of 22 joints in patients with RA, and to correlate the US parameters with disease activity (DAS-28).

Material and Methods: This is a cross-sectional study of 224 RA patients stratified based on their DAS-28 scores and assessed using eight preselected US examination protocols, including 22, 18, 16, 14, 10, 8 and two different combinations of 4 joints, respectively.

Results: We found a significant difference between different US hand scores regarding their ability to detect active inflammation and erosions. DAS-28 scores correlated very well with the Power Doppler (PD) scores generated by all eight US examination protocols (r=0.89-1,

P<0.05), irrespective of patients' disease activity. Simplified US scores missed information on presence of PD in 20.6 - 40.2% patients (P<0.05), and misdiagnosed non-erosive hand RA in 12 - 38.4% patients (P<0.05), depending on the number of joints excluded from US hand examination.

Conclusions: Preselected simplified US scores are less reliable in appreciating the disease burden when compared with an extended protocol for 22 joint US examination, raising clinicians' awareness regarding the need to comprehensively assess multiple hand joints to reliably rule out subclinical inflammation.

Keywords: hand, ultrasound, Power Doppler ultrasound.

Introduction

Formatted: Font: 14 pt

RA is a chronic inflammatory condition associated with well-recognised inflammatory joint features, which are amenable to US examination. The use of US facilitated a significant progress in the early diagnosis of RA, enabling a better assessment of the disease activity, prognosis and response to different therapeutic interventions. The implementation of US scoring systems in addition to clinical examination could help standardising the way RA is monitored; however, based on local availability of US and sonographer expertise, different scoring systems have been used in clinical practice. Despite significant research progress in supporting the role of US in RA, no consensus was reached with regard to what scoring system is the most useful. The OMERACT US Task Force defined the US pathology associated with RA (1), which combines tendon, joint and bone abnormalities (1, 2). The presence of Power Doppler (PD) is recognized as a reliable objective measure of active joint inflammation (3). Different semi-quantitative scoring systems are currently used for

assessing synovial hypertrophy (SH), joint effusion, tendon abnormalities and erosions (4), and protocols for hand and feet US examination are well-established (5).

A recent systematic review of the scoring systems used to evaluate synovitis in RA found difficult to determine the least number of joints that needed to be assessed for a global US score (1). The purpose of our study was to investigate how much we can simplify the US examination of hands in RA, without compromising the ability of a certain US scoring system to evaluate the disease activity and damage associated with hand RA. The authors focused on the US examination of hands as this is the most commonly used in routine clinical practice.

Material and M~~m~~ethods~~:~~

Formatted: Font: 14 pt

This is a real-life, cross-sectional study, which evaluated patients referred to our US rheumatology outpatient clinics, presenting with inflammatory sounding hand joint pains. The patients were referred based on clinician indication to have an US scan to help with identifying joint inflammation that was not confidently assessed clinically. We examined 604 patients between Jan 2012 and August 2015. For each patient, a set of demographic, clinical and laboratory data were recorded at the time of the scan. Of 604 patients referred to our clinic, 224 patients with RA were included in the study analysis based on their final diagnosis made using the 2010 ACR/EULAR classification criteria, following complete investigations and revision of the clinical notes. Fig~~ure~~ 1 details the patient selection and stratification based on DAS-28 scores.

This study evaluated the same set of reported outcomes and clinical and laboratory parameters for all the patients, to ensure homogeneity of the collected data. The following information was ~~analysed~~analyzed: disease duration (in months), hand tender joint count

(TJC) and swollen joint count (SJC), as well as a patient reported global disease assessment score (GVAS).

Additional data about the high sensitivity C-reactive protein (hsCRP), erythrocyte sedimentation rate (ESR), presence of rheumatoid factor (RF), anti citrullinated cyclic peptides antibodies (ACPA) and anti-nuclear antibodies (ANA) were collected at the time of the scan (needed to exclude associated pathology).

For each patient, a detailed record was compiled of their medication at the time of the US scan, including paracetamol and NSAIDs, disease-modifying drugs (DMARDs), biologic therapies and glucocorticoids, either oral or intramuscular depot injection.

The US protocol examination used included the extensor tendons and 22 joint assessments (dorsal longitudinal and transverse views of wrists, including extensor tendons, metacarpo-phalangeal – MCP joints, and proximal interphalangeal – PIP joints), as per our local clinic protocol. The same US examination protocol was used for each patient, irrespective of their hand symptoms. The US findings were scored according to the OMERACT scoring system (1). The hand US examination was performed by two clinicians (CC and LA) in the same session, and for each patient a consensus was obtained.

We used a Logiq S8 US machine (GE Healthcare, ~~Medical Systems US and Primary Care Diagnostics,~~ Wauwatosa, Wisconsin, WI, USA), equipped with a multi-frequency linear matrix array transducer (6-15 MHz). B-mode and PD machine setting were optimized and standardized for all our patients' US examinations. The settings used were: B-mode frequency 11-15 MHz depending on the depth of the anatomical area, Doppler frequency 7.5-15, depending on the depth of anatomical area; Doppler gain 18-20 dB, low wall filters and pulse repetition frequency around 800 Hz. In this study, we only used Power Doppler (PD) mode.

Comment [AS1]: Is Wauwatosa correct??

Comment [AS2]:

Comment [CC3]: Yes, this is correct.

The information collected comprised the following US parameters: SH grade (graded 1-3), erosions (present/absent), PD signal (graded 1-3), joint effusion (present/absent), osteophytes (present/absent), and tendon abnormalities (PD signal present/absent) using the US definition of joint pathology as defined by the OMERACT group (2) (Fig~~ure~~ 2 exemplifies two MCP joints with different SH and PD grades). Well controlled disease was defined as PD score zero (including joints and tendons).

To address our research question and assess how many joints would require scanning, and which joints are most likely to provide the answer as to whether or not there is active disease, we tested and compared the following scoring systems (bilateral examination):

- 22 joints (MCPs, PIPs, wrists)
- 18 joints (wrists, MCP 2-5 and PIP 2-5)
- 16 joints (MCP 2-5 and PIP 2-5)
- 14 joints (wrists, MCP 2-4 and PIP 2-4)
- 10 joints (wrists, MCP 2-3 and PIP 2-3)
- 8 joints (MCP 2-3 and PIP 2-3)
- 4 joints (wrists +MCP5)
- 4 joints (MCP 2-3)

The above joint combination score was selected based on our experience of performing US examination of hands in more than 1000 patients, which identified that the most affected joints in RA were the wrists, MCP 2,3 and 5, and PIP 2 and 3 (unpublished observation).

The SH grade 1 score was calculated as the total number of the joints with SH grade 1, the SH grade 2 score as the total number of the joints with SH grade 2, and the SH grade 3 score as the total number of the joints with SH grade 3 per patient. The total PD score was the sum

of all individual PD scores per patient, and the erosion score was calculated as the total number of erosions per patient.

Data about active inflammation affecting tendons overlying the above mentioned joints were also collected and reported separately. The total grey scale scores and PD scores for joints were calculated as a sum of the individual scores for all the joints included in the US examination protocol the score refers to. The duration of the US examination was approximately 25 minutes/patient. This 22 joint protocol is used routinely in our US clinics, which have 30 minute slots for clinical and US examination of patients with RA.

Descriptive statistics were used to characterize the RA population, and Student T test, Mann-Whitney U and Kuskal-Wallis tests were implemented for the assessment of different parameters and US scoring systems (IBM SPSS Statistics 22, IBM Corporation, 1 New Orchard Road, Armonk, New York 10504-1722, US). A P-value of <0.05 was considered a statistically significant result. Spearman's correlation coefficients were used to correlate permutations of pairs of US scores and the total PD scores with the disease activity, as assessed by the disease activity score assessing 28 joints (DAS-28).

The data were collected as standard of practice. The study ~~analysed~~analyzed cross-sectionally the results of the US examinations of patients seen in our US clinics over a defined period of time. No ethical approval or patient's consent were required as no patient information was used for teaching or new intervention research. The results of our study analysis had no impact on the clinical management of patients and their confidentiality was maintained.

Results~~:~~

Comment [AS4]: All kind of commercial equipment, please give full name of the product, the producer and its location (city, state (if in the US) and country).

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: 14 pt

To ~~characterise~~characterize in detail our RA cohort, we stratified patients based on DAS-28 (ESR) assessment of disease activity (Table 1). Demographic parameters were similar among different disease activity groups. As expected, patients with higher disease activity scores had significantly higher TJC, SJC, ESR and GVAS, while the CRP levels were similar between different groups. Both objective and subjective parameters included in the DAS-28 composite score were significantly increased in patients with active disease compared to moderate or low disease activity groups and with the group in remission.

There were no significant differences in the total US scores including the majority of US parameters, or in the disease duration or type of medication used (for both conventional and biologic DMARDs). The only significant difference was between the proportion of patients with SH grade 2 at the US examination of their hands, which was higher in patients with moderately-active and strongly-active RA ($P<0.05$) (Table 1). The SH grade 2 total score also correlated with the SJC ($r=0.89$, $P<0.05$).

The comparative analyse of the above-mentioned US scores showed no significant differences between the ability of the pre-selected US scores to capture information regarding SH grade 2 and 3, and the total PD scores per patient; however, the proportion of patients with no active disease at the US examination differed significantly based on the number of joints included in the examination protocol ($P<0.05$) (Table 2). Similarly, different US scores varied significantly in their ability to assess the total erosion score per patient and the proportion of patients with erosions ($P<0.05$). By simplifying the US examination of the hand in RA patients, active RA was underdiagnosed in a proportion of 20.6 to 40.2% of patients; similarly, the erosive burden was underappreciated in 12 - 28.4% RA patients (Table 2).

Strong correlations were found between the PD score generated by the 22 joint examination and all of the other US score combinations ($r = 0.68$ - 0.74, $P<0.05$). The scores that

correlated very strongly were those assessing 8, 10 and 14 joints (r = 0.92-0.96, P<0.05). The weakest correlation was found between the 8 and the 4 joint score (wrist and MCP 5 bilaterally) (r=0.28, P<0.05) (Suppl. Table 1).

The permutation comparisons between pairs of US scores related to their ability to detect the presence active joint inflammation found no significant differences between the total PD scores assessed by 8, 10, 12 and 16 joint US scores and 10, 12, 16 and 18 joint scores, respectively (Suppl. Table 2). Similarly, the total grey scale score (combining the total scores for SH grade 2 and 3) identified no significant differences between the permutation comparisons between the scores assessing 8, 10, 14, 16 and 18 joints (Suppl. Table 3).

The analysis was also focused on correlating the total PD scores derived from all the pre-set US examination protocols with the DAS-28 scores in patients stratified based on their disease activity, to identify if certain US hand examination protocols can be used differentially in patients with active disease compared to patients in remission. All the total PD scores derived from the eight US examination protocols correlated very strongly with DAS-28 assessment, irrespective of how well the disease was controlled (r = 0.88-1, P<0.005).

Discussion

Formatted: Font: 14 pt

~~In conclusion,~~ T~~t~~his is the first large cross-sectional study correlating different US examination protocols (derived from a 22-hand joint comprehensive score) with DAS-28 score in patients in RA, stratified based on their disease activity.

Quantitative and semi-quantitative US scores have been previously compared in RA (3), and US examination have been found to be sensitive to therapeutic interventions (4-8). A comprehensive study comparing several US score systems in RA found that all were sensitive

to change when assessing the response of RA patients to adalimumab (9, 10). In addition, simplified US scores (including 6 or 12 joints) have previously been compared with extensive US protocol examinations (assessing 12, and 44 joints respectively), and showed good sensitivity to change in three separate studies (11-13). However, none of these studies stratified patients based on their disease activity scores or included RA patients based on the clinical indication to have an US scan, as it is the case with our study. The need to use a comprehensive US scoring system, capturing both active and chronic inflammatory changes for assessment of RA disease activity, is supported by the good correlation between US and MRI findings (8, 14). The presence of SH and PD signal was found to be associated with structural damage in RA (15), even in patients in clinical remission (16), and was associated with risk of flares (17, 18). ~~There was a good correlation between US findings and clinical examination in one study examining 60 joints/patient (19); however, there are obvious limitations to implement in practice such a comprehensive US protocol.~~

The role and reliability of US in the disease activity assessment in patients with RA is supported by several studies (19-21). ~~It was previously proposed that a targeted US remission in early RA would inform clinicians better about the need of disease control optimisation~~optimization ~~compared to clinical assessment; however, this was not associated with long-term benefits in a recent randomised~~randomized ~~controlled trial (22).~~

Previous studies reported good correlation between hand US scores and DAS-28 assessment using three different US scores (23, 24), result that was also replicated by our study, which included a larger number of joint combinations, and also assessed US parameters stratified based on the DAS-28 scores.

Our comparative analysis of several US scoring systems showed that there is significant difference in terms of the equivalence of several US hand-scoring systems. ~~that can be used~~

~~in routine practice.~~ Our study found that age, duration of symptoms, duration of disease, type of medication and total PD score generated by US examination of hands were not able to inform about the inclusion of patients in one specific disease activity group, as patients stratified based on DAS 28 scores had similar parameters. ~~Only the proportion of patients with joints with SH grade 2 was different across different disease activity groups; however, this finding was no replicated in the case of severe SH (grade 3), disparity that can be explained by the lower number of patients with SH grade 3 included in our study.~~

In addition to previous studies, we have been interested in exploring the amount and significance of missed information related to the use of simplified US hand examination protocols. A significant proportion of patients have been diagnosed as having well-controlled or non-erosive hand RA ~~in our study~~ by using US protocols limiting the number of joints examined (10.6 - 40.2 % and 12-34.8%, respectively). Our study found that the assessment of our preselected 8, 10 and 14 joints captured comparable amounts of information regarding disease activity in RA (still misdiagnosing around 40% of patients as having well controlled disease, equivalent to PD score zero), while the two 4 joint scores missed significant information when compared to the others (around 60% patient were diagnosed in remission despite having active disease at least in one joint). The scores including 20 and 22 joints captured more information than the 8, 10 and 14 joint scores, even if all the eight US scores we explored correlated very well with the DAS-28 assessment. This is particularly relevant for our patient group, characterized by a small number of active joints and clinical indication to have an US scan to establish if their disease was well controlled or not. In this context, underdiagnosing active disease would have erroneously led to classifying our patients as being in remission. The clinical consensus is that we cannot predict which joints are the most likely to flare in patients with RA patients; therefore examining only the joints that previously flared using a patient-tailored US protocol is not justified.

~~In conclusion,~~ E~~e~~ven if a comprehensive hand joint score is time-consuming, it can provide significant additional information compared to a simplified score, as our study showed. As expected, all the scores correlated very well with each other, because they are derived from a comprehensive US hand score, while missing significant information proportional to the number of joints excluded from US examination. All the pre-selected US hand scores correlated with the disease activity scores, despite the fact that the patient groups stratified based on disease activity had similar median total PD scores. This showed that subclinical in inflammation can be find in similar proportion in RA patients, irrespective of their degree of chronic joint changes that are likely to influence their DAS-28 score.

Limitations: Our study did not have strict inclusion criteria: the patients were included based on clinical indication to exclude subclinical synovitis. Therefore, there is a significant selection bias, as the study did not capture patients with obvious active synovitis detected by clinical examination. In this particular clinical context, detection of active disease in at least one joint is clinically relevant, as US examination triggered treatment optimization to minimize joint damage (e.g. guided steroid injection targeting the active joints or escalation of therapy).

In conclusion, ~~Our study concluded that~~ even if simplified US scores for hand assessment of RA disease activity can be useful in practice, by examining additional joints, clinicians are able to detect subclinical inflammation, which is not captured by the simplified US scores. If previously studies re-assured clinicians that various US examination protocols correlated well with the DAS-28 assessment or were sensitive to change following therapy, our study showed that a significant proportion of patients can be misclassified as having well-controlled or non-erosive disease as a result of simplified US protocols. Further studies, including large longitudinal cohorts, are needed to establish the smaller number of joints needed to be

examined to ~~minimise~~minimize the risk of under detecting subclinical inflammation in patients with hand RA.

Conflict of interest:

The authors declared no conflicts of interest.

~~**Funding**~~

~~This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. CC was funded by a Biomedical Research Council Funding grant – BRCIII/001.~~

Comment [C5]: The details about funding were transferred to the title page, as required.

1. Mandl P, Naredo E, Wakefield RJ, et al. 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;38:2055-2062.
2. Wakefield RJ, Balint PV, Szkudlarek M, et al. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. J Rheumatol. 2005;32:2485-2487.
3. Ellegaard K, Terslev L, Christensen R, et al. Comparison of discrimination and prognostic value of two US Doppler scoring systems in rheumatoid arthritis patients: a prospective cohort study. Clin Exp Rheumatol. 2014;32:495-500.
4. Epis O, Scioscia C, Locaputo A, et al. Use of ultrasound in treatment decisions for patients with rheumatoid arthritis: an observational study in Italy. Clin Rheumatol. 2016;35:1923-1929.
5. Hau M, Kneitz C, Tony HP, et al. 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;61:55-58.

6. Terslev L, Torp-Pedersen S, Qvistgaard E, et al. Estimation of inflammation by Doppler ultrasound: quantitative changes after intra-articular treatment in rheumatoid arthritis. Ann Rheum Dis. 2003;62:1049-1053.
7. Epis O, Filippucci E, Delle Sedie A, et al. Clinical and ultrasound evaluation of the response to tocilizumab treatment in patients with rheumatoid arthritis: a case series. Rheumatol Int. 2014;34:737-742.
8. Rowbotham EL, Grainger AJ. Rheumatoid arthritis: ultrasound versus MRI. Am J Roentgenol. 2011;197:541-546.
9. Hammer HB, Kvien TK. Comparisons of 7- to 78-joint ultrasonography scores: all different joint combinations show equal response to adalimumab treatment in patients with rheumatoid arthritis. Arthritis Res Ther. 2011;13:R78.
10. Filippucci E, Iagnocco A, Salaffi F, et al. Power Doppler sonography monitoring of synovial perfusion at the wrist joints in patients with rheumatoid arthritis treated with adalimumab. Ann Rheum Dis. 2006;65:1433-1437.
11. Perricone C, Ceccarelli F, Modesti M, et al. The 6-joint ultrasonographic assessment: a valid, sensitive-to-change and feasible method for evaluating joint inflammation in RA. Rheumatology (Oxford). 2012;51:866-73.
12. Naredo E, Valor L, De la Torre I, et al. Ultrasound joint inflammation in rheumatoid arthritis in clinical remission: how many and which joints should be assessed? Arthritis Care Res (Hoboken). 2013;65:512-7.
13. Naredo E, Rodriguez M, Campos C, 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.

14. Schmidt WA, Schicke B, Ostendorf B, et al. Low-field MRI versus ultrasound: which is more sensitive in detecting inflammation and bone damage in MCP and MTP joints in mild or moderate rheumatoid arthritis? Clin Exp Rheumatol. 2013;31:91-96.

15. Dougados M, Devauchelle-Pensec V, Ferlet JF, et al. The ability of synovitis to predict structural damage in rheumatoid arthritis: a comparative study between clinical examination and ultrasound. Ann Rheum Dis. 2013;72:665-671.

16. Brown AK, Quinn MA, Karim Z, et al. Presence of significant synovitis in rheumatoid arthritis patients with disease-modifying antirheumatic drug-induced clinical remission: evidence from an imaging study may explain structural progression. Arthritis Rheum. 2006;54:3761-3773.

17. Nguyen H, Ruyssen-Witrand A, Gandjbakhch F, et al. Prevalence of ultrasound-detected residual synovitis and risk of relapse and structural progression in rheumatoid arthritis patients in clinical remission: a systematic review and meta-analysis. Rheumatology (Oxford). 2014;53:2110-2118.

18. Bellis E, Scire CA, Carrara G, et al. Ultrasound-detected tenosynovitis independently associates with patient-reported flare in patients with rheumatoid arthritis in clinical remission: results from the observational study STARTER of the Italian Society for Rheumatology. Rheumatology (Oxford). 2016;55:1826-1836.

19. Naredo E, Bonilla G, Gamero F, et al. Assessment of inflammatory activity in rheumatoid arthritis: a comparative study of clinical evaluation with grey scale and power Doppler ultrasonography. Ann Rheum Dis. 2005;64:375-381.

20. Mandl P, Balint PV, Brault Y, et al. Metrologic properties of ultrasound versus clinical evaluation of synovitis in rheumatoid arthritis: results of a multicenter, randomized study. Arthritis Rheum. 2012;64:1272-1282.

21. Szkudlarek M, Terslev L, Wakefield RJ, et al. Summary Findings of a Systematic Literature Review of the Ultrasound Assessment of Bone Erosions in Rheumatoid Arthritis. J Rheumatol. 2016;43:12-21.

22. Dougados M, Jousse-Joulin S, Mistretta F, et al. Evaluation of several ultrasonography scoring systems for synovitis and comparison to clinical examination: results from a prospective multicentre study of rheumatoid arthritis. Ann Rheum Dis. 2010;69:828-833.

23. Rosa J, Ruta S, Saucedo C, et al. Does a Simplified 6-Joint Ultrasound Index Correlate Well Enough With the 28-Joint Disease Activity Score to Be Used in Clinical Practice? J Clin Rheumatol. 2016;22:179-183.

Figure legends:

Figure 1. Flowchart of the study population.

Figure 2: Examples of MCP joint grading: SH grade 3 and PD grade 2 (above), and SH grade 2 and PD grade 3 (below)

~~**References:**~~

~~1. Mandl P, Naredo E, Wakefield RJ, Conaghan PG, D'Agostino MA, Force OUT. 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(9):2055-62.~~

~~2. 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(12):2485-7.~~

~~3. Ellegaard K, Terslev L, Christensen R, Szkudlarek M, Schmidt WA, Jensen PS, et al. Comparison of discrimination and prognostic value of two US Doppler scoring systems in rheumatoid arthritis patients: a prospective cohort study. Clin Exp Rheumatol. 2014;32(4):495-500.~~

Formatted: Indent: Left: 0"

Formatted: Font: Times New Roman, 12 pt

Formatted: Line spacing: Double

Formatted: Font: 14 pt

Formatted: Line spacing: Double

4. Epis O, Scioscia C, Locaputo A, Cappelli A, Maier A, Rocchetta PA, et al. Use of ultrasound in treatment decisions for patients with rheumatoid arthritis: an observational study in Italy. Clin Rheumatol. 2016;35(8):1923-9.
5. 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(1):55-8.
6. Terslev L, Torp-Pedersen S, Qvistgaard E, Danneskiold-Samsoe B, Bliddal H. Estimation of inflammation by Doppler ultrasound: quantitative changes after intra-articular treatment in rheumatoid arthritis. Ann Rheum Dis. 2003;62(11):1049-53.
7. Epis O, Filippucci E, Delle Sedie A, De Matthaeis A, Bruschi E. Clinical and ultrasound evaluation of the response to tocilizumab treatment in patients with rheumatoid arthritis: a case series. Rheumatol Int. 2014;34(5):737-42.
8. Rowbotham EL, Grainger AJ. Rheumatoid arthritis: ultrasound versus MRI. AJR Am J Roentgenol. 2011;197(3):541-6.
9. Hammer HB, Kvien TK. Comparisons of 7- to 78-joint ultrasonography scores: all different joint combinations show equal response to adalimumab treatment in patients with rheumatoid arthritis. Arthritis Res Ther. 2011;13(3):R78.
10. 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(11):1433-7.
11. Perricone C, Ceccarelli F, Modesti M, Vavala C, Di Franco M, Valesini G, et al. The 6-joint ultrasonographic assessment: a valid, sensitive to change and feasible method for evaluating joint inflammation in RA. Rheumatology (Oxford). 2012;51(5):866-73.
12. Naredo E, Valor L, De la Torre I, Martinez-Barrio J, Hinojosa M, Aramburu F, et al. Ultrasound joint inflammation in rheumatoid arthritis in clinical remission: how many and which joints should be assessed? Arthritis Care Res (Hoboken). 2013;65(4):512-7.
13. Naredo E, Rodriguez M, Campos C, Rodriguez-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(4):515-22.
14. Schmidt WA, Schicke B, Ostendorf B, Scherer A, Krause A, Walther M. Low-field MRI versus ultrasound: which is more sensitive in detecting inflammation and bone damage in MCP and MTP joints in mild or moderate rheumatoid arthritis? Clin Exp Rheumatol. 2013;31(1):91-6.
15. Dougados M, Devauchelle-Pensec V, Ferlet JF, Jousse-Joulin S, D'Agostino MA, Backhaus M, et al. The ability of synovitis to predict structural damage in rheumatoid arthritis: a comparative study between clinical examination and ultrasound. Ann Rheum Dis. 2013;72(5):665-71.
16. Brown AK, Quinn MA, Karim Z, Conaghan PG, Peterfy CG, Hensor E, et al. Presence of significant synovitis in rheumatoid arthritis patients with disease-modifying antirheumatic drug-induced clinical remission: evidence from an imaging study may explain structural progression. Arthritis Rheum. 2006;54(12):3761-73.
17. Nguyen H, Ruyssen-Witrand A, Gandjbakhch F, Constantin A, Foltz V, Cantagrel A. Prevalence of ultrasound-detected residual synovitis and risk of relapse and structural progression in rheumatoid arthritis patients in clinical remission: a systematic review and meta-analysis. Rheumatology (Oxford). 2014;53(11):2110-8.
18. Bellis E, Scire CA, Carrara G, Adinolfi A, Batticciotto A, Bortoluzzi A, et al. Ultrasound-detected tenosynovitis independently associates with patient-reported flare in patients with rheumatoid arthritis in clinical remission: results from the observational study STARTER of the Italian Society for Rheumatology. Rheumatology (Oxford). 2016;55(10):1826-36.

19. 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(3):375-81.
20. Mandl P, Balint PV, Brault Y, Backhaus M, D'Agostino MA, Grassi W, et al. Metrologic properties of ultrasound versus clinical evaluation of synovitis in rheumatoid arthritis: results of a multicenter, randomized study. Arthritis Rheum. 2012;64(4):1272-82.
21. Szkudlarek M, Terslev L, Wakefield RJ, Backhaus M, Balint PV, Bruyn GA, et al. Summary Findings of a Systematic Literature Review of the Ultrasound Assessment of Bone Erosions in Rheumatoid Arthritis. J Rheumatol. 2016;43(1):12-21.
22. Dale J, Stirling A, Zhang R, Purves D, Foley J, Sambrook M, et al. Targeting ultrasound remission in early rheumatoid arthritis: the results of the TaSER study, a randomised clinical trial. Ann Rheum Dis. 2016;75(6):1043-50.
23. Dougados M, Jousse Joulin S, Mistretta F, d'Agostino MA, Backhaus M, Bentin J, et al. Evaluation of several ultrasonography scoring systems for synovitis and comparison to clinical examination: results from a prospective multicentre study of rheumatoid arthritis. Ann Rheum Dis. 2010;69(5):828-33.
24. Rosa J, Ruta S, Saucedo C, Navarta DA, Catoggio LJ, Garcia-Monaco R, et al. Does a Simplified 6-Joint Ultrasound Index Correlate Well Enough With the 28-Joint Disease Activity Score to Be Used in Clinical Practice? J Clin Rheumatol. 2016;22(4):179-83.

Formatted: Line spacing: Double

Table 1- Comparison between RA patient groups stratified based on their DAS 28 scores using the 22 joint US scoring system as detailed above (Kruskal-Wallis test, $p<0.05$ shows a significant difference between the patient groups).

RA patients stratified based on disease activity	DAS28 >5.1	DAS 28 3.2-5.1	DAS 28 2.6-3.2	DAS 28 <2.6	P value
Age Mean ± SD	55.6 ± 13.8	53.2 ± 16.	54.2 ± 15.5	50.3 ± 15.3	P=0.44
% Female	80.0	89.5	71.4	75.6	P=0.11
Disease duration (months): Mean ± SD	120.3 ± 107	111.7± 135	70.5 ± 58.4	95.4 ± 169.9	P=0.51
% of patients on steroids at the time of the scan (all patients were on ≤ 10 mg daily)	38	43.3	29.0	36.8	P=0.36
% of patients on conventional DMARDs at the time of the scan	74	54.2	64.5	65.8	P=0.11

% of patients on biologic treatment at the time of the scan	24	19.3	29.0	26.3	P=0.67
CRP Mean ± SD	8.2 ± 10.8	6.1 ± 14.8	4.3 ± 7.9	4.3 ± 9.6	P=0.42
ESR Mean ± SD	31.1 ± 26.7	16.4 ± 15.1	11.5 ± 15.5	6.8 ± 6.7	**P<0.05**
SJC Mean ± SD	5.3 ± 4.6	2.4 ± 2.5	1.5 ± 1.9	0.5 ± 0.8	**P<0.05**
TJC Mean ± SD	17.1 ± 7.5	6.7 ± 5.2	4.5 ± 4.6	1.3 ± 2.0	**P<0.05**
GVAS Mean ± SD	74.6 ± 19.8	48.4 ± 26.4	34.4 ± 23.3	25 ± 26.2	**P<0.05**
Mean DAS 28 score ± SD	5.9 ± 0.75	3.9 ± 0.5	2.9 ± 0.16	1.7 ± 0.6	**P<0.05**
Total number of joints with SH grade 1 / patient Mean ± SD	2.4 ± 3.1	2.6 ± 3.7	1.7 ± 2.7	1.2 ± 1.8	P=0.52
Percentage of patients with joints with SH grade 1:	58.0	66.3	51.6	47.4	P=0.23
Total number of joints with SH grade 2 / patient Mean ± SD	2.4 ± 3.4	2.2 ± 3.2	1.6 ± 4.4	1.3 ± 2.5	P=0.36
Percentage of patients with joints with SH grade 2:	58.0	53.0	48.4	29.0	**P<0.05**
Total number of joints with SH grade 3 / patient Mean ± SD	2.1 ± 2.8	1.2 ± 1.7	1.1 ± 2.1	0.7 ± 1.3	P=0.49
Percentage of patients with joints with SH grade 3:	52.0	44.6	35.5	31.6	P=0.23
PD score Mean ± SD	1.9 ± 2.9	1.2 ± 2.4	1.1 ± 1.5	0.7 ± 1.3	P=0.58

Percentage of patients with PD signal	58.0	42.2	45.2	36.8	P=0.19
Total number of joints with erosions / patient Mean ± SD	6.5 ± 7.3	5.0 ± 5.5	4.3 ± 3.8	2.7 ± 4.1	P=0.17
Percentage of patients with erosions:	64.0	49.4	64.5	39.5	P=0.09
Percentage of patients with tendon abnormalities (GS score ≥ 2)	8.16	10.4	13.3	10.81	P=0.42
Percentage of patients with active tenosynovitis (PD score ≥ 1)	6.12	5.81	8.13	10.81	P=0.13

Table 2: Comparison between 8 different US scores (P<0.05 was considered significant).

US findings/ scores	22 joints (wrists, MCPs, PIPs)	18 joints (wrists, MCP 2-5, PIP 2-5)	16 joints (MCP 2-5, PIP 2-5)	14 joints (wrists, MCP 2-4, PIP 2-4)	10 joints (wrists, MCP2-3, PIP 2-3)	8 joints (MCP2-3, PIP 2-3)	4 joints (wrists, MCP5)	4 joints (MCP 2-3 bilaterally)	P value
SH grade 2 score /patient	Median: 0 IQR: 3	Median: 0 IQR:1	Median: 0 IQR:1	Median: 0 IQR: 1	Median: 0 IQR: 1	Median: 0 IQR: 0	Median: 0 IQR: 0	Median: 0 IQR:1	-
Mean SH grade 2 score ± SD:	2 ± 3.21	1.19 ± 2.32	0.9 ± 1.83	0.71 ± 1.4	0.7 ± 1.39	0.15 ± 0.4	0.43 ±0.92	1.19 ±2.32	0.43
Percentage of patients with no evidence of SH grade 2:	51.8	64.3	64.3	66.5	69.6	69.6	89.3	76.8	0.15
SH grade 3 score patient	Median: 0 IQR: 2	Median: 0 IQR:1	Median: 0 IQR:1	Median: 0 IQR:1	Median: 0 IQR: 1	Median: 0 IQR: 1	Median: 0 IQR: 0	Median: 0 IQR: 0	-
Mean SH grade 3 score ± SD:	1.3 ± 2.08	0.71 ± 1.41	0.7 ± 1.40	0.58 ±1.17	0.5 ± 0.97	0.5 ± 0.95	0.09 ±0.34	0.35±0.74	0.15
Percentage of patients with no evidence of SH grade 3:	57.1	69.2	69.2	71.4	74.1	74.1	93.3	77.7	0.15
PD score/patient	Median: 0 IQR: 2	Median: 0 IQR:1	Median: 0 IQR:1	Median: 0 IQR: 0	Median: 0 IQR: 0	Median: 0 IQR: 0	Median: 0 IQR: 0	Median: 0 IQR: 0	-
Mean PD score ± SD:	1.28 ± 2.31	0.69 ± 1.68	0.66 ± 1.66	0.61 ± 1.39	0.48 ±1.1	0.45 ± 1.07	0.09 ±0.36	0.32 ± 0.76	0.15
Percentage of patients with well controlled hand RA (PD score = 0)	54.0	74.6	75.9	75.4	77.7	79.0	94.2	81.3	**< 0.05**
Percentage of patients misdiagnosed with well controlled disease by the simplified US scores	N/A	20.6	21.9	21.4	23.7	25	40.2	27.3	**<0.05**
Erosion score/patient	Median: 2 IQR: 7.75	Median: 1 IQR: 4	Median: 1 IQR: 4	Median: 1 IQR:4	Median: 1 IQR: 3	Median IQR:3	Median: 1 IQR:1	Median: 1 IQR:2	**< 0.05**

Formatted Table

Percentage of patients with non-erosive hand RA:	30.4	42.4	42.4	46.9	46.9	46.9	68.8	55.4	< 0.05
Percentage of patients misdiagnosed with non-erosive hand RA by the simplified US scores	N/A	12	12	16.5	16.5	16.5	38.4	25	< 0.05

Formatted: Line spacing: Double

Formatted: Font: Times New Roman, 12 pt