

RESEARCH

Open Access



# Correlation and agreement between physical and ultrasound examination after a training session dedicated to the standardization of synovitis assessment in rheumatoid arthritis patients

Gerardo Quintana-López<sup>1,2,3\*</sup> , Kevin Maldonado-Cañón<sup>1</sup> , Jorge Bruce Flórez-Suárez<sup>1</sup> , Paul Méndez-Patarroyo<sup>1,3</sup>, Paola Coral-Alvarado<sup>1,3</sup> and Enrique Calvo<sup>4</sup>

## Abstract

**Objectives:** Assessing disease activity in rheumatoid arthritis (RA) patients requires comprehensive quantification of tender and swollen joints. We aimed to evaluate the correlation and agreement between rheumatologists after a training session dedicated to the standardization of synovitis assessment and compare its performance with a reference imaging modality such as musculoskeletal ultrasonography (MSUS).

**Methods:** In this cross-sectional study, a total of 28 and 10 joints in RA patients were evaluated by physical examination and ultrasound (US), respectively. After participating in a training session, individual joint assessment for tenderness and swelling was performed by three rheumatologists. MSUS examination was performed separately by an experimented radiologist in a standardized manner, evaluating findings according to the Outcome Measures in Rheumatology Clinical Trial (OMERACT) guidelines.

**Results:** A total of 80 RA patients were included, with a mean Disease Activity Score based on 28 joints (DAS28)-ESR of 4.02. The interobserver overall agreement and concordance rate in a total of 2240 joints assessed was 81.7% ( $k = 0.449$ ,  $p < 0.0001$ ) for tender joints and 66% ( $k = 0.227$ ,  $p < 0.0001$ ) for swollen joints. The overall concordance rate was fair (Fleiss' kappa = 0.21,  $p = 0.027$ ) with an overall agreement of 67.18% yet, more joints were found to be swollen by the US assessment, compared to the physical examination (43% vs 39%).

**Conclusion:** In our study population, joint tenderness showed better interobserver agreement, correlation, and concordance rate than joint swelling. When comparing the US assessment to the physical examination, a fair overall concordance rate supports the need for the implementation of training sessions dedicated to standardization in rheumatology clinics.

**Keywords:** Physical examination, Rheumatoid arthritis, Synovitis, Ultrasonography

## Introduction

Rheumatoid arthritis (RA) is a chronic systemic autoimmune disease characterized by hyperplasia and inflammation of synovial tissue with subsequent bone erosion and loss of joint space with a noteworthy

\*Correspondence: [gquintanal@unal.edu.co](mailto:gquintanal@unal.edu.co)

<sup>1</sup> Reumavance Group, Rheumatology Section, Department of Internal Medicine, Fundación Santa Fe de Bogotá University Hospital, Carrera 7 No. 117-15, 110111 Bogotá, Colombia

Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

association on functionality and quality of life [1]. In order to assess disease activity in RA patients, there is not an individual parameter to be used. It is thus necessary to achieve a comprehensive approach through the use of several individual clinical and or laboratory parameters and to develop quantitative indexes to be used in daily clinical practice by rheumatology health professionals [2–4].

Among the most commonly used methods, the Disease Activity Score (DAS), and its modified version, DAS28, are based on a tender and swollen joint count, combined with other parameters such as a patient global health assessment [4, 5]. Joint tenderness assessments estimate the patient's response to potentially painful stimulation. On the other hand, joint swelling assessment measures synovial inflammation or effusion recognized by fluid displacement. The examination technique should include exerting continuous and direct pressure on an affected joint with the thumb and index fingers until the examiner's nail bed turns pale; this compares with a pressure of approximately 4 kg/cm<sup>2</sup> [6, 7].

Although tender and swollen joint counts are considered fundamental parameters to estimate disease activity, its assessment is not as straightforward as assumed, thus exhibiting certain potential disadvantages. Amongst the prevailing concerns are poor reproducibility and substantial interobserver variability [6, 8]. One of the reasonable explanations could be the training and clinical experience gap among practitioners; therefore, finding a feasible solution through standardization.

Historically, training sessions focused on standardization of all DAS28 parameters assessment have been proposed, and once applied, have shown to provoke a substantial reduction regarding the variation between examiners [6–8]. The entire medical and nursing staff who will be assessing RA patients should be trained regularly (at least once every year) and should be trained together, allowing discussion about current standard procedures. There has been proposed as well the use of ultrasonography to demonstrate active synovitis in case of disagreement [6].

Prognostic factors, treatment decisions, monitoring, and complications are defined based on disease activity, thus consequently on a reliable tender and swollen joint count for each RA patient [9]. The aim of this study was to evaluate the correlation and agreement between rheumatologists after a training session dedicated to the standardization of synovitis assessment and compare its performance with a reference imaging modality such as musculoskeletal ultrasonography (MSUS).

## Materials and methods

### Patients

Patients were recruited from an outpatient clinic based in the Rheumatology Department at Fundación Santa Fe de Bogotá University Hospital, Colombia. Stratified random sampling was conducted, selecting patients with different disease activity states (according to the most recent clinical record) from a previously established ongoing cohort of 820 patients. The sample size was calculated for the desired correlation coefficient of 0.6, a population correlation coefficient of 0.8, a power of 0.8, and a confidence interval of 0.95. All those who were invited to participate fulfilled the 2010 ACR/European League Against Rheumatism (EULAR) classification criteria [10] and were at least 18 years old. Those who had a history of trauma, septic arthritis, joint replacement or synovectomy, joint deformity, and or crystal arthropathy were excluded.

The following data were registered at baseline: age, gender, treatment, erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), rheumatoid factor (RF), global health assessed by the patient (GH), Clinical Disease Activity Index (CDAI), and Simplified Disease Activity Index (SDAI).

### Training session dedicated to standardization

Our planned training session was structured to focus on standardization of the tenderness and swelling assessment of 28 joints (bilateral shoulders, elbows, wrists, knees, metacarpophalangeal (MCP), and interphalangeal (PIP)).

Three rheumatologists, with 10, 13, and 15 years of experience in clinical examination of RA patients, attended three sessions (separated by 1–2 weeks). Each session was divided into three steps: (1) individual joint assessment was performed by each rheumatologist (blinded to both clinical and other rheumatologist's data), (2) 20-min discussion on practice observations, difficulties, limitations, and facilitating factors during the physical examination, in order to reach agreement on uniform criteria and technique, and (3) joint individual reassessment based on the unified criteria. On the third session, disease activity indexes (DAS28-ESR, SDAI, and CDAI) were calculated individually by each rheumatologist.

### Ultrasound assessment

Twenty minutes after the last training session's third step, each patient was instructed to proceed to the ultrasound (US) assessment room. The US examination was performed by a radiologist with 15 years of experience and training in musculoskeletal radiology (blinded to physical examinations' data) on 10 joints (bilateral wrists, and 2nd to 5th MCPs), using a GE (General Electric) LOGICQ E ultrasound machine with a 6–13 Hz multifrequency

linear transducer. Findings were systematically graded according to the Outcome Measures in Rheumatology Clinical Trial (OMERACT) guidelines, evaluating synovitis as synovial hypertrophy defined by an abnormal nondisplaceable and poorly compressible hypochoic intraarticular tissue [11]. PIPs joints were considered as potential confounders due to the eventual overlapping of tenosynovitis, therefore were not assessed.

Synovitis grading was conducted based on a scoring method initially introduced by Szkudlarek et al., widely used for studies of this kind [12–14] and currently supported by the EULAR-OMERACT ultrasound task-force [15, 16] (0=absence of synovial thickening, 1 (mild)=minimal synovial thickening, filling the angle between the periarticular bones, without bulging over the line linking tops of the corresponding bones, 2 (moderate)=synovial thickening bulging over the line linking tops of the periarticular bones but without extension along the bone diaphysis, 3 (severe)=severe synovial thickening bulging over the line linking tops of the periarticular bones and with extension to at least one bone diaphysis). Normal distances between bone and joint capsule were acknowledged based on average population values proposed by Schmidt et al. [17].

**Statistical analysis**

Statistical analysis was performed using STATA software, version 15.0. Descriptive analysis was presented for continuous variables with central tendency measures as mean and standard deviation (SD) or as the median and interquartile range (IQR) for normally or nonnormally distributed data, respectively. For dichotomous variables, data were presented with percentages and absolute values.

Interobserver agreement and concordance were calculated through Cohen’s kappa (between two observers considering all the possible pairs, i.e., Observer A vs B, observer A vs C, observer B vs C), Fleiss’ kappa (between the three observers), and percentage of an overall agreement (percentage of observed exact agreements). The relative strength of agreement was described according to the following ranges of kappa (k) coefficients: <0.00=poor, 0.00–0.20=slight, 0.21–0.40=fair, 0.41–0.60=moderate, 0.61–0.80=substantial, and 0.81–1.00=almost perfect [18]. Linear correlation coefficients (Pearson correlation coefficient) were also calculated for tender and swollen joint counts.

**Results**

**Patient characteristics**

The demographic and clinical characteristics of the 80 selected patients are illustrated in Table 1. The majority

**Table 1** Clinical and demographic characteristics for all patients

Characteristics	Values (n = 80)
Age, years	55.3 (8.6)
Female, n (%)	68 (85)
Disease duration (years)	11.56 (7.93)
Positive RF, n (%)	71 (88)
ESR (mm/h)	19.32 (5.6)
CRP (mg/dL)	5.6 (12)
DAS28-ESR	4.02 (1.12)
SDAI	24.4 (8.59)
CDAI	18.7 (9.01)
<i>Number of DMARDs</i>	
1, n (%)	15 (18%)
≥ 2, n (%)	58 (72%)
bDMARD use, n (%)	16 (20)
Oral glucocorticoid use, n (%)	27 (33)
TJC-28, median	9
SJC-28, median	5

Values are the mean (SD) unless otherwise indicated

RF, rheumatoid factor; ESR, erythrocyte sedimentation rate; CRP, C-reactive protein; DAS28-ESR, Disease Activity Score based on 28 joint count and ESR; SDAI, Simplified Disease Activity Index; CDAI, Clinical Disease Activity Index; DMARDs, disease-modified anti-rheumatic drug; bDMARD, biological disease-modified anti-rheumatic drug; TJC, tender joint count; SJC, swollen joint count

of them were women (85%), mean age was 55.3 years, with 71 (88%) being seropositive for RF. The mean (± SD) DAS28-ESR, SDAI, and CDAI were 4.02 ± 1.12 (moderate activity), 24.4 ± 8.59, and 18.7 ± 9.01, respectively. All our patients received disease-modifying anti-rheumatic drugs (DMARDs) treatment, corresponding to mono, ≥ 2 DMARD, and biological (b) DMARD therapy in 18%, 72%, and 20% of cases. Twenty-seven (33%) patients were receiving oral glucocorticoid therapy (daily dose of 5–10 mg (prednisone equivalent)).

**Interobserver correlation and agreement**

A total of 2240 joints (i.e., 28 joints in 80 patients) were evaluated through physical examination. Compared with tenderness, swelling assessment showed a slight strength of agreement with lower percentage of overall agreement and concordance rates (66% and Fleiss kappa = 0.227; 95% CI 0.203–0.250, *p* < 0.0001 vs 81.7% and Fleiss kappa = 0.449; 95% CI 0.403–0.493, *p* < 0.0001) (see Table 2).

The linear correlation coefficient (Spearman’s correlation coefficient) was calculated, taking all the possible pairs (observer A–B, observer A–C, observer B–C), thus having an overall correlation coefficient of 0.81 and 0.431 for swollen and tender joints, respectively.

**Table 2** Interobserver agreement and concordance for physical examination (PE) tenderness and swelling assessment

Joint	Inter-observer agreement and concordance							
	PE tenderness assessment				PE swelling assessment			
	Fleiss kappa	95% CI	p value	Count	Fleiss kappa	95% CI	p value	Count
Right shoulder	0.521	0.494–0.652	<0.0001	83	0.249	0.124–0.322	0.0001	27
Left shoulder	0.446	0.198–0.546	<0.0001	64	0.134	0.092–0.184	0.018	23
Right elbow	0.4	0.308–0.537	<0.0001	71	0.146	0.092–0.287	0.012	69
Left elbow	0.403	0.258–0.446	<0.0001	64	0.245	0.166–0.392	0.0001	79
Right wrist	0.392	0.263–0.548	<0.0001	78	0.121	0.007–0.168	0.030	123
Left wrist	0.325	0.261–0.419	<0.0001	72	0.042	–0.192 to 0.234	0.259	145
Right MCP1	0.394	0.326–0.551	<0.0001	50	–0.166	–0.178 to 0.151	0.995	134
Right MCP2	0.298	0.225–0.362	<0.0001	53	0.13	0.109–0.140	0.023	158
Right MCP3	0.282	0.153–0.323	<0.0001	49	0.218	0.070–0.292	0.0004	109
Right MCP4	0.297	0.235–0.379	<0.0001	33	0.397	0.376–0.476	<0.0001	53
Right MCP5	0.379	0.236–0.477	<0.0001	36	0.306	0.272–0.387	<0.0001	63
Left MCP1	0.304	0.187–0.394	<0.0001	44	0.094	0.024–0.148	0.073	95
Left MCP2	0.444	0.391–0.619	<0.0001	36	0.262	0.133–0.314	<0.0001	110
Left MCP3	0.382	0.323–0.453	<0.0001	41	0.27	0.268–0.334	<0.0001	69
Left MCP4	0.315	0.086–0.375	<0.0001	32	0.177	0.059–0.232	0.003	45
Left MCP5	0.353	0.183–0.569	<0.0001	26	0.217	0.172–0.266	0.0004	71
Right PIP1	0.299	0.050–0.367	<0.0001	22	0.128	0.087–0.227	0.023	53
Right PIP2	0.353	0.310–0.379	<0.0001	28	0.042	0.015–0.094	0.258	95
Right PIP3	0.351	0.312–0.486	<0.0001	30	–0.017	–0.102 to 0.113	0.603	118
Right PIP4	0.232	0.111–0.250	<0.0001	28	0.074	0.037–0.117	0.126	82
Right PIP5	0.396	0.294–0.431	<0.0001	26	0.245	0.091–0.318	0.244	48
Left PIP1	0.165	0.018–0.304	<0.0001	21	0.02	–0.203 to 0.242	0.376	39
Left PIP2	0.167	0.085–0.299	<0.0001	24	–0.061	–0.236 to 0.2	0.827	50
Left PIP3	0.298	0.196–0.368	<0.0001	35	0.077	–0.071 to 0.219	0.116	101
Left PIP4	0.185	0.018–0.267	<0.0001	31	0.121	0.056–0.197	0.030	74
Left PIP5	0.339	0.270–0.573	<0.0001	18	0.113	0.041–0.267	0.040	46
Right knee	0.476	0.318–0.530	<0.0001	48	0.428	0.295–0.525	<0.0001	58
Left knee	0.379	0.350–0.471	<0.0001	57	0.214	0.038–0.362	0.0005	62
Overall agreement (%)	0.449 (81.7%)	0.403–0.493	<0.0001		0.227 (66%)	0.203–0.250	<0.0001	

MCP, metacarpophalangeal; PIP, proximal interphalangeal

**Agreement and concordance between physical and US examination**

What is interesting in Table 3 is that, from the then joints assessed by both methods, there has been a slight relative increase in the number of swollen joints through US examination (345 of 800 joints assessed by US (43%)), when compared to physical examination (946 of 2400 joints assessed by our three rheumatologist (39%)). Concordance strength of agreement was fair and ranged from Fleiss kappa 0.168–0.264. Likewise, the k coefficients between physical and US examination in the left MCP5 joint ( $k=0.168$ ,  $p$  0.06) followed by left MCP4 ( $k=0.177$ ,  $p$  0.053) joint stand out as the lowest values. By contrast, k coefficient in wrist was slightly higher ( $k=0.241$  ( $p$  0.007) and  $k=0.213$  ( $p$  0.023)). The percentage of overall agreement and

**Table 3** Agreement and concordance between physical and US examination swelling assessment

Joint	Overall agreement (%)	Fleiss kappa	p value
Right wrist	62.5	0.241	0.007
Left wrist	63.7	0.213	0.022
Right MCP 2	60.0	0.195	0.040
Right MCP 3	66.2	0.264	0.009
Right MCP 4	69.6	0.197	0.039
Right MCP 5	70.9	0.244	0.014
Left MCP 2	62.5	0.244	0.011
Left MCP 3	67.5	0.228	0.019
Left MCP 4	77.5	0.177	0.053
Left MCP 5	71.2	0.168	0.06

MCP, metacarpophalangeal; PIP, proximal interphalangeal

concordance rate was 67.18% and  $k=0.210$ ; ( $p=0.027$ ), respectively.

## Discussion

As explained in the introduction, it is clear that training sessions focused on the standardization of joint assessment techniques play a pivotal role in understanding the reliability of quantitative indexes daily used in RA patients. We aimed to evaluate the correlation and agreement between rheumatologists after such a training session and compare its performance with a reference imaging modality, namely, MSUS, to find a reasonable approach on behalf of the inevitable high intra- and interobserver variability described in the literature.

Contrary to expectations, despite the fact that all three participating rheumatologists attended the proposed training sessions, interobserver variability among them was still present. The wide range of perceived concordance rates suggests that the assessment of some individual joints may be particularly challenging on their own, thereby showing specific difficulties during standardization, with possible attribution to an underlying long and comprehensive learning process.

In terms of overall interobserver agreement and concordance rates, as well as of overall correlation, our findings propose that tenderness assessment was more homogenous than swelling assessment. A possible explanation for these results may be the fact that joint tenderness is inferred solely by the patient's response, while that of joint swelling by physicians' technique and interpretation of findings.

The importance of training sessions focused on standardization was first stated by Scott et al. in 1996 [19], whose findings suggested a considerably increased sensitivity of measurement for both tender and swollen joints and a reduction in the mean coefficient of variation for the number of swollen joints (82% vs 59%) after a 60-min training session based on the EULAR handbook for joint evaluation. Unlike Scott et al., on a multicentric cohort study evaluating standardization based on the aforementioned EULAR handbook, Grunke et al. [20] stated that even when consistency and variability significantly improved, the mean number of tender, as well as swollen joints decreased. Nonetheless, a reference imaging modality was not used.

In addition, clinical experience plays an essential part as recently proposed by El-Hadidi et al. [21] where after a consensus on joint assessment, the interobserver agreement was calculated to compare an experienced rheumatology professor (30 years of experience) with a young Rheumatology fellow (3 years of training). Although a high correlation between professor and candidate was described, specific results on joint assessment correlation

are similar to ours, showing a more robust correlation regarding tender joints compared to that of swollen joints.

When compared to MUSC, as previously stated, agreement and concordance were slightly lower for swelling assessment. This discrepancy could be attributed to the mean age of our patients (55 years); it seems possible that the older the patient, the more frequent synovial thickening and incipient joint deformities, thus being confounding for not only MSUS but also for PE assessment. Moreover, it is important to consider what has been widely held by recent reports, that is, the trend towards higher agreement and concordance rates regarding swollen joints count [22, 23]. These data must be interpreted with caution because of the use of the OMERACT synovitis definition and its implications of considering even a minimal amount of intraarticular tissue as an abnormal finding of synovial hypertrophy, thereby involving a potential overestimation.

It could be argued that the average age and the duration and activity of the disease could contribute to the differences between physical examination and US; nevertheless, those patients with joint deformity were excluded. It is relevant to bear in mind that, in those populations with a poorly controlled- and a longer- disease, joint assessment either by physical examination or by ultrasound remains a challenge when considering joint surface irregularities. It supports the pivotal role of conventional radiography as the first choice for the evaluation of structural changes such as erosions.

The limitations of this study include firstly the absence of power Doppler ultrasonographic assessment due to timing issues on behalf of the vast number of evaluated joints; prior studies have noted the role of power Doppler US in detecting subclinical and residual synovitis when assessing synovial vascularity, although, it was beyond the scope of our work to evaluate such conditions. Secondly, we did not count on another radiologist with sufficient experience in MSUS to be considered as a second evaluator; additionally, the US evaluation performed by a radiologist and not an articular US rheumatologist might be considered as a source of uncertainty; nonetheless, the vast experience, as well as the specific MSUS training of our radiologist, support the internal validity of the assessments. Very little was found in the literature on the question of performance differences between those scenarios (radiologist vs. rheumatologist), remaining as an intriguing issue for future research, especially for low and middle-income countries where training for such medical sub-specialties is not as frequent.

Thirdly, the absence of early RA patients, as well as of patients in clinical remission, and finally, that we did not report pretraining session joint count results in order to

propose a variability improvement. The latter consideration must be handled with care when considering that this decision was based on the fact that pretraining session correlation and agreement scores were certainly variable, close to the null value, and even negative. Taking this into consideration, the gain of the stated training session was to revert such tendency and yet turn it to positive values, even though the kappa coefficients were in the range of slight to moderate. It is worth mentioning that when adding more than two observers to the calculations, the slightest discrepancy substantially lowers the statistical estimator; thus, an adjusted threshold must be taken when interpreting such results. Finally, a reasonable approach could be to hypothesize that the increase in the total swollen joint count through US examination would have a substantial implication in the categorization of the activity of the disease; nevertheless, even when it was beyond our scope, we trust that future research on this behalf will rise as a natural progression of this work.

## Conclusions

Taken together, our findings indicate that joint tenderness assessment had a better interobserver agreement and concordance rate, moreover a higher overall correlation coefficient than joint swelling. The evidence from our results suggests a fair overall concordance rate when comparing the US examination to the physical examination, thus supporting the need for the implementation of training sessions dedicated to standardization in rheumatology clinics to guarantee reliability in identifying tender and swollen joints within rheumatologists.

## Acknowledgements

Not applicable.

## Authors' contributions

GQ-L: Conceptualization, methodology, formal analysis, investigation, writing-original draft preparation, writing-review and editing. KM-C: Writing-original draft preparation, writing-review and editing. JBF-S: writing-original draft preparation, writing-review and editing. PM-P: Conceptualization, methodology, investigation. PC-A: Conceptualization, methodology, investigation. EC: Investigation. All authors read and approved the final manuscript.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Availability of data and materials

The data underlying this article will be shared on reasonable request to the corresponding author.

## Declarations

### Ethics approval and consent to participate

This study was conducted with the approval of our Local Research Ethics Committee in accordance with the Declaration of Helsinki. Every patient gave written and spoken informed consent.

### Consent for publication

All authors approved the final version of this manuscript to be published.

## Competing interests

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We agree to bear the applicable publication charges if their manuscript is accepted for publication. All authors declare that they have no competing interest.

## Author details

<sup>1</sup>Reumavance Group, Rheumatology Section, Department of Internal Medicine, Fundación Santa Fe de Bogotá University Hospital, Carrera 7 No. 117-15, 110111 Bogotá, Colombia. <sup>2</sup>Department of Internal Medicine, School of Medicine, Universidad Nacional de Colombia, Bogotá, Colombia. <sup>3</sup>Department of Internal Medicine, School of Medicine, Universidad de Los Andes, Bogotá, Colombia. <sup>4</sup>Department of Radiology, School of Medicine, Universidad Nacional de Colombia, Bogotá, Colombia.

Received: 3 June 2021 Accepted: 5 November 2021

Published online: 18 November 2021

## References

- McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. *N Engl J Med*. 2011;365:2205–19.
- Backhaus M, Kamradt T, Sandrock D, Loreck D, Fritz J, Wolf KJ, et al. Arthritis of the finger joints: a comprehensive approach comparing conventional radiography, scintigraphy, ultrasound, and contrast-enhanced magnetic resonance imaging. *Arthritis Rheum*. 1999;42:1232–45.
- Naredo E, Bonilla G, Gamero F, Uson J, Carmona L, Laffon A. Assessment of inflammatory activity in rheumatoid arthritis: a comparative study of clinical evaluation with grey scale and power Doppler ultrasonography. *Ann Rheum Dis*. 2005;64:375–81.
- Van der Heijde D, Van't Hof M, Van Riel P, et al. Development of a disease activity score based on judgment in clinical practice by rheumatologists. *J Rheumatol*. 1993;20:579–81.
- Prevoe MLL, Van't Hof MA, Kuper HH, et al. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum*. 1995;38(1):44–8.
- Porter D, Gadsby K, Thompson P, White J, McClinton C, Oliver S. DAS28 and rheumatoid arthritis: the need for standardization. *Musculoskeletal Care*. 2011;9(4):222–7.
- Scott IC, Scott DL. Joint counts in inflammatory arthritis. *Clin Exp Rheumatol*. 2014;32(Suppl. 85):S7–12.
- Sokka T, Pincus T. Joint counts to assess rheumatoid arthritis for clinical research and usual clinical care: advantages and limitations. *Rheum Dis Clin N Am*. 2009;35(4):713–22.
- Smolen JS, Landewé RBM, Bijlsma JWJ, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann Rheumatic Dis*. 2020;79:685–99.
- Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology-European League Against Rheumatism collaborative initiative. *Ann Rheum Dis*. 2010;69:1580–8.
- Wakefield RJ, Balint PV, Szkudlarek M, Filippucci E, Backhaus M, D'Agostino MA, et al. OMERACT 7 special interest group. Musculoskeletal ultrasound including definitions for ultrasonography pathology. *J Rheumatol*. 2005;32:2485–7.
- Szkudlarek M, Wakefield R, Backhaus M, Terslev L. The discriminatory capacity of ultrasound in rheumatoid arthritis: active vs inactive, early vs advanced, and more. *Rheumatology*. 2012;51(suppl 7):vii6–9.
- Wakefield RJ, Green MJ, Marzo-Ortega H, Conaghan PG, Gibbon WW, McGonagle D, et al. Should oligoarthritis be reclassified? Ultrasound reveals a high prevalence of subclinical disease. *Ann Rheum Dis*. 2004;63:382–5.
- Szkudlarek M, Court-Payen M, Jacobsen S, Klarlund M, Thomsen H, Østergaard M. Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis. *Arthritis Rheum*. 2003;48(4):955–62.
- Bruyn GA, Iagnocco A, Naredo E, Balint PV, Gutierrez M, Hammer HB, et al. OMERACT definitions for ultrasonographic pathologies and

- elementary lesions of rheumatic disorders 15 years on. *J Rheumatol*. 2019;46(10):1388–93. <https://doi.org/10.3899/jrheum.181095>.
16. Terslev L, Naredo E, Aegerter P, et al. Scoring ultrasound synovitis in rheumatoid arthritis: a EULAR-OMERACT ultrasound taskforce-Part 2: reliability and application to multiple joints of a standardised consensus-based scoring system. *RMD Open*. 2017;3:e000427. <https://doi.org/10.1136/rmdopen-2016-000427>.
  17. Schmidt WA, Schmidt H, Schicke B, Gromnica-Ihle E. Standard reference values for musculoskeletal ultrasonography. *Ann Rheum Dis*. 2004;63:988–94.
  18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74.
  19. Scott DL, Choy EHS, Greeves A, Isenberg DM, Kassiror D, Rankin E, et al. Standardizing joint assessment in rheumatoid arthritis. *Clin Rheumatol*. 1996;15(6):579–82.
  20. Grunke M, Antoni C, Kavanaugh A, Hildebrand V, Dechant C, Schett G, et al. Standardization of joint examination technique leads to a significant decrease in variability among different examiners. *J Rheumatol*. 2010;37(4):860–4.
  21. El-Hadidi K, Gamal S, Saad S. Inter-observer agreement of standard joint count examination and disease global assessment in a cohort of Egyptian Rheumatoid Arthritis patients. *Reumatol Clín*. 2019;15(5):258–63.
  22. Hirata A, Ogura T, Hayashi N, Takenaka S, Ito H, Mizushima K, et al. Concordance of patient-reported joint symptoms, physician-examined arthritic signs, and ultrasound-detected synovitis in rheumatoid arthritis. *Arthritis Care Res*. 2017;69(6):801–6.
  23. Salaffi F, Filippucci E, Carotti M, Naredo E, Meenagh G, Ciapetti A, et al. Inter-observer agreement of standard joint counts in early rheumatoid arthritis: a comparison with grey scale ultrasonography: a preliminary study. *Rheumatology*. 2008;47:54–8.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

