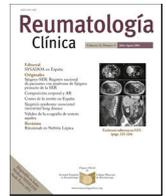




Sociedad Española  
de Reumatología -  
Colegio Mexicano  
de Reumatología

# Reumatología Clínica

www.reumatologiaclinica.org



## Original Article

# Inter-observer agreement of standard joint count examination and disease global assessment in a cohort of Egyptian Rheumatoid Arthritis patients



Khaled El-Hadidi<sup>a</sup>, Sherif M. Gamal<sup>a,\*</sup>, Sahar Saad<sup>b</sup>

<sup>a</sup> Rheumatology and Rehabilitation Department, Cairo University Hospital, Cairo, Egypt

<sup>b</sup> Rheumatology and Rehabilitation Department, Monera General Hospital, Cairo, Egypt

## ARTICLE INFO

### Article history:

Received 28 February 2017

Accepted 13 October 2017

### Keywords:

Inter-observer agreement  
Disease global assessment  
Egyptian patients

## ABSTRACT

**Objectives:** To assess the inter-observer agreement of standard joint count between experienced Rheumatology professor (Prof) and young Rheumatology fellow (candidate), and to compare disease global assessment between professor, young candidate and patients.

**Methods:** This study included one hundred rheumatoid arthritis patients. For all patients independent clinical evaluation was done by two rheumatologists (professor and candidate) for detection of tenderness in 28 joints and swelling in 26 joints. The study also involved global assessment of disease activity by the provider (Prof and candidate) (EGA) as well as by the patient (PGA). The EGA was determined without previous knowledge of the patient's laboratory test results.

**Results:** A highly significant accordance (correlation) between professor and candidate was found in both the number of tender joints ( $p < 0.001$ ) ( $r = 0.946$ ), and the number of swollen joints ( $p < 0.001$ ) ( $r = 0.797$ ). Regarding swollen joints, the highest agreement was in right knee (0.929), while poor agreement was found in the right 5th MCP (0.049). Regarding tender joints, the highest analogy was in the right elbow (0.899), in contrast to the left 3rd PIP (0.462) which showed the least congruence. Agreement study using kappa measurement for disease global assessment showed: moderate agreement (between professor and candidate) (0.405), fair agreement between (professor and patient) (0.213), fair agreement between (candidate and patient) (0.367).

**Conclusion:** Inter-observer reliability was better for TJCs than SJCs. Regarding SJCs agreement was better in large joints such as the knees compared to the small joints such as the MCPs. Disease global assessment may show discrepancy between patients and physicians.

© 2017 Elsevier España, S.L.U. and Sociedad Española de Reumatología y Colegio Mexicano de Reumatología. All rights reserved.

## Acuerdo entre observadores del examen estándar de recuentos articulares y la escala visual analógica en una cohorte de pacientes con artritis reumatoide egipcia

## RESUMEN

**Objetivos:** Evaluar el acuerdo interobservador del recuento de articulaciones estándar entre el profesor experimentado y el candidato joven, y comparar la evaluación global de la enfermedad entre el profesor, el candidato joven y los pacientes.

**Métodos:** Este estudio incluyó a 100 pacientes con artritis reumatoide. Para todos los pacientes, la evaluación clínica independiente fue realizada por 2 reumatólogos (profesor y candidato) para la detección de la sensibilidad en 28 articulaciones y la hinchazón en 26 articulaciones. También en este estudio evaluaciones globales de la actividad de la enfermedad por el proveedor (prof y candidato) Se realizó el paciente (PGA) para cada paciente. El EGA se determinó sin conocimiento previo de los resultados de la prueba de laboratorio del paciente.

### Palabras clave:

Acuerdo entre observadores  
Evaluación global de la enfermedad  
Pacientes egipcios

\* Corresponding author.

E-mail address: [sherif775@hotmail.com](mailto:sherif775@hotmail.com) (S.M. Gamal).

**Resultados:** Se encontró una concordancia altamente significativa (correlación) entre el profesor y el candidato tanto en el número de articulaciones sensibles ( $p < 0.001$ ) ( $r = 0.946$ ), y el número de articulaciones inflamadas ( $p < 0.001$ ) ( $r = 0.797$ ). En cuanto a las articulaciones inflamadas, el mayor acuerdo fue en la rodilla derecha (0,929), mientras que el acuerdo pobre fue encontrado en el 5to MCP derecho (0,049). En cuanto a las articulaciones sensibles, la mayor analogía fue en el codo derecho (0,899), encontraste con el tercer PIP izquierdo (0,462) que mostró la menor congruencia. El estudio del acuerdo utilizando la medición kappa para la evaluación global de la enfermedad mostró: Acuerdo moderado (entre profesor y candidato) (0,405), Acuerdo justo entre (profesor y paciente) (0,213), Acuerdo justo entre (candidato y paciente) (0,367).

**Conclusión:** La confiabilidad entre observadores fue mejor para TJs que SJs. Con respecto a los SJs, el acuerdo fue mejor en las articulaciones grandes, como las rodillas en comparación con las articulaciones pequeñas, como los MCP. La evaluación global de la enfermedad puede mostrar discrepancias entre los pacientes y los médicos.

© 2017 Elsevier España, S.L.U.

y Sociedad Española de Reumatología y Colegio Mexicano de Reumatología. Todos los derechos reservados.

## Introduction

Rheumatoid arthritis (RA) is the most common chronic inflammatory joint disease.<sup>1</sup> RA is characterized by chronic inflammation, synovial proliferation and excessive proinflammatory cytokine production, leading to cartilage and bone destruction.<sup>2</sup> It is clear that early diagnosis, referral, and treatment of patients with RA results in improvement in clinical signs and prevention of joint destruction.<sup>3</sup>

Achievement of low disease activity or even remission is an attainable target in RA.<sup>4</sup> Accurate assessment of disease activity and joint damage in RA is important in rheumatological practice to enable therapeutic decisions and to evaluate disease outcome and response to treatment.<sup>5,6</sup> Joint examination is a prerequisite to the diagnosis of RA, and quantitative counts of swollen and tender joints remain the most specific tools for patient assessment.<sup>7</sup> Disease severity is directly related to the number of swollen and tender joints. Consequently, research and clinical trials on the use of disease-modifying anti-rheumatic drugs and biologicals depend on reductions in the number of swollen and tender joints as key outcome measures.<sup>8</sup>

Joint counts are included in most indices of disease activity,<sup>9</sup> and rheumatologists should include a joint count assessment for each RA patient at every visit.<sup>8</sup> Although joint counts are crucial for the assessment of RA, their reliability however is an issue.<sup>10</sup> Several studies<sup>11–14</sup> reported considerable variations in joint counts between both individual observers and centres in routine clinical practice and even in clinical trials. Despite this, determining the number of swollen and tender joints is the most stable method of clinical assessment. This may explain why training to standardize methods,<sup>15</sup> increases the sensitivity of counting swollen and tender joints and reduces the variability of measurement, though it does not entirely abolish it.<sup>8</sup>

Patient global assessment (PGA) is one of the most widely used patient-reported outcomes in RA practice and is included in several composite scores for RA assessment. PGA is often assessed by a single question with a 0–10 or 0–100 scale.<sup>16</sup> However, in RA, assessment of disease activity may differ between patients and physicians.<sup>17</sup> Such discordance between patients and clinicians in their evaluation of RA activity has been widely noted.<sup>18,19</sup> Most often, patients rate their RA as more active than clinicians. This difference may raise questions about the validity of these measures, and about whose assessment should be used to guide treatment decisions.<sup>20</sup>

Thus the aim of the current study is to evaluate inter-observer agreement between professor and candidate concerning swollen and tender joint counts, and to determine the joints that show low inter-observer agreement. Another goal was to compare provider global assessment (EGA) performed by the professor versus EGA done by the candidate on one hand, and the provider global

assessment (EGA) to the patients global assessment (PGA) on the other hand.

## Patients and methods

This study included one hundred rheumatoid arthritis (RA) patients who fulfilled the ACR/EULAR 2010 Classification Criteria for RA.<sup>21</sup> Patients were attending the Rheumatology and Rehabilitation Department of the Kasr El-Aini Hospitals & Specialized Manial Hospital, Cairo University. Patients' ages ranged from 24 to 75 years with a mean of 45.12 years, mean disease duration was  $8.14 \pm 6.84$  years. Our patients were 6 males and 94 females who were consecutively collected over a period of 7 months starting from May 2014 until November 2014. All patients included in the study were subjected to full history taking, thorough clinical examination, including local joint examination.

Local joint evaluation was done by two rheumatologists who carried out a consensus on joint assessment before the study: the first, a skilled rheumatology professor (Prof) widely experienced in joint evaluation, with more than thirty years experience in University hospitals, and the second, a young Rheumatology fellow (candidate) who has been trained by different rheumatology professors and has performed more than 500 supervised joint count examinations in about three years training in a University hospital.

Clinical evaluation was performed independently and sequentially by the two rheumatologists in the same day for detection of tenderness in 28 joints and swelling in 26 joints. Patients evaluation was done by the candidate without being informed of the results of the professor's evaluation.

The following 26 joints were assessed bilaterally for tenderness and swelling: [10 metacarpophalangeal (MCPs), 10 proximal interphalangeal (PIPs) joints, both elbows, both wrists and both knees], while both shoulder joints were assessed for tenderness only. Clinical inter-observer agreement for both tenderness and swelling was calculated.

The sum of the tender joint count (TJC) as well as the swollen joint count (SJC) was recorded for each patient. Correlation was done regarding the total number of swollen, tender joints in each patient comparing the results of the professor to those of the candidate using spearman's rho test.

Also in this study, global assessment of disease activity for each patient was performed by both the provider (professor and candidate) (EGA) on one hand, and the patient (PGA) on the other hand. The EGA was determined without previous knowledge of the patient's laboratory test results. The question for determining the PGA was, "How do you estimate your disease activity today". The assessment was done using a 0–100 scale.<sup>22</sup> CDAI was used to assess disease activity by professor and candidate.<sup>23</sup>

### Statistical methods

Data management and statistical analysis were performed using Statistical Package for Social Sciences version 21 (SPSS Inc, Chicago, IL).

Numerical data were summarized using mean and standard deviation or median and range. Categorical data was summarized as percentages. Correlations were determined by using Spearman's rho test. As a measure of reliability, Cohen's kappa was used to assess the degree of agreement between professor and candidate. All *p*-values are two-sided. *p*-values < 0.05 were considered significant.

### Correlation

In statistics, the correlation coefficient *r* measures the strength and direction of a linear relationship between two variables on a scatter plot. The value of *r* is always between +1 and –1 (with +1 being exactly perfect uphill (positive) linear relationship and –1 which is exactly a perfect downhill (negative) linear relationship).

The magnitude of agreement by Kappa coefficient is graded/evaluated as follows: Almost perfect 0.8–1, substantial 0.6–0.8, moderate 0.4–0.6, fair 0.2–0.4, Poor 0–0.2, and no agreement for values less than zero.

### Results

#### Concerning tender and swollen joint examination

We compared the results of the examination performed by the professor versus that performed by the candidate regarding tenderness in 28 joints and swelling in 26 joints. The degree of agreement was measured by Cohen's kappa. (of 28 joints for tenderness and 26 joints for swelling between the professor and the candidate in all patients involved in this study and we measured the degree of agreement by Kappa measurement.)

Regarding tender joints, the highest agreement was in the right elbow (0.899), while the lowest agreement was in the left 3rd PIP (0.462) as shown in [Table 1](#).

Regarding swollen joints, the highest accordance was in the right knee (0.929), while there was no agreement in the right 5th MCP (0.049) as shown in [Table 2](#).

Details of agreement regarding tender and swollen joints are shown in [Tables 3–4](#) (in descending manner from highest to lowest agreement).

### Correlation

Correlation was done between the results of the professor and those of the candidate regarding the total number of swollen, tender joints in each patient using Spearman's rho test. The results showed highly significant correlation regarding tenderness ( $p < 0.001$ ), swollen joints ( $p < 0.001$ ); however the correlation was stronger regarding tender joints ( $r = 0.946$ ) compared to the swollen joints ( $r = 0.797$ ).

#### Disease global assessment

According to **EGA of the professor**: there were 18 patients with mild activity, 56 patients with moderate activity and 26 patients with severe activity.

According to **EGA of candidate**: there were 19 patients with mild activity, 50 patients with moderate activity and 31 patients with severe activity.

**Table 1**

Tender joints from highest to lowest degree of agreement.

	Kappa	Degree of agreement	<i>p</i> value
Rt. elbow	0.899	Almost perfect	<0.001
Rt. wrist	0.879	Almost perfect	<0.001
Lt. wrist	0.840	Almost perfect	<0.001
Rt. knee	0.832	Almost perfect	<0.001
Lt. elbow	0.820	Almost perfect	<0.001
Lt. knee	0.813	Almost perfect	<0.001
Lt. shoulder	0.798	Substantial	<0.001
Lt. 4th MCP	0.793	Substantial	<0.001
Rt. 2nd MCP	0.734	Substantial	<0.001
Rt. 3rd MCP	0.729	Substantial	<0.001
Rt. 5th MCP	0.684	Substantial	<0.001
Rt. shoulder	0.670	Substantial	<0.001
Rt. 1st PIP	0.662	Substantial	<0.001
Lt. 3rd MCP	0.638	Substantial	<0.001
Rt. 1st MCP	0.633	Substantial	<0.001
Rt. 4th MCP	0.629	Substantial	<0.001
Rt. 4th PIP	0.617	Substantial	<0.001
Lt. 2nd MCP	0.609	Substantial	<0.001
Rt. 3rd PIP	0.607	Substantial	<0.001
Rt. 2nd PIP	0.582	Moderate	<0.001
Lt. 2nd PIP	0.563	Moderate	<0.001
Lt. 1st PIP	0.530	Moderate	<0.001
Lt. 4th PIP	0.513	Moderate	<0.001
Lt. 5th MCP	0.505	Moderate	<0.001
Lt. 5th PIP	0.488	Moderate	<0.001
Rt. 5th PIP	0.480	Moderate	<0.001
Lt. 1st MCP	0.472	Moderate	<0.001
Lt. 3rd PIP	0.462	Moderate	<0.001

**Table 2**

Swollen joints from highest to lowest degree of agreement.

	Kappa	Degree of agreement	<i>p</i> value
Rt. shoulder		It was very difficult to assess in both reviewers	
Lt. shoulder			
Rt. knee	0.929	Almost perfect	<0.001
Lt. knee	0.825	Almost perfect	<0.001
Lt. wrist	0.739	Substantial	<0.001
Rt. 2nd PIP	0.718	Substantial	<0.001
Lt. 5th PIP	0.621	Substantial	<0.001
Rt. 5th PIP	0.589	Moderate	<0.001
Lt. 4th PIP	0.576	Moderate	<0.001
Rt. wrist	0.564	Moderate	<0.001
Rt. 3rd PIP	0.554	Moderate	<0.001
Lt. 3rd PIP	0.544	Moderate	<0.001
Rt. 2nd MCP	0.540	Moderate	<0.001
Lt. 3rd MCP	0.525	Moderate	<0.001
Lt. 1st PIP	0.525	Moderate	<0.001
Rt. 3rd MCP	0.497	Moderate	<0.001
Lt. 2nd MCP	0.443	Moderate	<0.001
Lt. 5th MCP	0.411	Moderate	<0.001
Lt. 4th MCP	0.406	Moderate	<0.001
Lt. elbow	0.392	Fair	<0.001
Lt. 2nd PIP	0.390	Fair	<0.001
Lt. 1st MCP	0.313	Fair	0.002
Rt. 4th PIP	0.313	Fair	0.001
Rt. 1st PIP	0.289	Fair	<0.001
Rt. elbow	0.273	Fair	<0.001
Rt. 4th MCP	0.272	Fair	0.002
Rt. 1st MCP	0.233	Fair	0.018
Rt. 5th MCP	0.049	Poor	0.625

According to **PGA of patient**: there were 12 patients with mild activity, 38 patients with moderate activity and 50 patients with severe activity.

Agreement study using kappa measurement showed:

- Moderate agreement between (professor and candidate) (0.405)
- Fair agreement between (professor and patient) (0.213)
- Fair agreement between (candidate and patient) (0.367)

**Table 3**  
Candidate EGA with patient PGA.

	Pt.PGA						Kappa	p value	Degree of agreement
	Mild activity		Moderate activity		Severe activity				
<i>Cand. EGA</i>									
Mild activity	10	10.0%	9	9.0%	0	0.0%	0.367	<0.001	Fair
Moderate activity	2	2.0%	24	24.0%	24	24.0%			
Severe activity	0	0.0%	5	5.0%	26	26.0%			
Total	12		38		50				

**Table 4**  
Prof. with patient EGA.

	Pt.PGA						Kappa	p value	Degree of agreement
	Mild activity		Moderate activity		Severe activity				
<i>Prof. EGA</i>									
Mild activity	7	7.0%	9	9.0%	2	2.0%	0.213	0.002	Fair
Moderate activity	5	5.0%	23	23.0%	28	28.0%			
Severe activity	0	0.0%	6	6.0%	20	20.0%			
Total	12		38		50				

**Table 5**  
Prof. with candidate EGA.

	Cand. EGA						Kappa	p value	Degree of agreement
	Mild activity		Moderate activity		Severe activity				
<i>Prof. EGA</i>									
Mild Activity	13	13.0%	5	5.0%	0	0.0%	0.405	<0001	Moderate
Moderate Activity	6	6.0%	35	35.0%	15	15.0%			
Severe Activity	0	0.0%	10	10.0%	16	16.0%			
Total	19		50		31				

**Table 6**  
Correlation regarding disease global assessment between Prof, candidate and patients.

Spearman's rho	Cand. EGA	Prof. PGA
<i>Pt. VAS</i>		
r	0.718	0.580
p value	0.000	0.000
N	100	100
<i>Cand. VAS</i>		
r		0.766
p value		0.000
N		100

- Details are shown in the following [Tables 3–5](#).

Highly significant correlation was found between (professor and candidate) (professor and patient) (candidate and patient); however the *r* value was highest in (professor and candidate) ( $r = 0.766$ ) and lowest in (professor and patient) ( $r = 0.580$ ) as shown in [Table 6](#).

By using CDAI for assessment of disease activity we found:

- Remission was found in 2 patients by professor and in 1 patient by candidate.
- Mild activity was detected in 12 patients by professor and 11 patients by candidate.
- Moderate activity was determined in 28 patients by professor and 30 patients by candidate.
- Severe activity was observed in 58 patients by both the professor and the candidate.

Accordance in CDAI between professor and candidate using kappa measurement was substantial (0.754).

## Discussion

Accurate assessment of disease activity in rheumatoid arthritis (RA) is crucial for establishing disease severity and monitoring response to treatment.<sup>24</sup> Counting the number of swollen joints is a clinical method of quantifying the amount of inflamed synovial tissue.<sup>25</sup> However, one of the most important limitations to the joint count is its poor reproducibility with a requirement to be performed by the same observer at each visit.<sup>7</sup> Also discordance between patient and physician ratings of RA severity occurs in clinical practice and correlates with pain scores and measurements of joint disease.<sup>17</sup>

Thus in this study we tried to assess the inter-observer agreement of standard joint count between experienced professor and young candidate and to show in which joints inter-observer reproducibility could be of value, and to compare provider and patient global assessment between professor, young candidate and patients.

In this study Kappa measurement was used for evaluation of degree of agreement between professor and candidate regarding joint tenderness and swelling.

Regarding swollen joints, agreement ranged from (poor agreement as in the right 5th MCP (0.049) to almost perfect agreement as in the right knee (0.929)), however agreement was reported as (Fair – moderate) in the vast majority of the joints. It is to be noted that agreement was much better in large joints as knees than in small joints as MCPs.

Regarding tender joints, agreement ranged from (moderate as in left 3rd PIP (0.462)) to almost perfect as in right elbow (0.899). However, agreement was reported as (moderate-substantial) in the vast majority of the joints, and it is also to be noted that agreement was much better in large joints as elbows, wrists, and knees as compared to small joint such as MCPs and PIPs.

Salaffi et al.<sup>9</sup> conducted a similar study to assess the inter-observer agreement of standard joint count and to compare clinical examination with grey scale ultrasonography (US). Their study was conducted on 44 patients with recent-onset RA (disease duration < 2 years), and they reported fair to moderate inter-observer agreement on individual joint counts for either tenderness or joint swelling apart from the glenohumeral joint.

Results reported by Salaffi et al.<sup>9</sup> were quite similar to our results. Nevertheless, inter-observer agreement was better in our study, which may be explained by the type of patients used in Salaffi's study (recent-onset RA) and may also be due to the lesser experience of the second examiner in Salaffi's study (who has examined only 300 joints before the study).

Regarding disease global assessment, by using kappa measurement we found, moderate agreement between (professor and candidate), while fair agreement was found between (candidate and patient), as well as (professor and patient). Such discrepancy between patients and physicians in their perceptions of rheumatoid arthritis disease activity was reported by Studenic et al.<sup>26</sup> in their study conducted on 646 rheumatoid arthritis patients. Also Khan et al.<sup>27</sup> found that nearly 36% of patients had discordance in RA activity assessment with their physicians. This may be explained by the fact that (when we know that) pain was overwhelmingly the single most important determinant of patient global assessment, followed by fatigue. In contrast, physician global assessment was most influenced by swollen joint count (SJC), followed by erythrocyte sedimentation rate (ESR).<sup>27</sup> Thus it is commonly reported that in RA, physicians tend to rate disease activity lower than patients; however, scoring differences in both directions have been reported.<sup>17,28</sup> Similarly Hernández-Cruz and Cardiel<sup>29</sup> reported that the reliability of most of the outcome measures for RA was good, especially for those measurements evaluated by a rheumatologist, while those requiring patient participation need to be improved.

Another study done<sup>24</sup> to evaluate the degree of patient – physician discordance in assessment of global disease severity in RA, showed that nearly one third of RA patients differed from their physicians by a meaningful degree in assessment of global disease severity, and they found that discordance was much higher in patients with depressive symptoms. Gvozdenovic et al.<sup>30</sup> found differences between patient global disease activity (PtGDA) and physician global disease activity (PhGDA) which vary from one country to another and attributed this to the culture and level of education of the each country.

Nikiphorou et al.<sup>16</sup> reported that the limitations in the use of PGA in RA may be due to the several possible ways of measuring PGA, including the intended assessment or underlying concept (i.e., global health versus disease activity) and variations in wording/phrasing may lead to differences in interpretation of PGA.

Finally, we can say that, although in our study inter-observer agreement for swollen joints ranged from fair to moderate, further training, standardization and specification may be required to improve inter-observer agreement, especially with availability of multiple treatment options and multiple biological treatment now days, as the use of such treatment, depend mainly on clinical evaluation and patients disease activity. Also attention should be paid to improve disease global assessment, especially the patients' version (PGA), to improve assessment of RA patients and therefore treatment decisions.

#### Conflict of interest

None declared.

#### References

- Hense S, Luque Ramos A, Callhoff J, Albrecht K, Zink A, Hoffmann F. Prevalence of rheumatoid arthritis in Germany based on health insurance data: regional differences and first results of the PROCLAIR study. *Z Rheumatol.* 2016;75: 819–27.
- Abdel-Wahab SM, Tharwat I, Atta DS, El-Sammak AA, Atef R. Serum level of interleukin-33 in rheumatoid arthritis patients and its association with bone erosion and interstitial lung disease. *Egypt Rheumatol.* 2016;38:99–104.
- Breedveld FC, Weisman MH, Kavanaugh AF, Cohen SB, Pavelka K, van Vollenhoven R. The PREMIER study: a multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. *Arthritis Rheum.* 2006;54:26–37.
- Aletaha D, Kerschbaumer A. Rheumatoid arthritis. *Z Rheumatol.* 2017. <http://dx.doi.org/10.1007/s00393-016-0251-7>.
- Ohrndorf S, Backhaus M. Advances in sonographic scoring of rheumatoid arthritis. *Ann Rheum Dis.* 2013;72 Suppl. 2:i69–75.
- Alcalde M, D'Agostino MA, Bruyn GA, Möller I, Iagnocco A, Wakefield RJ, et al. OMERACT Ultrasound Task Force. A systematic literature review of US definitions, scoring systems and validity according to the OMERACT filter for tendon lesion in RA and other inflammatory joint diseases. *Rheumatology (Oxford).* 2012;51:1246–60.
- Sokka T, Pincus T. Joint counts to assess rheumatoid arthritis for clinical research and usual clinical care: advantages and limitations. *Rheum Dis Clin North Am.* 2009;35:713–22, v–vi.
- Scott DL, Antoni C, Choy EH, Van Riel PC. Joint counts in routine practice. *Rheumatology (8-Oxford).* 2003;42:919–23.
- 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 (Oxford).* 2008;47:54–8.
- Cheung PP, Gossec L, Mak A, March L. Reliability of joint count assessment in rheumatoid arthritis: a systematic literature review. *Semin Arthritis Rheum.* 2014;43:721–9.
- Lewis PA, O'Sullivan MM, Rumpf WR, Coles EC, Jessop JD. Significant changes in Ritchie scores. *Br J Rheumatol.* 1988;27:32–6.
- Thompson PW, Hart LE, Goldsmith CH, Spector TD, Bell MJ, Ramsden MF. Comparison of four articular indices for use in clinical trials in rheumatoid arthritis: patient, order and observer variation. *J Rheumatol.* 1991;18:661–5.
- Scott DL, Panayi GS, van Riel PL, Smolen J, van de Putte LB. Disease activity in rheumatoid arthritis: preliminary report of the Consensus Study Group of the European Workshop for Rheumatology Research. *Clin Exp Rheumatol.* 1992;10:521–5.
- Scott DL. European preferences in assessing rheumatoid arthritis. *J Rheumatol.* 1993;20:542.
- Scott DL, Choy EH, Greeves A, Isenberg D, Kassiner D, Rankin E, et al. Standardising joint assessment in rheumatoid arthritis. *Clin Rheumatol.* 1996;15:579–82.
- Nikiphorou E, Radner H, Chatzidionysiou K, Desthieux C, Zabalán C, van Eijk-Hustings Y, et al. Patient global assessment in measuring disease activity in rheumatoid arthritis: a review of the literature. *Arthritis Res Ther.* 2016;28:251.
- Smolen JS, Strand V, Koenig AS, Szumski A, Kotak S, Jones TV. Discordance between patient and physician assessments of global disease activity in rheumatoid arthritis and association with work productivity. *Arthritis Res Ther.* 2016;18:114.
- Furu M, Hashimoto M, Ito H, Fujii T, Terao C, Yamakawa N, et al. Discordance and accordance between patient's and physician's assessments in rheumatoid arthritis. *Scand J Rheumatol.* 2014;43:291–5.
- Wolfe F, Michaud K, Busch RE, Katz RS, Rasker JJ, Shahouri SH, et al. Polysymptomatic distress in patients with rheumatoid arthritis: understanding disproportionate response and its spectrum. *Arthritis Care Res.* 2014;66:1465–71.
- Ward MM, Guthrie LC, Alba MI. Standards of comparison and discordance in rheumatoid arthritis global assessments between patients and clinicians. *Arthritis Care Res (Hoboken).* 2016;1. <http://dx.doi.org/10.1002/acr.23103>.
- Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham 3rd 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.
- Tugwell P, Boers M, on behalf of the OMERACT Committee. Developing consensus on preliminary core efficacy endpoints for rheumatoid arthritis clinical trials. *J Rheumatol.* 1993;20:555–6.
- Aletaha D, Nell VP, Stamm T, Uffmann M, Pflugbeil S, Machold K, et al. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. *Arthritis Res Ther.* 2005;7:R796–806.
- Barton JL, Imboden J, Graf J, Glidden D, Yelin EH, Schillinger D. Patient-physician discordance in assessments of global disease severity in rheumatoid arthritis. *Arthritis Care Res (Hoboken).* 2010;62:857–64.
- Sokka T, Pincus T. Quantitative joint assessment in rheumatoid arthritis. *Clin Exp Rheumatol.* 2005;23:S58–62.
- Studenic P, Radner H, Smolen JS, Aletaha D. Discrepancies between patients and physicians in their perceptions of rheumatoid arthritis disease activity. *Arthritis Rheum.* 2012;64:2814–23. <http://dx.doi.org/10.1002/art.34543>.

27. Khan NA, Spencer HJ, Abda E, Aggarwal A, Alten R, Ancuta C, et al. Determinants of discordance in patients' and physicians' rating of rheumatoid arthritis disease activity. *Arthritis Care Res (Hoboken)*. 2012;64:206–14.
28. Kaneko Y, Kuwana M, Kondo H, Takeuchi T. Discordance in global assessments between patient and estimator in patients with newly diagnosed rheumatoid arthritis: associations with progressive joint destruction and functional impairment. *J Rheumatol*. 2014;41:6–1061.
29. Hernández-Cruz B, Cardiel MH. Intra-observer reliability of commonly used outcome measures in rheumatoid arthritis. *Clin Exp Rheumatol*. 1998;16:459–62.
30. Gvozdencić E, Koevoets R, Wolterbeek R, van der Heijde D, Huizinga TW, Allaart CF, et al. Assessment of global disease activity in RA patients monitored in the METEOR database: the patient's versus the rheumatologist's opinion. *Clin Rheumatol*. 2014;33:461–6.