REVISTA BRASILEIRA DE REUMATOLOGIA

www.reumatologia.com.br

SOCIEDADE BRASILEIRA DE REUMATOLOGIA

Brief communication

Articular ultrasonography: interobserver reliability in rheumatoid arthritis☆

Melissa Cláudia Bisi[a,*], *Aline Defaveri do Prado*[a], *Cristina Rabelo*[b], *Flávia Brollo*[b], *Inês Guimarães da Silveira*[b], *José Alexandre de Mendonça*[c], *Henrique Luiz Staub*[b]

[a]*Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brazil*
[b]*Faculty of Medicine, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brazil*
[c]*Pontifícia Universidade Católica de Campinas, Campinas, SP, Brazil*

ARTICLE INFO

Article history:
Received on 17 May 2013
Accepted on 27 September 2013

Keywords:
Rheumatoid arthritis
Ultrasonography
Reproducibility, kappa

ABSTRACT

Introduction: Ultrasonography (US) has a recent use in Rheumatology, and the reliability of the method in rheumatoid arthritis (RA) patients has yet to be clarified.
Objective: To test, in a RA survey, the reproducibility of musculoskeletal US performed by rheumatologists with one-year training through re-analysis by a Rheumatologist experienced in the method.
Patients and methods: This cross-sectional study included consecutive RA patients from our tertiary center. US exam was performed in metacarpophalangeal joints, proximal interphalangeal joints, and wrists. Presence of synovitis, power Doppler (PD) signal, bone erosions, and cartilage changes comprised the US parameters evaluated. A kappa value in-between 0.20 and 0.40 was considered fair; in-between 0.41 and 0.60 was moderate; in-between 0.61 and 0.80 was good; and above 0.81 was excellent. Results: We analyzed 1,380 joints of 60 RA patients (78% females, 78% caucasoids). Mean age was 58 ± 11.56 years, mean disease duration was 9.98 ± 7.79 years, mean DAS28 was 3.82 ± 1.53, and mean HAQ was 0.91 ± 0.67. Kappa agreement for synovitis ranged from 0.30 to 0.70; for PD signal, from 0.53 to absolute agreement; for erosions, from 0.70 to 0.97; for cartilage changes, from 0.28 to 0.63.
Conclusion: Although good, moderate and excellent interobserver agreement were obtained for erosions and PD, concordance for synovitis and cartilage changes were less impressive in our patients with active RA. Further studies on standardization of scanning technique are necessary to improve musculoskeletal US reproducibility.

☆ *Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brazil.*

* *Corresponding author.*
E-mail: melibisi@yahoo.com.br (M.C. Bisi).

0482-5004/$ - see front matter.
http://dx.doi.org/10.1016/j.rbre.2013.09.002

Ultrassonografia articular: confiabilidade interobservadores em artrite reumatoide

Palavras-chave:
Artrite reumatoide (AR)
Ultrassonografia (US)
Reprodutibilidade Kappa

RESUMO

Introdução: A ultrassonografia (US) tem uso recente na reumatologia, e a confiabilidade do método em pacientes com artrite reumatoide (AR) ainda está por ser definida.
Objetivo: Testar, em uma pesquisa de AR, a reprodutibilidade da US musculosquelética realizada por reumatologistas com treinamento de um ano por meio da reanálise por um reumatologista com experiência no método.
Pacientes e métodos: Esse estudo transversal incluiu pacientes de AR consecutivos de nosso centro terciário. O exame US foi realizado nas articulações metacarpofalângicas, articulações interfalângicas proximais e pulsos. Os parâmetros avaliados foram: presença de sinovite, sinal de *power* Doppler (PD), erosões ósseas e alterações cartilaginosas. Um valor Kappa entre 0,20 e 0,40 foi considerado razoável; entre 0,41 e 0,60, moderado; entre 0,61 e 0,80, bom; e acima de 0,81, excelente.
Resultados: Analisamos 1380 articulações de pacientes com AR (78% mulheres, 78% caucasoides). Média de idade = 58 ± 11,56 anos, duração média da doença = 9,98 ± 7,79 anos, DAS28 média = 3,82 ± 1,53 e HAQ média = 0,91 ± 0,67. A concordância de Kappa para sinovite variou de 0,30-0,70; para sinal PD, de 0,53 até a concordância absoluta; para erosões, de 0,70-0,97; para alterações cartilaginosas, de 0,28-0,63.
Conclusão: Embora tenha sido obtida concordância interobservadores boa, moderada e excelente para erosões e PD, a concordância para sinovite e alterações cartilaginosas foi menos substancial em nossos pacientes com AR ativa. Há necessidade de novos estudos sobre a padronização da técnica de análise, objetivando a melhora da reprodutibilidade da US musculosquelética.

Introduction

Rheumatoid arthritis (RA) is a chronic autoimmune disease affecting mostly peripheral joints. Radiologically, articular involvement is characterized by cortical bone erosions, culminating with deformities.[1]

Currently, musculoskeletal ultrasonography (US) has become an important tool in the diagnosis and monitoring of rheumatic diseases, especially in RA. This method has shown better sensitivity than clinical evaluation and radiography for detection of rheumatoid synovitis and joint erosion.[2]

US have some advantages when compared to other imaging techniques, such as: it is noninvasive, fast, low-cost, and can display various joints in motion, in addition, can be repeated without major risks, and is well accepted by the patient.[3]

Despite these significant advantages, sonographic findings remain highly operator-dependent requiring professional knowledge in Anatomy, Pathology and techniques allowed by the US machine.[4] This is partly due to the subjective image's assessment and the low degree of standardization of the technique, due to the small number of multicenter studies evaluating the interobserver concordance.[5]

The current study aims to analyze the interobserver agreement of data obtained by two rheumatologists with one year-training in US, in comparison with those ones of an expert on US. This interobserver concordance among rheumatologists of different experiences in US has not been detailed in Brazilian RA patients to date.

Materials and methods

Patients

Patients with RA according to criteria of the American College of Rheumatology 1987 were recruited at Saint Lucas Hospital, [6] Pontifical University Catholic of Rio Grande do Sul (PUC-RS), Porto Alegre, Brazil; for this cross-sectional study, we excluded patients with a prior history of fracture or surgery in the dominant hand. The study was approved by the local ethics committee, and all patients signed a free consent.

Patients screened were submitted blindly to US examination by a rheumatologist. Another rheumatologist carried out the disease activity score (DAS28) calculation. This score defines remission when it is below 2.6; low activity from 2.6 to 3.2; moderate activity from 3.2 to 5.1; and severe activity > 5.1.[7] Patients also responded to the health assessment questionnaire (HAQ); in-between 0 and 1: mild limitation; greater than 1 to 2: moderate limitation; and greater than 2 to 3: severe limitation.[8]

Methods

US examination of wrist, second and third proximal interphalangeal and metacarpophalangeal joints were proceeded with high-resolution machine My Lab 60 (ESAOTE, Genoa, Italy) with high-frequency linear transducer (18 MHz).The PD frequency was from 8.0 to 10.0 MHZ, pulse

repetition frequency (PRF) from 0.5 to 1.0. The examination was performed on the dominant side dorsal and ventral in longitudinal and transverse scan, to evaluate the following parameters: presence of synovitis (qualitative and semi-quantitative), signal of power Doppler (PD, qualitative and semi-quantitative), presence of erosions (qualitative) and cartilage assessment (qualitative and semi-quantitative scores).

Images were recorded and archived in Dropbox site, so that all investigators obtained remote access. The examination was carried out by two rheumatologists with the same one-year level of US training (EULAR basic and intermediate courses), one of them the main author of this study. Each rheumatologist examined independently, at different times, 30 different patients (total database of 60 patients). In addition, each of the two played their own images and recorded separately assessments to be re-analyzed by a rheumatologist expert on musculoskeletal US. This PhD expert has over five years of experience in musculoskeletal US and is a national reference in the field. None of the three involved in the evaluations knew the interpretation of the other. Statistical testing was proceeded using the total data from the two rheumatologists and re-analysis by the expert.

In the US analysis, synovitis was scored by gray scale US as: 0 = absence; 1 = mild (discrete hypoechoic image/anechoic in the joint capsule); 2 = moderate (the joint capsule is elevated parallel to the joint area); and 3 = severe (important distention of the joint capsule).[9]

Quantitative evaluation of synovial inflammatory activity through the PD was classified as: 0 = absence (no signal PD, no intra-articular color signal); 1 = mild (up to 3 color signals or 2 single and 1 confluent signal in the intra-articular area); 2 = moderate (greater than grade 1 to < 50% of the intra-articular area filled with color signals); and 3 = severe (> 50% flow intra-articular area filled with color signals).[10]

The presence of erosions was evaluated in the transverse and longitudinal plane and rated as follows: 0 = no erosion; 1 = very small (< 1mm); 2 = small (1-2 mm); 3=moderate (2-4 mm); and 4 = large (> 4 mm). [11]

Cartilage assessment was divided in: 0 = normal hyaline cartilage; 1 = loss of sharpness of the superficial margin of the hyaline cartilage; 2 = partial thickness defect of cartilage layer; 3 = thickness defect of cartilage with normal subchondral bone; and 4 = complete loss of cartilage layer and subchondral bone involvement.[12]

Statistics

Kappa values were utilized to assess interobserver concordance of variables. The weight kappa was calculated when the linear correlation was below 50%. The PABAK (prevalence-adjusted bias-adjusted kappa) was utilized for linear correlations above 50%.[13] Confidence intervals were obtained using the standard error (SE) of weight kappa (wk) (nonzero) as follows: [interval lower kappa = -1.96 *SE (wk)] and [high range kappa = + 1.96 * SE (wk)].[13]

Kappa values were divided in: < 0.20: poor concordance; between 0.21 and 0.40: fair; between 0.41 and 0.60: moderate; between 0.61 and 0.80: good; and between 0.81 and 1: excellent.[13] The significance level for statistical tests was 5%. Statistical programs used were SPSS 12.1 and WinPepi for different kappa calculation.

Results

Out of the 60 RA patients, 47 (78%) were females, and also 78% were caucasian. Mean age was 58 ± 11.56 years, while the mean disease duration was 9.98 ± 7.79 years. Forty-two patients (70%) tested positive for rheumatoid factor. The mean DAS28 was 3.8, pointing to moderate disease activity, while the mean HAQ (0.91) indicated mild limitations to our patients.

A total of 1,380 images of the 60 patients were scanned by the two investigators. Tables 1 and 2 show the agreement rates of data from the two rheumatologists and the expert. The kappa values disclosed good to excellent agreement for erosion (0.70-0.97); moderate to excellent to PD (0.53-1), here including absolute agreement in the third metacarpophalangeal ventral; and fair to good concordance for synovitis (0.30-0.70) and cartilage changes (0.28-0.63).

Table 1 – Agreement and kappa values for synovitis and power Doppler

	Synovitis			Power Doppler		
	Agreement (%)	Kappa	95% CI	Agreement (%)	Kappa	95% CI
Wrist dorsal	58	0.44[a]	0.27 - 0.61	68	0.53[a]	0.32 - 0.74
Wrist ventral	64	0.47[a]	0.25 - 0.69	90	0.87[a]	0.84 - 0.90
2 MCP dorsal	63	0.51[a]	0.35 - 0.67	91	0.89[a]	0.76 - 1.02
2 MCP ventral	45	0.30[b]	0.13 - 0.47	91	0.83[a]	0.46 - 1.20
2 MCP radial	-	-	-	-	-	-
3 MCP dorsal	66	0.56[a]	0.38 - 0.74	95	0.93[a]	0.75 - 1.11
3 MCP ventral	55	0.33[a]	0.10 - 0.56	100	1[a]	NC
2 PIP dorsal	65	0.53[a]	0.31 - 0.75	85	0.77[a]	0.52 - 1.02
2 PIP ventral	81	0.63[a]	0.35 - 0.91	97	0.93[a]	0.91 - 0.95
3 PIP dorsal	73	0.64[a]	0.41 - 0.87	84	0.77[a]	0.55 - 0.99
3 PIP ventral	80	0.70[a]	0.38 - 1.02	98	0.97[a]	0.97 - 0.98

MCP, metacarpophalangeal joints; PIP, proximal interphalangeal; CI, confidence interval; NC, not calculated.
[a]When agreement > 50%, considered PABAK.
[b]When agreement < 50%, considered weighted kappa.

Table 2 – Agreement and kappa values for erosions and cartilage changes

	Erosion			Cartilage		
	Agreement (%)	Kappa	95% CI	Agreement (%)	Kappa	95% CI
2 MCP dorsal	93	0.87[a]	0.73 - 1.01	67	0.58[a]	0.37 - 0.79
2 MCP ventral	95	0.90[a]	0.90 - 0.90	-	-	-
2 MCP radial	88	0.77[a]	0.60 - 0.94	-	-	-
3 MCP dorsal	85	0.70[a]	0.46 - 0.94	70	0.63[a]	0.45 - 0.81
3 MCP ventral	98	0.97[a]	0.35 - 1.59	-	-	-
2 PIP dorsal	95	0.90[a]	0.36 - 1.44	-	-	-
2 PIP ventral	93	0.87[a]	0.37 - 1.37	-	-	-
3 PIP dorsal	95	0.90[a]	0.36 - 1.44	-	-	-
3 PIP ventral	97	0.93[a]	0.32 - 1.54	-	-	-

MCP, metacarpophalangeal joints; PIP, proximal interphalangeal; CI, confidence interval; NC, not calculated.
[a]When agreement > 50%, considered PABAK.

Discussion

The usefulness of US in monitoring structural changes of rheumatoid joints has been previously reported.[14,15] Technological advances have improved the definition of US images, expanding the spectrum of the method in Rheumatology and other areas.[16,17]

The main objective of our study was to evaluate the interobserver concordance of musculoskeletal US in RA patients, an issue yet to be explored. The central idea was to analyze data from two rheumatologists trained in basic and intermediate US courses with re-analysis by a rheumatologist expert in musculoskeletal US.

The great majority (about three quarter) of our survey of 60 RA patients was of caucasian women. The ratio female-male was similar to that described in Europe and United States.[18] Mean age of our patients was around 60 years, with mean disease duration of approximately 10 years. Age of disease onset, in our survey by 50 years, was higher than previously reported.[18]

As a whole, our RA population showed active disease (mean DAS28 3.8, configuring moderate activity). In fact, only two patients were in remission (DAS28 ≤ 2.6). As for the HAQ (mean value 0.91) our survey showed mild functional limitation; only four patients had severe limitations (HAQ > 2.0).

We analyzed a total of 1380 images as to the presence of synovitis, PD signal, bone erosion and cartilage changes. The usual Cohen kappa coefficient was not appropriate for our study, since we dealt with ordered semiquantitative variables and great heterogeneity in the prevalence of such variables.[19] We then set up to use the weight kappa when the linear correlation was below 50%, and PABAK for concordances above 50%.[20]

The highest concordance in our study (good to excellent kappas; 0.70 to 0.97) concerned the presence of bone erosions. Kappas for PD were moderate to excellent (0.53 to 1.00), including absolute agreement in the third MCP. By looking at the pictures, however, we noticed that no patient had positive PD at this location. As to synovitis, data were well less impressive, with kappas varying from fair to good at the most (0.30 to 0.70). Worthy of note, synovitis and PD are variables that must be analyzed dynamically in the US examination; a subtle change in the transducer angle may spoil the interpretation of these parameters.

The interobserver agreement for musculoskeletal US was evaluated by Naredo et al. in 2006.[21] This project ("Teach the Teachers") included 22 rheumatologists and one experienced radiologist. In hands and wrists, mean kappa value for synovitis was 0.73, just higher than ours; as to erosions, their kappa value (0.64), although conceptually moderate, was lower than the one we obtained.

Iagnocco et al. reported kappa values for synovitis, tenosynovitis and erosions between 0.73 and 0.89; again, concordance for synovitis, but not for erosions, was higher than in our study.[22] According to Gutierrez et al., 4-week training for rheumatologists with no experience in US was enough to achieve moderate to excellent concordance for bone erosions.[23]

Kappas for cartilage changes can also be interpreted as a negative surprise in our study (performance fair to good at the maximum, 0.28 to 0.63). In theory, images of cartilage should have been more reproducible, since their interpretation is static. Cartilage evaluation was only recently standardized;[24] this implies difficulties in training professionals for this parameter. Knowingly, basic and intermediate US courses tend to emphasize training for the synovitis and erosion parameters. Slightly differently from our data, Filippucci et al. reported moderate to good interobserver concordance for cartilage changes (0.56 to 0.76).[24]

The US has been considered an operator-dependent test. For this reason, US studies of interobserver reliability are of great importance. European rheumatologists highly experienced in musculoskeletal US formatted the EULAR Standing Committee for Education and Training in US, in order to spread US knowledge in Rheumatology to different countries.[25] As long as the US standardization takes place, the amount of evidence-based multicenter studies will naturally grow.

Despite the overall good concordance obtained for the majority of variables herein evaluated, we recognize that the knowledge of musculoskeletal US has a learning curve that is dependent on the increasing experience of the examiner and parameters standardization. The US, an emerging extension of physical exam for the rheumatologist, stands not only as a diagnostic tool, but also as a parameter of disease monitoring.

Our study presents logistical limitations that should be mentioned. The RA sample could be larger, for a more reliable statistics. The US procedures were not carried out simultane-

ously. Lastly, we were not able, for the moment, to compare US variables with scores of activity and functional impairment.

In summary, in our survey of active RA patients, the majority of the US variables proved reproducible. There was fair to good interobserver reliability for synovitis and cartilage changes, moderate to excellent for PD and good to excellent for bone erosions. Newer studies should better define the usefulness and reproducibility of musculoskeletal US in RA and other related rheumatic disorders.

Conflicts of interest

The authors declare no conflicts of interest.

REFERENCES

1. Van den Berg WB, Van Lent PL, Joosten LA, Abdollahi-Roodsaz S, Koenders MI. Amplifying elements of arthritis and joint destruction. Ann Rheum Dis 2007;66:iii45-8.
2. Kane D, Balint PV, Sturrock RD. Ultrasonography is superior to clinical examination in the detection and localization of knee joint effusion in rheumatoid arthritis. J Rheumatol 2003;30:966-71.
3. Joshua F, Lassere M, Bruyn GA, Szkudlarek M, Naredo E, Schmidt WA *et al*. Summary findings of a systematic review of the ultrasound assessment of synovitis: proceedings of OMERACT 8. J Rheumatol 2007;34:839-47.
4. Grassi W, Cervini C. Ultrasonography in Rheumatology: an evolving technique. Ann Rheum Dis 1998;57:268-71.
5. 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: 515-22.
6. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
7. Ostergaard M, Ejbjerg B, Szkudlarek M. Imaging in early rheumatoid arthritis: roles of magnetic resonance imaging, ultrasonography, conventional radiography and computed tomography. Best Pract Res Clin Rheumatol 2005;19:91-116.
8. Fries JF, Spitz P, Kraines G, Holman H. Measurement of Patient Outcome in Arthritis. Arthritis Rheum 1980;23:137-45.
9. Bruce B, Fries JF. The Health Assessment Questionnaire (HAQ). Clin Exp Rheumatol 2005;23:S14-8.
10. Szkudlarek M, Court-Payen M, Stranberg C, Klarlund M, Klausen T, Ostergaard M. The Power Doppler ultrasonography for assessment of synovitis in metacarpophalangeal joints of patients with rheumatoid arthritis. A Comparison with dynamic magnetic resonance imaging. Arthritis Rheum 2001;44:2018-23.
11. Wakefield RJ, Balint PV, Szkudlarek M, Filippucci E, Backhaus M, D'Agostino MA *et al*. OMERACT 7 Special Interest Group. Musculoskeletal ultrasound including definitions for Ultrasonographic pathology. J.Rheumatol 2005;32:2485-7.
12. Disler DG, Raymond E, May DA, Wayne JS, McCauley TR. Articular cartilage defects: in vitro evaluation of accuracy and interobserver reliability for detection and grading with US. Radiology 2000;215:846-51.
13. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions, 3rd ed. Hoboken NJ, USA: John Wiley & Sons; 2003.
14. Mendonça JA. Yazbek MA, Laurindo IM, Bertolo MB. Wrist ultrasound analysis of patients with early rheumatoid arthritis. Braz J Med Biol Res 2011;44:11-5.
15. Naredo E, Bonillag, 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.
16. Filippucci E, Iagnocco A, Meenagh G, Riente L, Delle Sediei A, Bombardieri S, et al. Ultrasound imaging for the rheumatologist. Clin Exp Rheumatol 2006;24:1-5.
17. Grassi W, Salaffi F, Filippucci E. Ultrasound in rheumatology. Best Pract Res Clin Rheumatol 2005;19:467-85.
18. Mody GM, Cardiel MH. Challenges in the management of rheumatoid arthritis in developing countries. Best Practice & Research Clinical Rheumatology 2008;22:621-41.
19. Guggenmoos-Holzmann, I. The Meaning of Kappa: Probabilistic Concepts of Reliability and Validity Revisited. J Clin Epidemiol 1996;49:775-82.
20. Vieira AJ, Garrett JM. Understanding Interobserver Agreement: The Kappa Statistic. Family Medicine 2005;37:360-3.
21. Naredo E, Möller I, Maragues C, de Agustin JJ, Scheel AK, Grassi W et al. Interobserver reliability in musculoskeletal ultrasonography: results from a "Teach the Teachers" rheumatologist course. Ann Rheum Dis 2006;65:14-9.
22. Iagnocco A, Ossandon A, Coari G, Conti F, Priori R, Alessandri C, et al. Wrist joint involvement in systemic lupus erythematosus. An ultrasonographic study. Clin Exp Rheumatol 2004;22:621-4.
23. Gutierrez M, Filippucci E, Ruta S, Salaffi F, Blasetti P, Geso LD, Grassi W. Inter-observer reliability of high-resolution ultrasonography in the assessment of bone erosions in patients with rheumatoid arthritis: experience of an intensive dedicated training programme. Rheumatology 2011;50:373-80.
24. Filippucci E, Luz KR, Di Geso L, Salaf F, Tardella M, Carotti M *et al*. Interobserver reliability of ultrasonography in the assessment of cartilage damage in rheumatoid arthritis. Ann Rheum Dis 2010;69:1845-8.
25. Scheel AK, Schmidt WA, Hermann KG, Bruyn GA, D'Agostino MA, Grassi W, et al. Interobserver reliability of rheumatologist performing musculoskeletal ultrasonography: results from a EULAR "Train the Trainer" course. Ann Rheum Dis 2005;64:1043-9.