RHEUMATOLOGY

doi:10.1093/rheumatology/keq441
Advance Access publication 25 January 2011

Concise report

Inter-observer reliability of ultrasound detection of tendon abnormalities at the wrist and ankle in patients with rheumatoid arthritis

Mihaela C. Micu[1], Sara Serra[2], Daniela Fodor[3], Manuel Crespo[4] and Esperanza Naredo[4]

Abstract

Objective. To assess inter-observer reliability in US detection of tendon inflammatory and structural changes at wrists and ankles in RA patients.

Methods. Fourteen consecutive RA patients underwent bilateral US assessment of the extensor carpi ulnaris (ECUT) and tibialis posterior tendons (TPTs) by two blinded rheumatologists, with different level of experience in musculoskeletal (MS) US. Grey scale and power Doppler (PD) US assessment was focused on detection of tenosynovitis, tenosynovial and intra-tendon PD signal and structural lesions (i.e. tendinosis, tendon erosion, partial or total rupture).

Results. The frequency of US findings detected by Investigator 1 was 28.6% for inflammatory changes and 51.8% for structural damage changes while Investigator 2 detected 34 and 53.6% for the corresponding abnormalities. A high overall agreement (82.7%) was found for inflammatory pathology and 89.7% for structural lesions in all tendons. Mean kappa (κ) values for all tendons and pathology was moderate ($\kappa = 0.42$), with fair level of agreement for the wrist region (0.27–0.34) and moderate to good values for the ankle region ($\kappa = 0.47$–0.62). Subclinical abnormalities were detected in 37.5% of the tendons by Investigator 1 and 28.6% of the tendons by Investigator 2.

Conclusions. MSUS showed high overall agreement and fair to moderate inter-observer κ-values between investigators with different levels of experience in detection of tendon pathology at the wrist and ankle in RA patients. Further standardization of scanning method and pathology definitions may improve MSUS reproducibility.

Key words: Musculoskeletal ultrasound, Interobserver reliability, Tendon, Tenosynovitis, Ankle, Wrist, Rheumatoid arthritis.

CLINICAL SCIENCE

Introduction

Musculoskeletal ultrasound (MSUS) is a very useful tool in assessing periarticular and IA anatomical structures involved in rheumatic diseases. MSUS assessment is easy, safe, can be performed immediately after the clinical examination (CE) at the bedside and is very well accepted by patients [1, 2]. Moreover, MSUS allows rapid collection of dynamic multiplanar information about the involved anatomic structures that helps to make an accurate diagnosis and also guides interventional diagnostic and/or therapeutic procedures [3–7].

In symptomatic and asymptomatic patients, MSUS has higher sensitivity in detecting synovitis and erosions in comparison with CE and conventional radiology [8–12]. Moreover, repeated examinations are feasible in patients without radiation exposure. However, because MSUS is still an operator-dependent technique, the experience and anatomy knowledge of the operator along with their capacity to exploit the software possibilities of the US

[1]Rehabilitation II Department, Division of Rheumatology, Rehabilitation Clinical Hospital, Cluj-Napoca, Romania, [2]Rheumatology Department, Hospitais da Universidade Coimbra, Portugal, [3]Internal Medicine Department, Medicala II Clinic, University of Medicine and Pharmacy 'Iuliu Hatieganu', Cluj-Napoca, Romania and [4]Department of Rheumatology, Hospital Universitario Severo Ochoa, Madrid, Spain.

Submitted 10 September 2010; revised version accepted 6 December 2010.

Correspondence to: Mihaela C. Micu, Division of Rheumatology, Clinical Rehabilitation Hospital, Cluj-Napoca, 400306, Viilor str. no. 46–50, Romania. E-mail: mcmicu@yahoo.com

machine greatly influences the acquisition and interpretation of the images. In the literature, variable inter-observer reliability between two or more operators has been reported over recent years [13–16]. This variability may be perceived as an Achilles tendon and limit US application for diagnostic purposes. For this reason, the achievement of optimal inter-observer agreement is of paramount importance for MSUS implementation in rheumatological daily practice. This study aimed at assessing the inter-observer reliability between two rheumatologists with different experience level in using MSUS assessment in daily practice, in detection of tendon inflammation and structural damage in patients with RA.

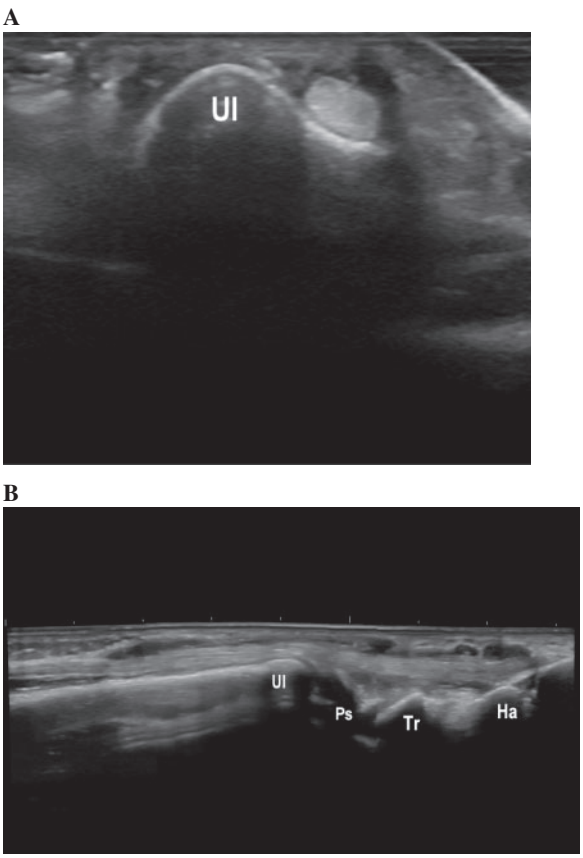
Patients and methods

Fourteen patients with RA (fulfilling the 1987 ACR criteria) who consecutively attended the outpatient rheumatology clinic at the Hospital Universitario Severo Ochoa in March 2010 were prospectively recruited [17]. The mean (S.D.) age of the patients was 61.4 (14.6) years (range 34–79), and the mean (S.D.) disease duration was 6.4 (4) years (range 1–12). The study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of the Severo Ochoa Hospital Madrid. Written informed consent was obtained from each patient.

The patients underwent blinded clinical and power Doppler (PD) US assessment. Two rheumatologists—one highly experienced in MSUS (Investigator 1 with 15 years experience) and one less experienced in MSUS (Investigator 2 with 2 years experience) performed the PDUS evaluation with a commercially available real-time scanner (Logiq 9; GE Medical Systems Ultrasound & Primary Care Diagnostics, LLC; Wauwatosa, WI, USA) using multifrequency linear array transducers (8–14 MHz). Before starting the study, the two investigators spent 30 min to reach consensus on the US examination method and the criteria for pathology of the studied tendons. For each patient, both rheumatologist ultrasonographers consecutively and blindly performed a longitudinal and transverse grey-scale (GS) and PD bilateral examination of the extensor carpi ulnaris tendon (ECUT) at the ulnar stiloid and carpal bones and the tibialis posterior tendon (TPT) at the medial malleolus level. These tendons were chosen because of their frequent involvement in RA [18–20].

The ultrasonographers assessed the presence of tenosynovitis, tenosynovial and intratendon PD signal, tendinosis, tendon erosion and partial and complete tendon tear. Inflammatory and structural pathology was identified according to Outcome Measures in Rheumatology Clinical Trials (OMERACT) definitions and published descriptions of tendon US pathology [21–23]. Tenosynovitis was defined as the presence of hypoechoic or anechoic thickened tissue with or without fluid within the tendon sheath, which is seen in two perpendicular planes and which may exhibit Doppler signal. Tendinosis was defined as loss of normal internal echotexture. Tendon erosion was defined as focal tendon defect. Partial tear was defined as partial-thickness interruption of the tendon fibres.

FIG. 1 (**A**) Transverse US image of tenosynovitis of the ECUT at the ulnar stiloid level. An abnormal hypoechoic thickening of the synovial sheath can be seen. (**B**) Longitudinal US image of tenosynovitis of the ECUT at the ulnar stiloid and carpal bones level. An abnormal hypoechoic thickening of the synovial sheath can be seen. The bony landmarks from proximal (left) to distal (right) are: ulna (Ul), pisiform (Ps), triquetrum (Tr) and hamate (Ha) bones.



Complete tear was defined as full-thickness interruption of the tendon fibres.

GS and PD machine settings were standardized among investigators before the study to optimize PDUS scanning of superficial and deep anatomic areas. These settings were as follows: dynamic range of 72 dB, GS frequency of 12–14 MHz, Doppler frequency of 6.3–7.5 MHz, GS gain of 66 dB, colour gain of 41 dB, low-wall filters and pulse repetition frequency of 500–750 Hz. Flow was additionally demonstrated in two planes and confirmed by pulse wave Doppler spectrum to exclude artefacts.

At the same visit, the patients were evaluated clinically by a third rheumatologists who was unaware of the US findings. The following data were recorded: age, sex, disease duration, RF and CCP antibody status and presence of radiographic erosions. The clinical assessment included the 28-joint DAS (DAS-28) and the presence of

pain and swelling at both wrists and ankles. ESR and CRP values were obtained from laboratory tests performed within 3 days of the visit.

Statistical analysis

Statistical analysis was performed using the software package SPSS, version 9.0.1 (Chicago, IL, USA). Quantitative variables were presented as the mean (S.D.) and range. Categorical variables were presented as absolute frequencies and percentages. Overall agreement, defined as the percentage of exact agreement observed, was calculated for each US abnormality at each region and at all regions. For testing inter-observer reliability, tendon pathology was divided into two main categories: inflammation (tenosynovitis and PD signal) and structural damage (tendinosis, tendon erosions, partial and complete tear). Inter-observer reliability was calculated for each tendon and for each category of pathology using the unweighted κ, calculated with the Landis and Koch method [24]. Values of $\kappa < 0.20$ reflected poor inter-observer agreement, between 0.21 and 0.40 fair, between 0.41 and 0.60 moderate, 0.61 and 0.80 good and >0.80 excellent agreement.

Results

Patient demographic, clinical, laboratory and radiological data were as follows (supplementary table 1, available as supplementary data at *Rheumatology* Online): ratio men/women 3/11 (21.4/78.8%), mean age in the group of 61.4 (S.D.) 14.6 years and mean disease duration of 6.4 (S.D.) 4 years; positive RF and/or CCP antibodies both in 85.7% of the cases and radiographic erosions present in 21.4% of the patients; the mean values of DAS-28 and CRP were 3.15 (1.2) and 7 (7.4), respectively.

Eleven patients out of 14 (78.6%) showed MSUS abnormalities. Investigator 1 found inflammatory changes in 16 (28.6%) tendons and structural damage in 29 (51.8%) tendons while Investigator 2 found tendon inflammatory changes in 19 (34%) and structural damage in 30 (53.7%) (supplementary table 2, available as supplementary data at *Rheumatology* Online). A representative US image of tenosynovitis is shown in Figure 1.

Inter-observer reliability

The overall agreement for US abnormalities in each region and in all regions is presented in Table 1. High overall agreement was obtained for inflammatory abnormalities (82.7%) as well as for structural lesions (89.7%), the lowest agreement being registered in ECUT tenosynovitis (53.8%), whereas absolute agreement (100%) was reached for the ECUT complete tear.

The inter-observer reliability expressed by the unweighted κ-values for all abnormalities in each tendon and for pathologic category at both regions is shown in Table 2. Overall inter-observer agreement was moderate ($\kappa = 0.420$), with a fair level of agreement for the wrist region (0.27–0.37) and moderate to good values for the ankle region ($\kappa = 0.47$–0.62). For the inflammatory findings, we obtained a $\kappa = 0.27$ indicating a fair level of agreement, in comparison with a moderate agreement level for damage findings with $\kappa = 0.52$. In 21 (37.5%) out of 56 tendons, Investigator 1 detected subclinical abnormalities as well as subclinical inflammatory changes in 15 (28.8%) of the tendons. Investigator 2 detected subclinical abnormalities in 16 (28.6%) out of 56 tendons and subclinical inflammatory changes in 9% (16.7) of the tendons.

Discussion

The inter-observer reliability between two or more observers is still an open debate because of the large variability of the results published in the literature [13–16, 25–27]. In recent years, a constant effort has been made to obtain better standardization of the US scanning methods to diminish the disadvantages generated by operator dependence. Moreover, it is well known that wrist and ankle are complex areas, difficult to assess because of the high number of anatomical structures in close relationship. For this reason, agreement even between experienced operators has been rather low in the past. Few studies have previously evaluated the overall reliability of MSUS in detecting combined joint and tendon pathology in RA

Table 1 Overall agreement for each region and all regions for US abnormalities and pathology category

Abnormality on US/region	Wrist (ECUT), %	Ankle (TPT), %	All regions, %
Tenosynovitis	53.6	82.1	67.9
Tenosynovial PD signal	85.7	89.3	87.5
Intra-tendon PD signal	92.96	92.96	92.96
Inflammation			82.7
Tendinosis	71.4	78.6	75.0
Tendon erosion	85.7	100	92.96
Partial rupture	96.4	92.9	94.7
Complete tear	100	96.4	96.2
Damage			89.7

Table 2 Overall κ-values for each tendon and for each pathology category

Wrist	κ
R-ECUT	0.27
L-ECUT	0.34
Ankle	
R-TPT	0.47
L-TPT	0.62
Overall	0.42
Inflammation	0.27
Damage	0.52

R-ECUT: right extensor carpi ulnaris tendon; L-ECUT: left extensor carpi ulnaris tendon; L-TPT: left tibialis posterior tendon; R-TPT: right tibialis posterior tendon.

patients at the wrist and/or ankle level. In a study conducted by Scheel *et al.* [13], between experts, κ-values were moderate ($\kappa = 0.59$) for hand and wrist region with overall agreement of 73% and fair ($\kappa = 0.28$) for ankle region with high overall agreement of 82%. Naredo *et al.* [14], found good κ-values, 0.61 for hand *vs* 0.54 for ankle region with overall agreement between 84 and 91%. Wakefield *et al.* [15] found high overall inter-observer agreement for the detection of abnormalities of the TPT (80%) but with poor agreement for the whole hindfoot area. Another study on tendons conducted by Bruyn *et al.* [16], focused on shoulder area, found moderate inter-observer agreement ($\kappa = 0.46$) for biceps tenosynovitis and good ($\kappa = 0.45$–0.76) for detection of structural damage in the rotator cuff. This variability was explained mostly by the different background of the operators concerning the scanning methods and types of US machine used in daily practice.

As far as we know, our study is the first to evaluate the inter-observer reliability of US assessment of tendon abnormalities in RA. Overall, US evaluation of the tendons is difficult and may expose a not very experienced operator to erroneous interpretations of anisotropy artefact or normal anatomic structures (e.g. retinacula) in complex areas such as the wrist and ankle. Anisotrophy is a US artefact characteristic of tendons, which may appear falsely hypoechoic when the US beam is not perpendicular to the tendon fibres. This false hypoechogenicity can be mistaken for tenosynovitis or tendon tear. Normal retinacula at the wrist and ankle appear at US as focal thickening of the tendon sheath that can be misinterpreted as tenosynovitis. In our work, however, despite the different level of experience and background between the two investigators, high overall agreement was obtained in evaluating both inflammatory changes and structural damage in both anatomic areas. In spite of the high overall agreement, we obtained an overall moderate κ-value ($\kappa = 0.42$), with better results in assessment of the structural damage ($\kappa = 0.52$) as compared with inflammatory changes. This apparently disappointing result is a very well known phenomenon in the literature. In fact it is known as the paradox of the low κ and high agreement and is generated by the low prevalence of the lesions in some cases, creating unbalanced marginal totals on calculation [28]. Maximal expression of this phenomenon occurred in our case in the detection of inflammatory features, especially at wrist level. This effect was favoured by the small number of patients with mean low disease activity and with 6 (42.85%) of them with disease duration <3 years and low prevalence of tenosynovitis. However, the number and characteristics of patients were not predefined but the study included all consecutive patients attending over a 3-week period of daily clinical practice.

Although it was not the objective of our study, in accordance with previous results on synovitis, we found a high frequency of subclinical tendon abnormalities in our RA population [9, 10]. In conclusion, our results showed that US evaluation of the tendons can be a reliable method in detection of inflammatory and structural tendon changes in RA. However, further standardization of the scanning methods and in particular more consensus on definitions are necessary and may improve the reproducibility of MSUS in daily practice and clinical trials.

Rheumatology key messages

- MSUS assessment is easy, safe, can be performed immediately after the CE at bedside.
- MSUS can be a reliable method in detection of inflammatory and structural tendon changes in RA.

Disclosure statement: The authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* Online.

References

1 Grassi W, Cervini C. Ultrasonography in rheumatology: an evolving technique. Ann Rheum Dis 1998;57:268–70.

2 Wakefield RJ, Gibbon WW, Emery P. The current status of ultrasonography in rheumatology. Rheumatology 1999;38:195–201.

3 Fillipucci E, Iagnocco A, Meenagh G *et al*. Ultrasound imaging for the rheumatologist II. Ultrasonography of the hand and wrist. Clin Exp Rheumatol 2006;24:118–22.

4 Meenagh G, Fillipucci E, Kane D, Taggart A, Grassi W. Ultrasonography in rheumatology: developing its potential in clinical practice and research. Rheumatology 2007;46:3–5.

5 Koski JM. Ultrasound guided injections in rheumatology. J Rheumatol 2000;27:2131–8.

6 Micu MC, Bogdan GD, Fodor D. Steroid injection for hip osteoarthritis: efficacy under ultrasound guidance. Rheumatology 2010;49:1490–4.

7 Iagnocco A, Naredo E. Ultrasound-guided corticosteroid injection in rheumatology: accuracy or efficacy? Rheumatology 2010;49:1427–8.

8 Wakefield RJ, Gibbon WW, Conaghan PG *et al*. The value of sonography in the detection of bone erosions in patients with rheumatoid arthritis. Arthritis Rheum 2001;43:2762–70.

9 Wakefield RJ, Green MJ, Marzo- Ortega H *et al*. Should oligoarthritis be reclassified? Ultrasound reveals a high prevalence of subclinical disease. Ann Rheum Dis 2004;63:382–5.

10 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.

11 Szkudlarek M, Narvestad E, Klarlund M *et al*. Ultrasonography of the metatarsophalangeal joints in rheumatoid arthritis: comparison with magnetic resonance

imaging, conventional radiography and clinical examination. Arthritis Rheum 2004;50:2103–12.

12 Karim Z, Wakefield RJ, Qiunn M *et al*. Validation and reproducibility of ultrasound in the detection of synovitis in the knee. Arthritis Rheum 2004;50:387–94.

13 Scheel AK, Schmidt WA, Hermann KG *et al*. Interobserver reliability of rheumatologists performing musculoskeletal ultrasonography: results from a EULAR ''Train the trainers'' course. Ann Rheum Dis 2005;64:1043–9.

14 Naredo E, Moller I, Moragues C *et al*. Interobserver reliability in musculoskeletal ultrasonography: results from a ''Teach the Teachers'' rheumatologist course. Ann Rheum Dis 2006;65:14–9.

15 Wakefield RJ, Freeston JE, O'Connor P *et al*. The optimal assessment of the rheumatoid arthritis hindfoot: a comparative study of clinical examination, ultrasound and high field MRI. Ann Rheum Dis 2008;67:1678–82.

16 Bruyn GAW, Pineda C, Hernandez-Diaz C *et al*. Validity of ultrasonography and measures of adult shoulder function and reliability of ultrasonography in detecting shoulder synovitis in patients with rheumatoid arthritis using magnetic resonance imaging as a gold standard. Arthritis Care Res 2010;62:1079–86.

17 Arnett FC, Edworthy SM, Bloch DA *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988; 31:315–24.

18 Riente L, Delle Sedie A, Iagnocco A *et al*. Ultrasound imaging of the rheumatologist V. Ultrasonography of the ankle and foot. Clin Exp Rheumatol 2006;24:493–8.

19 Grassi W, Fillipucci E, Farina A, Cervini C. Sonographic imaging of tendons. Arthritis and Rheum 2000;43:969–76.

20 Fillipucci E, Iagnocco A, Meenagh G *et al*. Ultrasound imaging for the rheumatologist II. Ultrasonography of the hand and wrist. Clin Exp Rheumatol 2006;24: 118–22.

21 Wakefield R, Balint PV, Szkudlarek M *et al*. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. J Rheumatol 2005;32:2485–7.

22 Bouysset M, Tavernier T, Tebib J *et al*. CT and MRI evaluations of tenosynovitis of the rheumatoid hind foot. Clin Rheumatol 1995;14:303–7.

23 Jernberg ET, Simkin P, Kravette M, Lowe P, Gardner G. The posterior tibial tendon and the tarsal sinus in rheumatoid flat foot: magnetic resonance imaging of 40 feet. J Rheumatol 1999;26:289–93.

24 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33: 159–74.

25 Szkudlarek M, Court-Payen M, Jacobsen S *et al*. Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis. Arthritis Rheum 2003;48:955–62.

26 Iagnocco A, Ossandan A, Coari G *et al*. Wrist joint involvment in systemic lupus erythematosus. An ultrasonographic study. Clin Exp Rheumatol 2004;22: 621–4.

27 Filippucci E, Farina A, Carotti M *et al*. Grey scale and power Doppler sonographic changes induced by intra-articular steroid injection treatment. Ann Rheum Dis 2004; 63:740–3.

28 Feinstein AR, Cicchetti DV. High agreement but low K: I. The problems of two paradoxes. J Clin Epidemiol 1990;43: 543–9.