RHEUMATOLOGY

Rheumatology 2012;51:1655–1661
doi:10.1093/rheumatology/kes103
Advance Access publication 23 May 2012

Original article

Reliability testing of tendon disease using two different scanning methods in patients with rheumatoid arthritis

George A. W. Bruyn[1], Ingrid Möller[2], Jesus Garrido[3], David Bong[4], Maria-Antonietta d'Agostino[5], Annamaria Iagnocco[6], Zunaid Karim[7], Lene Terslev[8], Nanno Swen[9], Peter Balint[10], Paul Baudoin[1], Dick Siewertsz van Reesema[11], Carlos Pineda[12], Richard J. Wakefield[13] and Esperanza Naredo[14]

Abstract

Objective. To assess the intra- and interobserver reliability of musculoskeletal ultrasonography (US) in detecting inflammatory and destructive tendon abnormalities in patients with RA using two different scanning methods.

Methods. Thirteen observers examined nine patients with RA and one healthy individual in two rounds independently and blindly of each other. Each round consisted of two consecutive examinations, an anatomy-based examination and a free examination according to personal preferences. The following tendons were evaluated: wrist extensor compartments 2, 4 and 6, finger flexor tendons 3 and 4 at MCP level, tibialis posterior tendon and both peronei tendons. Overall, positive and negative agreements and κ-values for greyscale (GS) tenosynovitis, peritendinous power Doppler (PPD) signal, intratendinous power Doppler (IPD) signal and GS tendon damage were calculated.

Results. Intraobserver κ-value ranges were 0.53–0.55 ($P < 0.0005$) for GS tenosynovitis, 0.61–0.64 ($P < 0.0005$) for PPD signal, 0.65–0.66 ($P < 0.0005$) for IPD signal and 0.44–0.53 ($P < 0.0005$) for GS tendon damage. For interobserver reliability, substantial overall agreement ranged from 80 to 89% for GS tenosynovitis, 97 to 100% for PPD signal, 97 to 100% for IPD signal and 97 to 100% for GS tendon damage. Results were independent of scanning technique.

Conclusion. Intraobserver reliability for tenosynovitis and tendon damage varied from moderate for GS to good for PD. Overall interobserver reliability for tenosynovitis and tendon damage was excellent both for GS and PD. This qualitative scoring system may serve as the first step to a semi-quantitative score for tendon pathology.

Key words: ultrasonography, rheumatoid arthritis, tenosynovitis, reliability testing, adult rheumatology.

[1]Department of Rheumatology, MC Groep, Lelystad, The Netherlands, [2]Department of Rheumatology, Instituto Poal de Reumatologia, Barcelona, [3]Department of Methodology of Psychology, Universidad Autonoma de Madrid, [4]Instituto Poal de Reumatologia, Barcelona, Spain, [5]Department of Rheumatology, Hopital Ambroise Pare, Paris, France, [6]Dipartimento di Medicina Interna e Specialità Mediche: Reumatologia, Sapienza Università di Roma, Rome, Italy, [7]Department of Rheumatology, Mid Yorkshire NHS Trust, Wakefield, UK, [8]Department of Rheumatology, Copenhagen University Hospital at Glostrup, Copenhagen, Denmark, [9]Department of Rheumatology, Medisch Centrum Alkmaar, The Netherlands, [10]Department of Rheumatology, National Institute of Rheumatology and Rehabilatation, Budapest, Hungary, [11]Department of Rheumatology, Deventer Ziekenhuizen, Deventer, The Netherlands, [12]National Institute of Rehabilitation, Mexico City, Mexico, [13]Department of Rheumatology, Chapel Allerton Hospital, Leeds, UK and [14]Department of Rheumatology, Hospital Universitario Severo Ochoa, Madrid, Spain.

Submitted 29 December 2011; revised version accepted 26 March 2012.

Correspondence to: George A. W. Bruyn, Department of Rheumatology, MC Groep, 8300 AA Lelystad, The Netherlands.
E-mail: gawbruyn@wxs.nl

Introduction

Tendon involvement may be the first sign of inflammatory arthritis, as was shown in a murine model [1] and in patients with early RA [2]. In RA, synovial inflammation may lead not only to tenosynovitis and functional impairment, but also to tendon rupture [3]. Taking the severity of morbidity and the serious complications of tendon pathology for patients with RA into account, assessment of tendon inflammation in an early stage is a key issue for prevention of irreversible damage and has been given priority by the OMERACT US task force. Ultrasonography (US) is an imaging modality that is now widely accepted in both rheumatology research and clinical practice to visualize joints and soft tissues in patients with various rheumatic diseases. US is able to image not only the damage of cartilage and bone, but also to identify tendon pathology and tendon sheath inflammation. Patients with RA are likely to have tendon disease if US is abnormal [4–7].

Despite the increasing use of US, the technique is regarded as examiner dependent. Results depend not only on the interpreting of the US images, but also on the scanning technique. This perception is compounded by a lack of data regarding its validity, reproducibility and responsiveness to change [8–10].

The aim of the present study was to assess the intra- and inter-reader reliability of US detection of tendon disease in RA patients among rheumatologists with experience in musculoskeletal US. In addition, we compared two different scanning techniques, one using fixed anatomical locations and one using a so-called free-hand examination.

Patients and methods

Patient selection

Nine consecutive patients with RA and one healthy person were selected from the outpatient rheumatology department of the MC Groep, Lelystad, The Netherlands. These included five women and four men, with a median age of 42 years (range 27–63 years); disease stage varied from early to established. One healthy person was also included, with the objective that the cohort include all types of tendons, either pathological or normal, as is seen in daily clinical practice. The demographic features are listed in Table 1. All patients had RA according to the ACR 1987 criteria for RA. All patients gave informed consent and the study was approved by the Ethical Review Board of MC IJsselmeerziekenhuizen. All patients were investigated twice, i.e. the procedure was repeated during the afternoon session, with rearrangement of the patients in a different order and on a different location.

Observers

Observers consisted of a group of 13 rheumatologists from seven countries with variable expertise (median experience 10 years, range 3–16 years) in musculoskeletal US. All were members of the OMERACT US group. The observers met for two consecutive days to perform the investigation. The sonographers performed the US investigation independently from each other. The observers were blinded to the clinical details. All investigators met for a training session the day before the exercise. Before the investigation, members of the OMERACT US group were invited to send images representative of tendon pathology to the principal investigators (G.A.W.B., E.N.). One hundred images were collected, reviewed and commented during the training session. During the training session, the observers also reviewed the definitions for tendon pathologies and standardized the scanning method.

Table 1 Patients'[a] disease characteristics

Demographic	Median (min–max)	Frequency
Age, years	42 (27–63)	
Sex (M/F)		3/6
Disease duration, months	60 (2–120)	
IgM RF$^+$		3
CCP$^+$		3
ESR, mm/h	28 (14–43)	
CRP, mg/l	16 (4–28)	
DAS-28	2.9 (1.1–5.2)	
HAQ	0.7 (0.2–1.0)	
Current DMARD therapy		
MTX		8
LEF		1
Current biologic therapy		
Tociluzimab		1
Infliximab		2
Golimumab		1
Adalimumab		1

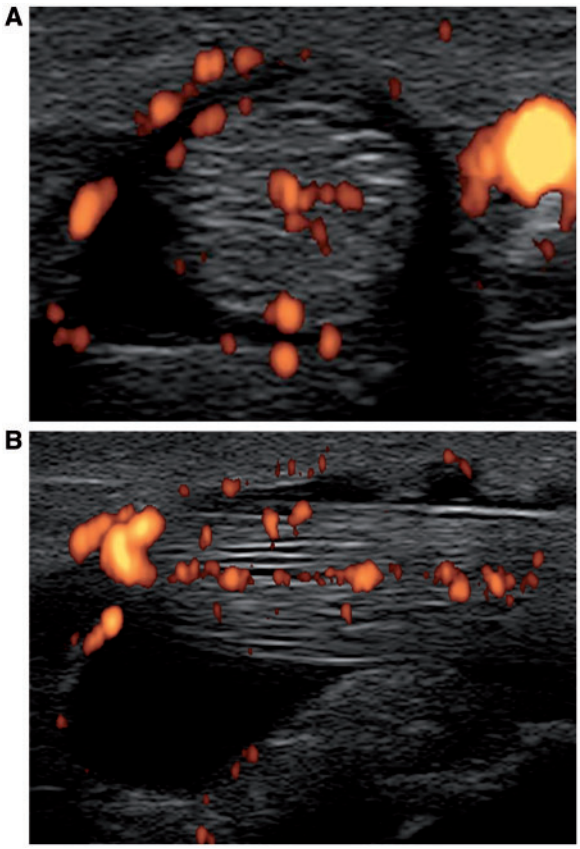
[a]Nine patients with RA and one healthy person were examined. ESR: erythrocyte sedimentation rate.

US imaging

Ultrasonography was performed with five Esaote MyLab Twice (Genoa, Italy) machines with a 6–18 MHz linear array transducer. All five machines had identical settings concerning greyscale (GS) gain and Doppler gain, which were not allowed to change. The frequency of the GS US examination was set at 14 MHz, the frequency of the power Doppler (PD) examination was 12.5 MHz. The US investigation consisted of a dichotomous assessment (presence/absence) in GS and with the PD technique of tendon inflammation (i.e. tenosynovitis) and tendon damage (i.e. partial or complete tear) in both longitudinal and transverse planes.

The following definitions were used for tenosynovitis and tendon tears. The OMERACT definition for tenosynovitis was used: hypoechoic or anechoic thickened tissue within the tendon sheath which is seen in two perpendicular planes and which may exhibit Doppler signal (Fig. 1). As to tendon tears, we defined partial thickness tear as an interruption of tendon fibres with or without hypoechoic material filling the defect, and

FIG. 1 US images of flexor tenosynovitis showing PD signal.



Transverse (**A**) and longitudinal (**B**) US scans of the third finger flexor tendon at the MCP level, showing GS tenosynovitis and a positive PD signal.

full-thickness tear interruption of the tendon fibres with or without hypoechoic materal filling the defect [11]. No ultrasonographic distinction was made between effusions and synovial hypertrophy, and these abnormalities were taken together for the analyses.

PD assessment of selected tendons was carried out with settings standardized to a pulse repetition frequency of 400–500 Hz and low wall filtres. The PD gain was adjusted to a level just below the disappearance of colour signs under the bony cortex [12].

Anatomical criteria for tendon selection and scanning location

To select specific anatomic regions at the hand and ankle with the intent to obtain maximal diagnostic capability and reliability, a literature review was performed [13–18]. Subsequently, five hand–wrist and five foot–ankle previously frozen unembalmed human specimens (Department of Anatomy Bellvitge Campus, University of Barcelona) were scanned with a real-time US machine (Logiq E9, General Electric, Kyunnggi-do, Korea), using a 6–15 MHz linear array transducer. Subsequently the following criteria for tendon selection were developed: (i) presence of a synovial sheath; (ii) minimal anatomical variations; (iii) absence of sesamoid bones that could potentially create an acoustic shadow; (iv) minimal tendon trajectory angulation in the region of interest to minimize anisotropy; (v) avoidance of regions of retinacular compression that would minimize the detection of fluid in early disease such as the region of the A1 annular pulley of the finger flexor tendons. Finally, we selected the following tendons for the study: wrist extensor compartment 2, 4 and 6, finger flexor tendons 3 and 4 at MCP level, tibialis posterior tendon at the medial malleolus and both peroneus tendons at the lateral malleolus (Table 2).

US scanning method

Each US investigation round included two consecutive scanning methods. Each sonographer was given 10 min time per scanning method, divided into 5 min for the wrist/hand and 5 min for the ankle, to examine one patient. First, the tendons were examined at the set anatomic locations described in Table 2 and in two perpendicular planes. After finishing the first anatomy-based examination and recording the findings on the sheet, the investigators carried out a second examination of the same patient using a free scanning method. The free-hand examination was defined as the daily clinical practice US examination according to the examiner's own personal preferences. The free scanning started at the same bony landmarks established for the anatomy-based scanning but allowed the examiner to assess the whole synovial sheath-covered area of the tendons.

Statistical analysis

Demographic data are shown as median (min–max) for quantitative variables or frequency for qualitative variables (Table 1). Prevalence of pathological findings are derivates from positive and negative scores given by the observers. Intraobserver concordance was assessed by Cohen's kappa index (κ). Since Cohen's κ is artificially low in case of high or low prevalence of an abnormality, low κ-values may be obtained even in the presence of high agreement [19–21]. To improve understanding of the results, omnibus κ-values are accompanied with proportions of overall agreement and positive and negative agreement [22]. Interobserver concordance was assessed by Light's κ [17], that is, the average of the κ-values obtained for the m $(m-1)/2$ different pairs of observers between m observers. In order to show the proportion of positive and negative agreement, we defined interobserver substantial agreement if 2/3 of observers gave the same score, i.e. at least 9 out of 13 observers. Agreement indexes were interpreted as follows: 0.81–1.00, excellent agreement; 0.61–0.80, good agreement; 0.41–0.60, moderate agreement; 0.21–0.40, fair agreement; 0.00–0.20, slight agreement and <0.00, poor agreement.

TABLE 2 Selected tendons and probe position

Tendon	Bony landmarks	Background	Probe position	Background
Flexor digitorum superficialis and profundus	Proximal to the metacarpal head of digit 3 and 4	Common tendon sheath; lack of sesamoids; avoids more distal vincular vessels	Transverse; proximal to A1 pulley or between A1 and A2 pulleys	Beneath pulleys synovial sheath and accompanying vasculature is attenuated
Extensor carpi radialis brevis and longus	Lister's tubercle	Common tendon sheath; no crossover	Transverse; at Lister's tubercle or slightly proximal	Easily localized for standardization; common sheath; no crossovers at this point
Extensor digitorum communis (EDC) and extensor indicis proprius	Lister's tubercle	Typical tendon sheath with no major vessels; few confounding variants of EDC	Transverse at Lister's tubercle	Easily localized for standardization
Extensor carpi ulnaris (ECU)	Ulnar styloid groove	Although compartment formed by subsheath of ECU, sheath itself is characteristic	Transverse at ulnar groove	Easily localized for standardization
Tibialis posterior	Proximal to the retromalleolar region	Characteristic sheath with vinculae but retromalleolar region relatively avascular	Transverse proximal and distal to medial malleolus	Easily localized for standardization; avoid retromalleolar region due to relative avascularity and fibrocartilaginous content of tendon
Peroneus longus and brevis	Proximal to the lateral malleolus	Common tendon sheath; tendons fibrocartilaginous with relative avascular region at malleolus	Transverse just proximal to lateral malleolus	Easily localized and proximal to relative avascular and fibrocartilaginous region

Results

Patient characteristics

Characteristics of the nine patients are shown in Table 1. Patients had long-standing disease but mild disease activity. All patients were receiving at least one DMARD. Biologics were administered in five of nine patients. None of the patients were on concurrent steroid therapy. Of the patients, 66% had a DAS-28 <2.6. The healthy person was female and was 34 years of age.

Prevalence

Theoretically, all findings had to be assessed 2275 times by the 13 observers, but some sheets were incompletely filled in, resulting in missing data. GS tenosynovitis was assessed 2097 times, signal-PD in tenosynovitis was assessed 2029 times, intratendinous-PD was assessed 2007 times and tendon damage was assessed 2000 times. Findings were scored as pathological 427 times for tenosynovitis, 167 times for signal-PD in tenosynovitis, 96 times for intratendinous-PD and 136 times for tendon damage. Considering these numbers as an approach to the true prevalence of pathological findings in the sample, prevalence of tenosynovitis was 20%, whereas prevalences for PPD signal, IPD signal and tendon damage were 8, 5 and 7%, respectively. These low prevalences prompt a critical appraisal of κ-values and underscore the need for overall proportion of agreement as well as proportions of positive and negative agreement.

Reproducibility

Table 3 lists the intraobserver proportion of overall agreement, positive and negative agreement as well as the corresponding κ-values and their statistical significance. Excellent overall intraobserver agreement was observed for GS tenosynovitis, PPD signal, IPD signal as well as partial and complete tendon tear. Positive agreement was found in two-thirds of the observations for GS tenosynovitis and PPD signal, whereas it was found in $\sim$50% of all observations for tendon damage. The proportion of negative agreement was $>$90% in all the cases. Nevertheless, the κ-values for intraobserver reproducibility for tenosynovitis were moderate (0.53–0.55), as were the κ-values for tendon damage (0.44–0.53). The κ-values for PPD signal were good (0.61–0.64), as were those for IPD signal (0.66–0.66). Proportions of overall, positive and negative agreement were almost identical for the anatomy-based assessment (between round 1 and round 2) and the free assessment (between round 3 and round 4).

Table 4 lists the proportion of substantial agreement (9 concordant scores out of 13 observers) and proportion of positive and negative agreement derivatives. Light's κ-values for interobserver agreement and their statistical significance are shown. The proportion of substantial agreement was high, $\geqslant$80%, in all cases. This high level of substantial agreement was obtained mainly from negative agreement that was always $>$90%. Positive agreement ranged from 100%, perfect positive agreement, to

Table 3 Intraobserver agreement of US of tenosynovitis in RA

Pathological finding	Rounds	Overall agreement, mean, %[a]	Positive agreement proportion, %	Negative agreement proportion, %	Cohen's κ, mean[a]	*P*
Tenosynovitis	R1R2	87	63	92	0.55	<0.0005
	R3R4	86	61	92	0.53	<0.0005
PD tenosynovitis	R1R2	95	63	98	0.61	<0.0005
	R3R4	95	67	98	0.64	<0.0005
IPD	R1R2	98	67	99	0.66	<0.0005
	R3R4	98	67	99	0.65	<0.0005
Tendon damage	R1R2	95	56	98	0.53	<0.0005
	R3R4	95	47	98	0.44	<0.0005

[a]Mean of 13 intraobserver values, one for each observer; R1: first anatomy-based round; R2: second anatomy-based round; R3: first free examination round; R4: second free examination round.

Table 4 Interobserver agreement for US of tendon disease in RA

Pathological finding	Rounds	Substantial agreement proportion, %	Positive agreement proportion, %	Negative agreement proportion, %	Light's κ	*P*
Tenosynovitis	R1	87	40	93	0.25	<0.0005
	R2	89	44	94	0.32	<0.0005
	R3	80	22	89	0.27	<0.0005
	R4	80	22	89	0.27	<0.0005
PD tenosynovitis	R1	97	67	99	0.29	<0.0005
	R2	100	100	100	0.47	<0.0005
	R3	97	67	99	0.36	<0.0005
	R4	97	67	99	0.53	<0.0005
IPD	R1	99	67	99	0.35	<0.0005
	R2	98	0	99	0.32	<0.0005
	R3	97	67	99	0.36	<0.0005
	R4	100	100	100	0.59	<0.0005
Tendon damage	R1	97	0[a]	99	0.11	<0.0005
	R2	98	0[a]	99	0.31	<0.0005
	R3	100	–[b]	100	0.15	<0.0005
	R4	100	–[b]	100	0.06	0.222

[a]Substantial agreement is declared if there are at least nine equal scores out of 13 observers; 0 means that there were positive findings but <9 out of 13. [b]No positive findings observed.

0%, no positive agreement at all. Interobserver κ-values ranged from slight to moderate, although κ-values were in all cases, with the exception of tendon damage in round 4, statistically significant. Again, there was a lack of influence of the scanning method (i.e. anatomy-based or free) on interobserver agreement.

Discussion

In this study we evaluated the intra- and interobserver agreement of GS and PD US among rheumatologists–ultrasonographers for tendon pathology in patients with RA. The reason for performing this exercise was based on research suggestions made by the OMERACT US task force, where tenosynovitis was indicated as one of the future research areas. We assessed both tenosynovitis and tendon damage of specific flexor and extensor tendons at three predilection areas in RA, i.e. wrist, MCPs and ankle. The specific tendons were selected on the basis of anatomical considerations. As a first step towards a semi-quantitative scoring system, we chose a simple qualitative presence/absence scoring system.

Our study shows that overall intraobserver reproducibility for tenosynovitis in RA was excellent (86–7%). Furthermore, our results show that overall intraobserver reproducibility for tendon damage is also excellent (95%). The overall agreement is much higher than one would expect from the calculated κ-value, because Cohen's κ is artificially low due to the low prevalence of 20% of tenosynovitis and the even lower presence of

other pathological US findings. Low prevalences of lesions probably has lowered Cohen's κ-values, a statistical phenomenon known as the paradox of high agreement but low κ [19, 20]. According to guidelines set out by Ciccheti and Feinstein [20], we used omnibus values of κ as well as the overall proportion of agreement and proportions of positive and negative agreement, which facilitates the understanding of how positive and negative agreement separately contribute to observed overall agreement.

Overall interobserver reproducibility for tenosynovitis also showed excellent agreement, although variation among the examiners was higher (80–89%) than for the intraobserver reliability exercise. Although the overall interobserver agreement was excellent, the κ-values were substantially lower than those for the intraobserver reliability exercise. Besides the previously mentioned reason of the paradox of high agreement and low κ, the κ-statistic also depends on the degree of discordance of the marginal distributions between observers, i.e. κ is artificially low when one observer yields systematically higher or lower evaluations than the other in a pairwise comparison. This was the case in our study.

Few studies have addressed the intra- and interobserver reliability of US in RA patients [4, 23, 24]. Our results are in accordance with these studies. A limitation of most of these studies is the fact that fewer observers carried out the US investigation, making reproducibility difficult to interpret. In addition, we chose to compare two scanning methods, a novel approach reflecting daily clinical practice. The advantage of a scanning method based on fixed anatomical points is that it can easily be standardized; however, the disadvantage is that it may not reflect daily clinical practice. Interestingly, we did not find any difference between the two as to reproducibility. A caveat must be made, as it is not known whether differences will be seen when more positive findings, i.e. tendon abnormalities, are present.

There are limitations to this study. The low prevalence of positive lesions came somewhat as a surprise, as most of the patients had active RA when they were invited to participate in the study a few weeks earlier. However, the fact that disease activity subsides is probably inevitable in this era of biologic therapy. Another limitation of our study is the small number of patients. Since there were five available machines and 13 sonographers, the experiment was set up in such a way that as much information as possible would be obtained in a single working day. A larger sample including both a homogeneous cohort of healthy persons and a cohort of RA patients with a higher prevalence of pathological findings and with various degrees of tendon disease could yield more accurate information regarding the index of agreement between observers and test–retest reliability, but this would require a longer experiment. A third limitation is the extent of previous training. Although extensive prestudy work including the collection of 100 US scans of tendon pathology was done, a longer training session to secure a broad consensus regarding pathology in advance might be necessary for the next step towards a semi-quantitative scoring system.

In conclusion, this analysis of reproducibility of US in detecting tenosynovitis and structural tendon damage in patients with RA suggests that a semi-quantitative scoring system for US is a feasible objective in the future. For the next step, careful consideration has to be paid to the amount of training and standardization, a larger group of healthy persons and patients and the presence of tendon disease in patients. A Delphi process to address the issue of standardization is now under way as part of the OMERACT process.

Rheumatology key messages

- Intra- and inter-observer reliability of US examination of tenosynovitis in patients with RA is generally good.
- US examination of tenosynovitis in RA does not differ between standard or free examination.

Acknowledgements

The authors are grateful for the valuable help of Ria de Kort and Marian de Waal, Rheumatology Research Nurses. The authors are also grateful to all the colleagues participating in the OMERACT Ultrasound Task Group, who sent US images of tendon disease in RA for review.

Funding: The study was made possible by an unrestricted scientific grant from Roche Pharmaceuticals BV. The five US machines were made available for the study by Esaote Netherlands BV.

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

References

1 Hayer S, Redlich K, Korb A *et al*. Tenosynovitis and osteoclast formation as the initial preclinical changes in a murine model of inflammatory arthritis. Arthritis Rheum 2007;56:79–88.

2 Eshed I, Feist E, Althoff CE *et al*. Tenosynovitis of the flexor tendons of the hand detected by MRI: an early indicator of rheumatoid arthritis. Rheumatology 2009;48:887–91.

3 Swen WA, Jacobs JW, Hubach PC *et al*. Comparison of sonography and magnetic resonance imaging for the diagnosis of partial tears of finger extensor tendons in rheumatoid arthritis. Rheumatology 2000;39:55–62.

4 Wakefield RJ, O Connor PJ, Conaghan PG *et al*. Finger tendon disease in untreated rheumatoid arthritis: a comparison of ultrasound and magnetic resonance imaging. Arthritis Rheum 2007;57:1158–64.

5 Martinoli C, Derchi LE, Pastorino C *et al*. Analysis of echotexture of tendons with US. Radiology 1993;186:839–43.

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

7 Grassi W, Tittarelli E, Blasetti P, Pirani O, Cervini C. Finger tendon involvement in rheumatoid arthritis. Evaluation with high-frequency sonography. Arthritis Rheum 1995;38:186–94.

8 Joshua F, Lassere M, Bruyn GA *et al*. Summary findings of a systematic review of the ultrasound assessment of synovitis. J Rheumatol 2007;34:839–44.

9 Bruyn GA, 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.

10 Bruyn GA, Naredo E, Moller I *et al*. Reliability of ultrasonography in detecting shoulder disease in patients with rheumatoid arthritis. Ann Rheum Dis 2009;68:357–61.

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

12 Torp-Pedersen ST, Terslev L. Settings and artefacts relevant in colour/power Doppler ultrasound in rheumatology. Ann Rheum Dis 2008;67:143–9.

13 Ranade A, Rai R, Prabhu L *et al*. Incidence of extensor digitorum brevis manus muscle. Hand 2008;3:320–3.

14 Benjamin M, Kaiser E, Milz S. Structure-function relationships in tendons: a review. J Anat 2008;212:211–28.

15 Hauger O, Chung C, Lektrakul N *et al*. Pulley system in the fingers: normal anatomy and simulated lesions in cadavers at MR imaging, CT, and US with and without contrast material detention of the tendon sheath. Radiol 2000;217:201–2.

16 Stewart DA, Smitham PJ, Gianoutsos MP *et al*. Biomechanical influence of the vincula tendinum on digital motion after isolated flexor tendon injury: a cadaveric study. J Hand Surg 2007;32:1190–4.

17 Petersen W, Bobka T, Stein V, Tillmann B. Blood supply of the peroneal tendons: injection and immunohistochemical studies of cadaver tendons. Acta Ortho Scand 2000;71:168–74.

18 Petersen W, Hohmann G, Stein V, Tillmann B. The blood supply of the posterior tibial tendon. J Bone Joint Surg 2002;84:141–4.

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

20 Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 1990;43:551–8.

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

22 Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. Psychol Bull 1971;76:365–77.

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

24 Micu MC, Serra S, Fodor D *et al*. Inter-observer reliability of ultrasound detection of tendon abnormalities at the wrist and ankle in patients with rheumatoid arthritis. Rheumatology 2011;50:1120–4.