Clinical and Experimental Rheumatology 2012; 30: 566-569.

The most reliable probe position in the ultrasonographic examination of the wrist in rheumatoid arthritis

N. Dopazo González[1, 3], D.F. Ten Cate[1], W.A.A. Swen[2], A. Mera Varela[3], S.A. Insua Vilariño[3], E. Perez-Pampin[3], J.J. Gómez-Reino[3], J.J. Luime[1]

[1]Rheumatology, Erasmus University Medical Centre, Rotterdam, The Netherlands;
[2]Rheumatology, Medical Centre Alkmaar, the Netherlands;
[3]Rheumatology, University Clinical Hospital of Santiago de Compostela, Spain.

Noelia Dopazo González
David F. Ten Cate, MD
Wijnand A.A. Swen, MD PhD
Antonio Mera Varela, MD
Santos A. Insua Vilariño, MD
Eva Perez-Pampin, MD
Juan J. Gómez-Reino, MD, PhD
Jolanda J. Luime, PhD

Please address correspondence and reprint requests to:
D.F. Ten Cate,
Department of Rheumatology,
Erasmus MC, Room Ee-963,
P.O. Box 2040,
3000CA Rotterdam, The Nertherlands.
E-mail: d.tencate@erasmusmc.nl

Received on July 5, 2011; accepted in revised form on December 12, 2011.

Key words: ultrasound, wrist, reliability, rheumatoid arthritis, intraclass correlation coefficient, agreement

Competing interests: none declared.

ABSTRACT

Objective. *This study aims to evaluate the inter-observer reliability of the ultrasonographic examination of the wrist in RA patients between 3 examiners and 3 probe positions.*

Methods. *Fifty-three RA patients were recruited at the University Clinical Hospital of Santiago de Compostela in Spain for ultrasonographic examination of the wrist. Ultrasonography (US) was performed on both wrists using a GE LOGIQ 9 machine, using three probe positions: Lister's Tubercle to digit II (position 1), Lister's Tubercle to digit III (position 2) and ulnocarpal (position 3), from the anatomic medial orientation. Three examiners (2 experienced ultrasonographers and 1 junior ultrasonographer) scored synovitis according to a 0–3 semiquantitative scoring system. Inter-observer reliability was expressed using the ICC (A,1).*

Results. *For grey-scale ultrasound (GSUS) the inter-observer reliability (ICC(A,1)) (single measure, agreement definition) ranged from 0.35 for the ulnocarpal joint, position 3, to 0.60 in both position 1 for the radiocarpal joint and position 2 for the inter-carpal joint. Using power Doppler ultrasound (PDUS) the inter-observer reliability (ICC(A,1)) ranged from 0.36 in position 3, to 0.52 both in position 1 and 2 for the radiocarpal joint.*

Conclusion. *The reliability of the GSUS-examination of the wrist joints of RA patients with GSUS shows highest, moderate reliability using the anatomical landmarks Tubercle of Lister and digit III (position 2). The reliability of the PDUS examination was similar and moderate in both position 1 (Lister's Tubercle to digit II) and position 2 (Lister's Tubercle to digit III). The reliability was poorest for position 3 (the anatomic medial view of the ulnocarpal joint) in both the GSUS and PDUS examination. This study suggests that position 2 should be used in clinical trials and daily practice.*

Introduction

Rheumatoid arthritis (RA) is a chronic and progressive inflammatory disorder primarily affecting synovium, cartilage and bone. Early intervention and tight disease control is crucial, improving the prognosis of RA substantially (1-2). Further improvement might be expected by early detection of synovitis and erosions, using modern imaging techniques such as ultrasonography (US) (3-4) and magnetic resonance imaging (MRI) (5).

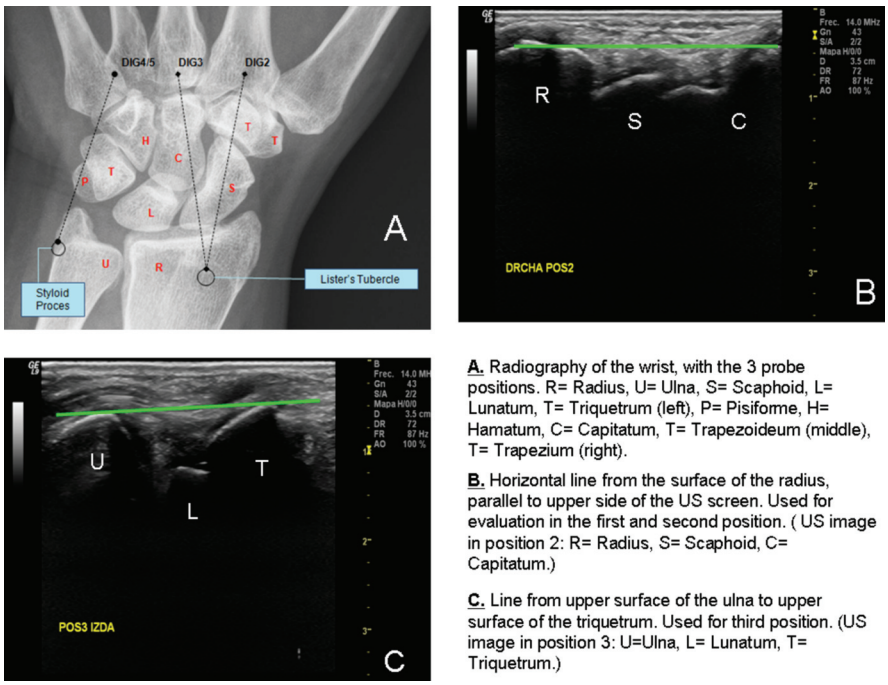
US is a growing imaging modality with great potential as an outcome measure in RA. But at the same time, US is perceived as imperfect and operator dependent (3-6). The operator dependency and the lack of clear criteria for interpretation of US images cause challenges in the interpretation and acquisition of US images. The wrist is one of the most difficult joints of the body to assess ultrasonographically (7). For this joint, MRI outperforms US (8-9). The complex anatomy of the wrist, the layer of soft tissue between the probe and the joint, minimal tissue contrast and lack of a reproducible window for visualising these structures complicates imaging of the wrist (7). Because of these difficulties, reliability is an issue. However previous studies have usually shown reliability data on a combination of joints and not for the wrist only (10). They also did not describe distinct landmarks for probe positioning, leaving the exact location where pathology was evaluated open for discussion. These issues, in combination with a lack of consensus of probe positioning, has led to little and equivocal data on intra- and inter-observer reliability of US scanning of the wrist (7, 10-11). This issue is recognised by the OMERACT (4).

Using distinct anatomical landmarks may improve reliable probe positioning, providing more consistent results between rheumatologists when evaluating the wrist. We therefore evaluated the inter-observer reliability of 3 different ultrasonographic probe positions of the dorsal and anatomic medial side of the wrist of RA patients using Grey-scale Ultrasound (GSUS) and Power Doppler Ultrasound (PDUS) by 3 examiners.

Patients and methods

Study design

A cross-sectional study, approved by the local ethics committee, was conducted at the University Clinical Hos-



A. Radiography of the wrist, with the 3 probe positions. R= Radius, U= Ulna, S= Scaphoid, L= Lunatum, T= Triquetrum (left), P= Pisiforme, H= Hamatum, C= Capitatum, T= Trapezoideum (middle), T= Trapezium (right).

B. Horizontal line from the surface of the radius, parallel to upper side of the US screen. Used for evaluation in the first and second position. (US image in position 2: R= Radius, S= Scaphoid, C= Capitatum.)

C. Line from upper surface of the ulna to upper surface of the triquetrum. Used for third position. (US image in position 3: U=Ulna, L= Lunatum, T= Triquetrum.)

Fig 1.

pital of Santiago de Compostela, Spain. Patients were invited if diagnosed with RA-ACR1987, aged over 18 years and attending the outpatient rheumatology clinic, or day clinic between May 2010 and July 2010. Subjects' written informed consent was obtained according to the Declaration of Helsinki before participation.

Exclusion criteria were a history of fractured wrist bones or ulna/radius, an intra-articular injection with glucocorticosteroids in the wrist <6 months ago and/or malformation of the wrist inhibiting correct probe positioning. No requirements were made regarding the presence of clinical wrist synovitis.

US examination

Both wrists of the patients were scanned from the dorsal side using three probe positions. The positions were chosen because they are defined by anatomically distinct landmarks (Fig. 1a).

- Position 1: From Lister's Tubercle to the proximal end of the second metacarpal bone (DIG II); imaging the radiocarpal joint (RCJ) and intercarpal joint (ICJ).
- Position 2: From Lister's Tubercle to the proximal end of the third metacarpal bone (DIG III); imaging the RCJ and the ICJ.
- Position 3: From the anatomic medial side of the ulnar styloid process to the region between the proximal end of the fourth (DIG IV) and fifth metacarpal bone (DIG V); imaging the ulnocarpal joint (UCJ). (Note that the position of the hand should be different for position 3 than presented in the figure))

The US scans were made using a GE LOGIQ 9 machine with a 5–13MHz multi-frequency linear array transducer. Patients were set in front of the examiner, with the palm of the hand placed flat on the table in a neutral position for the first two positions. For the third position the hand was placed in maximum abduction, seen from the anatomical position, ensuring good probe positioning. All patients were examined in the same room with a constant temperature. Evaluation of one patient took ~10 minutes including documentation.

Synovitis was scored real-time, using a semiquantitative scoring system (grade 0=absence, grade 1=mild, grade 2=moderate, and grade 3=severe synovitis), using a horizontal line from the upper surface of the distal radius parallel to the upper side of US screen (Fig. 1b). For position 3 this line ran from the upper surface of the ulna to the upper surface of the triquetrum (Fig. 1c). A distension of the joint capsule beneath this horizontal line was graded 1, at or above this line was graded 2 and bulging above the line was graded 3.

For Power Doppler (PD) we used the following settings: colour gain at disappearance of colour noise, Pulse Repetition Frequency (PRF) at 0.5 kHz, frequency at 7.5 MHz and wall filter (WF) at 83 Hz. We adjusted the size and position of the colour box to include the subcutaneous tissue to recognise artefacts caused by vessels above the joint. The PD signal was evaluated using the semiquantitative PD scoring system of Szkudlarek (3).

Ultrasonographers

Each patient was evaluated by two rheumatologists experienced in musculoskeletal US and one junior doctor trained for this occasion (>70 supervised US-wrist examinations prior to this study). The scans were performed independently, and blinded for clinical examination. Consensus regarding the scoring protocol was reached prior to the study and documentation was standardised per patient.

Additional data collection

Data on age, gender, disease duration, rheumatoid factor (RF), anti- cyclic citrullinated peptide (anti-CCP), Sharp-van der Heijde score, HAQ-score (12), DAS28 and current treatment were collected after US examination. Data on C-reactive protein (CRP) concentration and erythrocyte sedimentation rate (ESR) were obtained using each patient's most recent laboratory results. An independent rheumatologist, blinded to US findings, obtained a DAS28 (13).

Statistical analysis

The sample size was calculated according to Streiner (14). With an expected ICC(A,1) of 0.8, a confidence interval of 0.1 and 3 observations per patient, 46 patients were needed. Descriptive statistics were performed for patient and disease characteristics. Reliability was calculated by the intraclass correlation coefficient (ICC (A,1)) (15), single measure, agreement definition in R (version 2.7.1). The inter-observer reliability ICC (A,1) was determined per

position (1, 2, or 3) and per joint (radiocarpal (RC) and inter-carpal (IC)). The scores for the right and left wrist were combined. This study was conducted following the STROBE statement for observational studies (16).

Results

Patient characteristics (Table I)

The 53 RA patients had a median disease duration of 57 months [range: 0–354 months], a heterogeneous disease activity (median DAS28 of 2.83 [range 0.19–6.41]) and were reasonably self-sufficient (median HAQ of 0.88 [range: 0–2.8]). DAS28 indicated remission in 17 patients (32.1%). Four patients (7.5%) had severe disease activity. Clinically, 32 wrists were tender, of which 24 were also swollen. Four additional wrists were only swollen. 38 patients (71.6%) were treated with DMARDs and 31 (58.5%) with biologicals.

Inter-observer reliability

For GSUS the ICC (A,1) ranged between 0.35 (95%CI:0.23–0.48) in position 3 and 0.60 (95%CI: 0.50–0.70) in the RC joint of position 1 and also 0.60 (95%: 0.50–0.70) in the IC joint of position 2. For PDUS the ICC (A,1) ranged between 0.36 (95%CI: 0.23–0.49) in position 3 and 0.52 (95%CI: 0.41–0.63) in the RC joint of position 1 and 2. Details for each position are presented in Table II.

The inter-observer reliability between the inexperienced ultrasonographer and one of the two experienced ultrasonographers with whom she had trained intensively ranged from 0.66 to 0.74 in GSUS and from 0.62 to 0.72 in PDUS. Whereas these values ranged from 0.39 to 0.40 in GSUS and from 0.20 to 0.47 in PDUS for the inexperienced ultrasonographer and the other experienced ultrasonographer with whom she had trained less.

Table I. Patient disease characteristics.

Characteristics	n=53
Sex, female (%)	40 (75)
Age, years mean (SD)	58 (14)
Disease duration, months median (min – max)	57 (0–354)
CRP, mg/dL median (min – max)	0.40 (0.20–15.5)
ESR, mm median (min – max) (n=50)	14 (1–80)
RF positivity (%)	31 (58)
Sharp- v/d Heijde score, median (min – max) (n=50)	40 (3–186)
HAQ score, median (min – max)	0.88 (0–2.75)
DAS28, median (min – max) (n=50)	2.89 (0.19–6.45)
Current therapy, number (%)	
NSAID	30 (57)
DMARD	40 (75)
Biological	32 (60)

SD: Standard deviation; CRP: C-reactive protein; HAQ: Health Assessment Questionnaire; NSAID: non-steroidal anti-inflammatory drug; DMARD: Disease-modifying anti-rheumatic drug; DAS28: Disease Activity Score of 28 joints.

Table II. Inter-observer reliability ICC (A,1).

	Inter-observer reliability		
	Position 1	Position 2	Position 3
RC (GS)	0.60	0.52	0.35(UC)
IC (GS)	0.44	0.60	–
RC (PD)	0.52	0.52	0.36(UC)
IC (PD)	0.40	0.40	–

Right and left wrist together. Per examiner per joint.
(RC: radiocarpal joint; IC: intercarpal joint; GS: Grey scale; PD: Power Doppler; UC: ulnocarpal).

Discussion

Wrist examination with US is made difficult due to the complex anatomy of the wrist and the amount of anatomical structures to evaluate. Therefore it is important to use distinct anatomical landmarks to describe probe positioning. Previous studies did not use these distinct landmarks, leaving the exact location where pathology was evaluated open for discussion. Using distinct anatomical landmarks improves reliable probe positioning providing consistent results between rheumatologists.

In this study with distinct bony landmarks, easy to locate on ultrasonographic examination, we showed that, when taking into account both radiocarpal and inter-carpal joints, GSUS examination of the wrist for the presence of synovitis seems most reliable in position 2 (Lister's Tubercle to DIG III). For evaluating only the RC-joint, position 1 (Lister's Tubercle to DIG II) performed best. For PDUS, the inter-observer reliability was similar for position 1 and 2. The ulnocarpal orientation (position 3) showed the poorest reliability in both GSUS and PDUS.

Although it is generally recognised that US underperforms in the wrist due to inter-examiner variation, only one study presented data on the reliability of only the wrist resulting in a kappa for PDUS of 0.69 (1). Most US reliability studies provide data for a combination of joints and not for the wrist separately (7, 10-11). In our study PDUS reliability was fair. This may partly be explained by features of the ICC itself. Because point estimates of the ICC depend on the heterogeneity of the patients in the study it is necessary to have sufficient variation in the presence of pathology (17, 18). We selected RA patients with a heterogeneous clinical involvement of the wrist. This seemed not to be sufficient for the PD scores in our study. Many of our patients had a zero score on PD causing too little variation between patients, which is a reason for low ICCs (A,1). One way to cope with this in future studies is to select patients based on their PD status by an independent US examiner not involved in the reliability assessment.

To further improve inter-observer reliability in US wrist examination calibration between ultrasonographers is helpful. This was highlighted by the higher reliability between the two ultrasonographers who had trained most

intensively together in this study. This same principle was described in an important MRI study where calibration exercises were used to improve inter-reader reliability of MRI scoring considerably (19). Another way to improve wrist evaluation is to scan healthy subjects getting familiar with the normal anatomical variation. This is especially true for the ulnocarpal joint in which the complex anatomy causes structures to appear pathologic due to their echogeneity. The moderate inter-observer reliability could be improved somewhat by using the multiplanar method, taking position 1 and 2 together. The ICC (A,1) would rise to 0.66 in both the RC and the IC-joint for GSUS and to 0.62 for PDUS in the RC-joint (data not presented).

In conclusion, US-examination of the wrist joints of RA patients with GSUS shows highest, moderate reliability using the anatomical landmarks Tubercle of Lister and digit III (position 2). The reliability for examination with PDUS is similar in both position 1 (Lister's Tubercle to digit II) and position 2 (Lister's Tubercle to digit III). Evaluation on the ulnar side (position 3) showed poor inter-observer reliability. This study suggests that position 2 should be used in clinical trials and in daily practice.

Acknowledgements

We would like to thank Esperanza Naredo and the Ultrasound School of the Spanish Society of Rheumatology for providing us with a grey-scale scoring system.

References

1. STRUNK J, HEINEMANN E, NEECK G, SCHMIDT KL, LANGE U: A new approach to studying angiogenesis in rheumatoid arthritis by means of power Doppler ultrasonography and measurement of serum vascular endothelial growth factor. *Rheumatology* (Oxford) 2004; 43: 1480-3.
2. COMBE B, LANDEWÉ R, LUKAS C *et al*.: EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis* 2007; 66: 34-45.
3. SZKUDLAREK M, COURT-PAYEN M, JACOBSEN S, KLARLUND M, THOMSEN HS, OSTERGAARD M: Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis. *Arthritis Rheum* 2003; 48: 955-62.
4. WAKEFIELD RJ, D'AGOSTINO MA, IAGNOCCO A *et al*.: The OMERACT Ultrasound Group: status of current activities and research directions. *J Rheumatol* 2007; 34: 848-51.
5. CONAGHAN PG, MCQUEEN FM, PETERFY CG *et al*.: The evidence for magnetic resonance imaging as an outcome measure in proof-of-concept rheumatoid arthritis studies. *J Rheumatol* 2005; 32: 2465-9.
6. WAKEFIELD RJ, BALINT PV, SZKUDLAREK M *et al*.: Musculoskeletal ultrasound including definitions for ultrasonographic pathology. *J Rheumatol* 2005; 32: 2485-7.
7. KOSKI JM, SAARAKKALA S, HELLE M *et al*.: Assessing the intra- and inter-reader reliability of dynamic ultrasound images in power Doppler ultrasonography. *Ann Rheum Dis* 2006; 65: 1658-60.
8. HAAVARDSHOLM EA, OSTERGAARD M, EJBJERG BJ *et al*.: Reliability and sensitivity to change of the OMERACT rheumatoid arthritis magnetic resonance imaging score in a multireader, longitudinal setting. *Arthritis Rheum* 2005; 52: 3860-7.
9. TAOULI B, ZAIM S, PETERFY CG *et al*.: Rheumatoid arthritis of the hand and wrist: comparison of three imaging techniques. *AJR Am J Roentgenol* 2004; 182: 937-43.
10. 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.
11. NAREDO E, BONILLA G, GAMERO F, USON J, CARMONA L, LAFFON A: Assessment of inflammatory activity in rheumatoid arthritis: a comparative study of clinical evaluation with grey scale and power Doppler ultrasonography. *Ann Rheum Dis* 2005; 64: 375-81.
12. SIEGERT CE, VLEMING LJ, VANDENBROUCKE JP, CATS A: Measurement of disability in Dutch rheumatoid arthritis patients. *Clin Rheumatol* 1984; 3: 305-9.
13. PREVOO ML, VAN 'T HOF MA, KUPER HH, VAN LEEUWEN MA, VAN DE PUTTE LB, VAN RIEL PL: Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995; 38: 44-8.
14. STREINER DL NG: Health Measurement Scales. A Practical Guide To Their Development And Use. 4th Revised edition ed.: Oxford University Press; 2008.
15. MCGRAW K, WONG S: Forming inferences about some intraclass correlation coefficients. *Psycological Methods* 1996; 1: 30-46.
16. VON ELM E, ALTMAN DG, EGGER M *et al*.: The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008; 61: 344-9.
17. TEN CATE DF, LUIME JJ, HAZES JM, JACOBS JW, LANDEWÉ R: Does the intraclass correlation coefficient always reliably express reliability? Comment on the article by CHEUNG *et al*. *Arthritis Care Res* (Hoboken) 2010; 62: 1357-1358; author reply 1358.
18. LUKAS C, BRAUN J, VAN DER HEIJDE D *et al*.: Scoring inflammatory activity of the spine by magnetic resonance imaging in ankylosing spondylitis: a multireader experiment. *J Rheumatol* 2007; 34: 862-70.
19. BIRD P, JOSHUA F, LASSERE M, SHNIER R, EDMONDS J: Training and calibration improve inter-reader reliability of joint damage assessment using magnetic resonance image scoring and computerized erosion volume measurement. *J Rheumatol* 2005; 32: 1452-8.