

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

G. A. W. BRUYN,¹ C. PINEDA,² C. HERNANDEZ-DIAZ,² L. VENTURA-RIOS,³ C. MOYA,⁴ J. GARRIDO,⁵ H. GROEN,⁶ A. PENA,² R. ESPINOSA,² I. MÖLLER,⁷ E. FILIPPUCCI,⁸ A. IAGNOCCO,⁹ P. V. BALINT,¹⁰ D. KANE,¹¹ M.-A. D'AGOSTINO,¹² M. ANGULO,² R. PONTE,² J. M. FERNANDEZ-GALLARDO,¹³ AND E. NAREDO¹³

Objective. To assess the intra- and interobserver reproducibility of musculoskeletal ultrasonography (US) in detecting inflammatory shoulder changes in patients with rheumatoid arthritis, and to determine the agreement between US and the Shoulder Pain and Disability Index (SPADI) and the Disabilities of the Arm, Shoulder, and Hand (DASH) questionnaire, using magnetic resonance imaging (MRI) as a gold standard.

Methods. Eleven rheumatologists investigated 10 patients in 2 rounds independently and blindly of each other by US. US results were compared with shoulder function tests and MRI.

Results. The positive and negative predictive values (NPVs) for axillary recess synovitis (ARS) were 0.88 and 0.43, respectively, for posterior recess synovitis (PRS) were 0.36 and 0.97, respectively, for subacromial/subdeltoid bursitis (SASB) were 0.85 and 0.28, respectively, and the NPV for biceps tenosynovitis (BT) was 1.00. The intraobserver kappa was 0.62 for ARS, 0.59 for PRS, 0.51 for BT, and 0.70 for SASB. The intraobserver kappa for power Doppler US (PDUS) signal was 0.91 for PRS, 0.77 for ARS, 0.94 for SASB, and 0.53 for BT. The interobserver maximum kappa was 0.46 for BT, 0.95 for ARS, 0.52 for PRS, and 0.61 for SASB. The interobserver reliability of PDUS was 1.0 for PRS, 0.1 for ARS, 0.5 for BT, and 1.0 for SASB. *P* values for the SPADI and DASH versus cuff tear on US were 0.02 and 0.01, respectively; all other relationships were not significant.

Conclusion. Overall agreements between gray-scale US and MRI regarding synovitis of the shoulder varied considerably, but excellent results were seen for PDUS. Measures of shoulder function have a poor relationship with US and MRI. Improved standardization of US scanning technique could further reliability of shoulder US.

INTRODUCTION

Shoulder involvement is a critical issue in patients with rheumatoid arthritis (RA), with literature data reporting radiographic damage in 50% of patients after 2 years of disease to 96% of patients with 12 years of disease (1–4). Ongoing synovial inflammation is the primary causal event, resulting not only in humeral head erosions, but also in rotator cuff rupture, further compromising shoulder function.

Ultrasonography (US) is an imaging modality now widely available in both scientific research and clinical rheumatology practice for visualizing joints and soft tissues in patients with rheumatic diseases. US is able to image not only the damage to cartilage and cortical bone, but also to identify and quantify tendon pathology and synovial inflammation. Despite increasing efforts on the validation and reliability of US in the evaluation of small joints of the hands and feet, evidence on intra- and inter-

Supported by the Mexican College of Rheumatology. Dr. Bruyn's work was supported by Roche Pharmaceuticals, Basel, Switzerland, and Abbott Laboratories, Abbott Park, Illinois.

¹G. A. W. Bruyn, MD, PhD: Medisch Centrum Leeuwarden, Leeuwarden, The Netherlands; ²C. Pineda, MD, C. Hernandez-Diaz, MD, A. Pena, MD, R. Espinosa, MD, MSc, M.

Angulo, MD, R. Ponte, MD: Instituto Nacional de Rehabilitación, Mexico City, Mexico; ³L. Ventura-Rios, MD: Hospital Central Sur de Alta Especialidad, PEMEX, Mexico City, Mexico; ⁴C. Moya, MD: Private Practice Rheumatology, Mexico City, Mexico; ⁵J. Garrido, PhD: Universidad Autónoma de Madrid, Madrid, Spain; ⁶H. Groen, MD, PhD: University of Groningen, Groningen, The Netherlands; ⁷I.

observer variation of larger joints is still limited (5–9). Taking into account the burden of disability, functional impairment, morbidity, the irreversibility of lesions, and the sequelae of shoulder pathology for patients with RA, the Outcome Measures in Rheumatology Clinical Trials (OMERACT) subtask group on ultrasonographic examination of the shoulder previously examined patients with RA and shoulder disease in order to address the question of whether US is able to detect shoulder disease reliably as compared with magnetic resonance imaging (MRI) (10). After review of these data, our primary goal was set on how to improve earlier findings while primarily focusing on synovitis instead of erosions.

Our secondary goal was to investigate the relationship between clinical measures and US. In RA, on an individual joint basis, there is a poor correlation between clinical signs of synovitis, i.e., joint swelling and tenderness, and US assessment of synovial disease, where US may detect more synovial hypertrophy than palpable and power Doppler US may demonstrate that not every clinically inflamed joint is necessarily hypervascular (11). To our knowledge, these relationships have never been investigated for large joints such as the shoulder; furthermore, it has not been examined whether there is a reliable relationship between US and validated shoulder surveys for adults such as the Disabilities of the Arm, Shoulder, and Hand (DASH) instrument and the Shoulder Pain and Disability Index (SPADI) (12,13).

PATIENTS AND METHODS

Patient selection. Nine RA patients with clinical shoulder involvement were recruited in the outpatient rheumatology clinic of the Instituto Nacional de Rehabilitación of Mexico City, Mexico. These patients consisted of 1 man and 8 women, with a median age of 65 years (range 55–76 years) and a median disease duration of 2 years (range 0–15 years). During the study, all of the patients took medication, including nonsteroidal antiinflammatory drugs and a combination of the following disease-modifying antirheumatic drugs: methotrexate, sulfasalazine, and

hydroxychloroquine. No patient was receiving biologic therapy. One healthy control with no shoulder symptoms was also enrolled. All of the patients had established RA according to the American College of Rheumatology (formerly the American Rheumatism Association) 1987 criteria for RA (14). The study was conducted in accordance with the Declaration of Helsinki and was approved by the local ethic committee. All of the subjects gave informed consent.

Observers. The observers consisted of a group of 11 rheumatologists from 5 countries with variable expertise in musculoskeletal US (median experience 10 years, range 1–17 years). The observers met for 2 days to perform the investigation. The 10 persons included in the study were divided into 2 groups of 5 and were sonographically investigated by all of the observers in 2 rounds, one in the morning and one in the afternoon. The morning procedure was repeated during the afternoon session, with rearrangement of the patients in a different order. All of the observers were blinded to the clinical details and MRI results. All of the investigators met for a training session before the exercise to review the scoring method and for initial training of observers not familiar with some aspects of the scoring system. A statistician was on hand to receive the filled score sheets. The score sheets from the morning session were sealed in envelopes until the second session was concluded.

Clinical and laboratory assessment. All persons, including the healthy control, were clinimetrically evaluated by 2 rheumatologists who were blinded to the patient's details and did not participate in the US examination. The following data were recorded for each patient at study entry: age, sex, disease duration, and presence of rheumatoid factor. The clinimetric evaluation consisted of taking the Disease Activity Score in 28 joints (DAS28), a visual analog scale for pain, and determination of measures of adult shoulder function, including the DASH and SPADI instruments. The DASH contains 30 questions, of which 5 are related to symptoms and 25 are related to functional tasks. The SPADI is a questionnaire containing 5 items scoring pain on a scale of 0–10 and 8 items related to function and disability on a scale of 0–10. Erythrocyte sedimentation rate and C-reactive protein level were also recorded and obtained within 48 hours of the US examination. The disease activity index was estimated by calculating the DAS28 for all of the patients.

US examination of the glenohumeral joint, the subacromial/subdeltoid bursa, the biceps tendon, and the rotator cuff. All of the scans were performed using Siemens Acuson Antares (Siemens) machines with 7.5–13 MHz linear array transducers. The shoulder scoring system assessed elements of inflammation, as well as structural tendinous damage. Rotator cuff tendons were investigated for the presence of total or partial tears in a longitudinal and transverse plane on both static and dynamic maneuvers. The synovial structures of the shoulder, including the subacromial/subdeltoid bursa, the sheath of the long biceps tendon, and the axillary and posterior recess of the

Möller, MD: Instituto Poal de Reumatologia, Barcelona, Spain; ⁸E. Filippucci, MD, PhD: Università della Politecnica della Marche, Ancona, Italy; ⁹A. Iagnocco, MD: Sapienza Università di Roma, Rome, Italy; ¹⁰P. V. Balint, MD, PhD, FRCP: National Institute of Rheumatology and Physiotherapy, Budapest, Hungary; ¹¹D. Kane, PhD, FRCPI: Adelaide and Meath Hospital & Trinity College, Dublin, Ireland; ¹²M.-A. D'Agostino, MD, PhD: Hôpital Ambroise Paré, Paris, France; ¹³J. M. Fernandez-Gallardo, MD, E. Naredo, MD: Hospital Severo Ochoa, Madrid, Spain.

Dr. Garrido has received consultant fees, speaking fees, and/or honoraria (less than \$10,000 each) from Roche Pharmaceuticals and Schering-Plough, and (more than \$10,000) from Abbott Laboratories.

Address correspondence to G. A. W. Bruyn, MD, PhD, Department of Rheumatology, Medisch Centrum Leeuwarden, Leeuwarden, The Netherlands. E-mail: gawbruyn@wx.nl.

Submitted for publication September 24, 2009; accepted in revised form February 23, 2010.

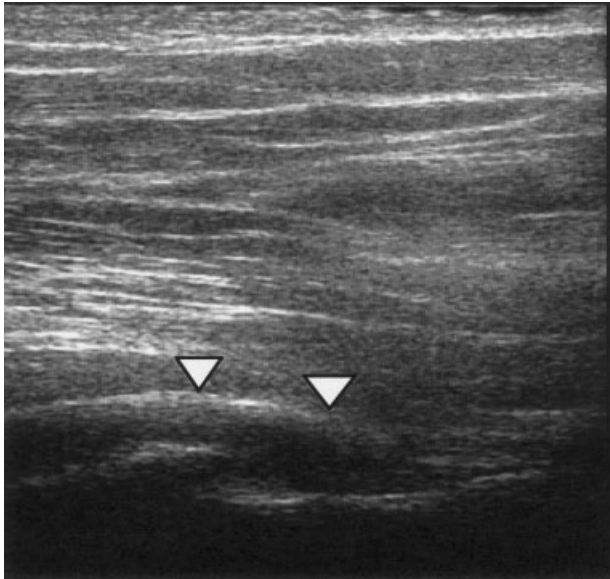


Figure 1. Short axis scan of posterior glenohumeral joint showing anechoic structure corresponding to effusion (arrowheads).

glenohumeral joint, were examined for the presence of effusions and synovial hypertrophy (Figures 1 and 2). Power Doppler assessment of selected synovial sites, including the biceps sheath, the subacromial/subdeltoid bursa, and the axillary and posterior recesses, was carried out with settings standardized to a pulse repetition frequency of 610 Hz, a Doppler frequency of 7.5 MHz, and low wall filters. The power Doppler gain was adjusted to a level just below the disappearance of artifacts under the bony cortex as recommended by Rubin et al (15). The OMERACT US definitions for tenosynovitis, synovitis, synovial hypertrophy, and effusion were applied (16), with minor modifications: a hypoechoic area of at least 3 mm around the long head of the biceps tendon was considered as tenosynovitis of the long biceps tendon; bursal thickness >3 mm or effusion was considered as effusion/synovial hypertrophy of the subacromial/subdeltoid bursa; >3 mm of effusion/synovial hypertrophy at the posterior recess superior to the glenoid labrum was considered as synovitis; and >3 mm of effusion/synovial hypertrophy at the axillary recess was considered as synovitis. No ultrasonographic distinction was made between effusions and synovial hypertrophy, and these abnormalities were taken together for the analyses.

MRI. Assessment of the affected shoulder by MRI took place within 15 working days prior to the US investigation in all of the patients. MRI was performed with a 1.5-T unit (Signa Excite) using a 4-channel shoulder array coil. The following sequences were used: T1-weighted spin-echo sequence (repetition time [TR] 450 msec, echo time [TE] 13.6 msec, slice thickness 3 mm, matrix 256 × 192, and field of view 140–160 mm) in an axial, transverse, and oblique coronal slice orientation parallel to the course of the supraspinatus tendon; and T2-weighted fat-suppressed images in a coronal, axial, and a sagittal plane (TR 3,000 msec, TE 36.4 msec). The MRIs were evaluated by 2 radiologists who were in consensus and had no knowledge of

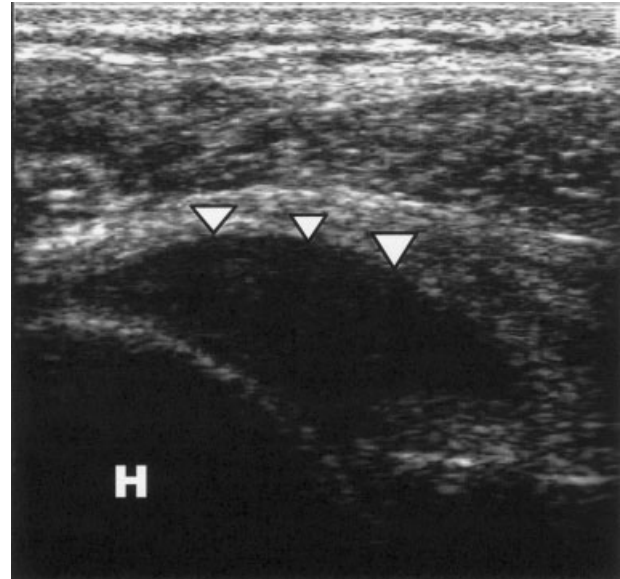


Figure 2. Long axis scan of axillary pouch showing anechoic structure corresponding to effusion as well as showing a spur of humeral head (H; arrowheads).

the results of the US. The MRIs were analyzed for the presence or absence of the same pathologic structures, e.g., fluid in synovial-covered spaces that were visualized by US. The MRI criterion for effusion was an intraarticular or intrabursal area with a high signal on T2-weighted sequences. The criterion for synovitis was enhancing material seen on the fat-suppressed T1-weighted sequences.

Statistical analysis. Overall agreement between US and MRI, defined as the percentage of observed exact agreements, was calculated for each observer. Averaged overall agreement and the kappa index are shown. Since Cohen's kappa is artificially low in case of high or low prevalence (17–20), we have used kappa adjusted by prevalence and bias instead of the kappa standard. Furthermore, the mean positive and negative percentages of agreement were calculated.

Intraobserver reliability is presented as the overall agreement between the first and second round for each scan and the kappa adjusted by prevalence and bias. According to Landis and Koch (21), 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, 0–0.20 = poor agreement, and <0 = no agreement.

Interobserver reliability was studied by calculating the maximum kappa, which is a modification of Cohen's kappa, wherein the maximum possible value of observed agreement is substituted for the value of 1 in the denominator of Cohen's calculation kappa (22). The maximum kappa reported is the mean of values obtained for each pair of observers. Comparison between mean clinical values from patients with findings classified as normal or abnormal by MRI or the majority of US observers was developed with the Student's *t*-test. *P* values less than or equal to 0.05 were considered as significant.

Table 1. Demographic and clinical characteristics of the patients*

	Value
Age, mean (range) years	51 (29–65)
Sex, M/F	1/8
Disease duration, mean (range) months	133 (2–183)
Rheumatoid factor, no.	9
Pain VAS, mm	53 ± 23
DASH	36 ± 25
SPADI	50 ± 28
DAS28	5.1 ± 1.1
ESR, mm/hour	29 ± 11
CRP level, mg/liter (normal value <5)	16 ± 18

* Values are the mean ± SD unless otherwise indicated. VAS = visual analog scale; DASH = Disabilities of the Arm, Shoulder, and Hand questionnaire; SPADI = Shoulder Pain and Disability Index; DAS28 = Disease Activity Score in 28 joints; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein.

RESULTS

The demographic and clinical characteristics of the patients are shown in Table 1. The pathologic shoulder findings on MRI are shown in Table 2. In the majority of patients, a subacromial/subdeltoid bursitis or a rotator cuff tear was found. Axillary or posterior synovitis was found in only a minority of patients. One patient had a biceps

tendon tear. No cases of biceps tenosynovitis were seen on MRI.

Validation. Table 2 summarizes the mean overall agreement and kappa values between MRI and US, pooling the 2 rounds. According to the kappa values, the presence or absence of axillary synovitis and biceps tendon tear was found with good agreement, fair agreement between MRI and US was found for posterior synovitis and bursitis, and moderate agreement for rotator cuff tear and poor agreement was found with respect to biceps tenosynovitis. Regarding glenohumeral joint synovitis, the mean agreement between MRI and US assessment was better for axillary synovitis (80%) than for posterior recess synovitis (66%).

Table 3 lists the specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV; if available) of the US examination for each shoulder abnormality, considering MRI findings as the gold standard.

Reproducibility. Table 4 lists the mean intraobserver agreement and the corresponding mean kappa values. The mean overall agreements for intraobserver reproducibility ranged from moderate to excellent. Excellent intraobserver agreement was observed for tears of the biceps tendon, supraspinatus and infraspinatus tendon, and power Doppler signals regarding the posterior glenohumeral joint and bursa. The mean kappa value for intraobserver reproducibility

Table 2. Agreement between ultrasonography (US) and magnetic resonance imaging (MRI) in 9 patients with rheumatoid arthritis and 1 healthy control

Pathologic finding	No. present according to MRI in 9 patients	Overall agreement between US and MRI, mean %	Kappa adjusted by prevalence and bias, mean	Positive agreement, mean %	Negative agreement, mean %
Axillary synovitis	3	80	0.60	62	86
Posterior synovitis	2	66	0.31	59	69
Subacromial/subdeltoid bursitis	7	61	0.21	67	44
Biceps tenosynovitis	0	52	0.03	–	65
Biceps tendon tear	1	88	0.75	6	93
Rotator cuff tear	6	71	0.42	47	76

Table 3. Sensitivity, specificity, PPV, and NPV of US findings compared with MRI as the gold standard*

	Axillary synovitis†	Posterior synovitis‡	Biceps tenosynovitis§	Biceps tendon tear¶	Subacromial/subdeltoid bursitis#	Rotator cuff tear**
Sensitivity	0.60 (0.53–0.67)	0.93 (0.86–1.00)	NA	0.14 (0.00–0.27)	0.54 (0.39–0.68)	0.92 (0.87–0.98)
Specificity	0.88 (0.78–0.97)	0.49 (0.36–0.61)	0.48 (0.38–0.59)	0.99 (0.96–1.01)	0.61 (0.45–0.78)	0.47 (0.27–0.67)
PPV	0.88 (0.80–0.96)	0.36 (0.31–0.42)	NA	0.75 (0.26–1.24)	0.85 (0.80–0.90)	0.79 (0.73–0.85)
NPV	0.43 (0.41–0.45)	0.97 (0.94–1.00)	1.00 (1.00–1.00)	0.89 (0.87–0.91)	0.28 (0.21–0.34)	0.68 (0.43–0.93)

* Values are the mean (95% confidence interval). PPV = positive predictive value; NPV = negative predictive value; US = ultrasound; MRI = magnetic resonance imaging; NA = not available.

† MRI finding: 3 present, 6 absent.

‡ MRI finding: 2 present, 7 absent.

§ MRI finding: 0 present, 9 absent.

¶ MRI finding: 1 present, 8 absent.

MRI finding: 7 present, 2 absent.

** MRI finding: 6 present, 3 absent.

Table 4. Intraobserver agreement on ultrasound findings in 10 patients with rheumatoid arthritis (including 1 healthy control)

	Pairs of observations	Overall agreement, %	Cohen's adjusted κ^*
Biceps tenosynovitis	110	73	0.51
Biceps tenosynovitis power Doppler	109	74	0.52
Biceps tendon tear	110	95	0.98
Synovitis			
Axillary	92	87	0.62
Posterior	109	77	0.59
Synovial power Doppler			
Axillary	99	94	0.77
Posterior	110	92	0.91
Bursitis subdeltoid/subscapular	110	82	0.70
Bursitis subdeltoid/subscapular power Doppler	110	94	0.94
Rotator cuff tear			
Supraspinatus	110	84	0.76
Infraspinatus	110	89	0.85
Subscapularis	110	70	0.45

* Adjusted for bias and prevalence according to Byrt et al (20).

ibility for bursitis was good (0.70). The mean kappa values for glenohumeral synovitis were moderate (0.59) and good (0.62) for posterior recess and axillary recess synovitis, respectively. The mean kappa values for synovial power Doppler flow in the joint recesses varied from good to excellent, whereas that for the power Doppler flow in the bursa was excellent. According to the kappa values, the intraobserver reproducibility for rotator cuff tears varied from moderate to excellent.

Table 5 lists the maximum kappa values for interobserver agreement. Increasing kappa values signify better agreement between the 11 observers. The mean interobserver kappa value for tenosynovitis of the long biceps tendon was moderate (0.46), as was the kappa value for the power Doppler signal within the tendon sheath. Mean kappa values for glenohumeral joint synovitis ranged from moderate (0.52) to excellent (0.97), with excellent interobserver mean maximum kappa values

for the presence of a power Doppler signal for the posterior joint recess. The interobserver kappa for bursitis was good (0.61), with an excellent mean kappa value for power Doppler signal within the bursa. The kappa values for rotator cuff tear scored a moderate to good agreement.

Clinical evaluation. Table 6 summarizes the mean clinical values between patients showing normal or abnormal MRI findings for axillary recess synovitis, subacromial/subdeltoid bursitis, and biceps tendon sheath tenosynovitis structures. Moreover, mean clinical values for patients showing normal or abnormal US findings for posterior recess synovitis and rotator cuff rupture are listed. For both items, US was classified as normal or abnormal according to the opinion of the majority of US observers. Due to the small size of the groups, meaningful comparisons

Table 5. Interobserver reliability in either the first or second round*

Pathologic finding	Overall agreement, %	Kappa max	SD	P
Biceps tenosynovitis	59	0.46	0.56	< 0.001
Power Doppler tenosynovitis present	51	0.52	0.58	< 0.001
Biceps tendon tear		-0.11		NS
Synovitis				
Axillary recess	79	0.95	0.19	< 0.001
Posterior recess	51	0.52	0.53	< 0.001
Synovial power Doppler signal				
Axillary recess	94	0.07	0.63	NS
Posterior recess	95	0.97	0.13	< 0.001
Bursitis: subdeltoid/subscapular bursa	41	0.61	0.50	< 0.001
Power Doppler bursa present	77	1.00	0.07	< 0.001
Rotator cuff tear				
Subscapularis	81	0.82	0.34	< 0.001
Supraspinatus	64	0.50	0.36	< 0.001
Infraspinatus tendon	81	0.84	0.36	< 0.001

* NS = not significant.

Table 6. Mean clinical values of various shoulder findings classified as normal or abnormal by MRI or US*

	DAS28				SPADI				DASH			
	N	Normal	N	Abnormal	N	Normal	N	Abnormal	N	Normal	N	Abnormal
MRI												
Axillary recess synovitis	8	4.5 ± 2.1	2	5.0 ± 1.6	8	41 ± 32	2	61 ± 29	8	29 ± 26	2	47 ± 33
Subacromial/subdeltoid bursitis	3	3.6 ± 3.1	7	5.0 ± 1.2	3	53 ± 46	7	41 ± 26	3	45 ± 39	7	27 ± 20
Biceps tendon sheath tenosynovitis	5	5.8 ± 0.4	5	3.4 ± 2.1†	5	57 ± 29	5	33 ± 31	5	41 ± 25	5	24 ± 27
Majority US												
Posterior recess synovitis	7	4.3 ± 2.2	3	5.2 ± 1.1	7	35 ± 30	3	67 ± 22	7	23 ± 21	3	55 ± 27
Rotator cuff tear	4	4.0 ± 2.7	6	5.0 ± 1.3	4	18 ± 17	6	63 ± 24†	4	10 ± 7	6	47 ± 23†

* US classification was considered normal or abnormal according to the majority of 11 US observers. MRI = magnetic resonance imaging; US = ultrasonography; DAS28 = Disease Activity Score in 28 joints; SPADI = Shoulder Pain and Disability Index; DASH = Disabilities of the Arm, Shoulder, and Hand questionnaire.
† $P < 0.05$.

were only possible for the findings of rotator cuff tear and posterior synovitis. Patients with abnormal rotator cuff findings according to the majority of US observers showed significantly higher clinical values of SPADI and DASH than those classified as normal. All of the power Doppler comparisons were not significant.

DISCUSSION

To our knowledge, this is the first study undertaken to date that specifically focuses on validation of shoulder synovitis detected by gray-scale and power Doppler US in patients with early and established RA. We used MRI as the reference imaging technique. The results show that US can reliably assess inflammation of the subacromial/subdeltoid bursa and the glenohumeral joint. In addition, our study shows that US can reliably assess signs of rotator cuff impairment in patients with RA. Furthermore, this study indicates that clinical measures of adult shoulder function have a poor correlation with imaging modalities and, therefore, have a limited role in assessing shoulder function. These findings are in agreement with previous studies that also have demonstrated the superiority of US compared with physical examination in patients with RA and shoulder pain (23–25).

Various studies on patients with shoulder disease in RA have demonstrated that US is comparable with MRI in being more sensitive than radiography in detecting bone erosion (26,27). Our first study on shoulder US confirmed the high agreement level between US and MRI for detecting erosions of the humeral head (10). However, since erosions are a relatively late sign in inflammatory joint disease, the aim of the present study was to assess the agreement of US and MRI in detecting inflammatory changes, i.e., glenohumeral joint synovitis, bursitis, and long biceps tendon sheath tenosynovitis. In studying synovitis with gray-scale US, 2 findings are of interest: detection of synovial proliferation and effusion. In addition, US has the advantage that it is able to detect the presence of hyperemia in the target area by the power Doppler mode. Although all of the inflamed structures were found more frequently by MRI than by gray-scale US, our study indicates that US is able to detect enlargement of the posterior recess and of the anterior recess with good to

excellent agreement compared with MRI and with moderate to good kappa values. The overall agreements of 66% and 80% of the observations for posterior recess synovitis and axillary synovitis, respectively, also compared favorably with 64% and 31%, respectively, found in a previous study (10). The particularly improved agreement for synovitis of the axillary recess is probably due to more rigid definitions of effusion (>3 mm) and a longer training session. The agreements for the detection of synovitis/effusion of the glenohumeral joint in 2 other studies (28,29) varied from moderate (50%) to excellent (89%). The PPV for axillary recess synovitis and the NPV for posterior recess synovitis were also excellent (Table 3). With regard to the subacromial/subdeltoid bursitis, there was a 64% agreement between gray-scale US and MRI, indicating a good agreement, with a fair kappa. The PPV for subacromial/subdeltoid bursitis was excellent. Adding power Doppler US to the number of cases of synovitis found on gray-scale US, intra- and interobserver agreement improved and kappa values increased (Tables 4 and 5). The validity of power Doppler US could not be assessed in this study because MRI has no comparable modality and we did not take histologic specimens. The agreement between US and MRI with regard to long biceps tendon sheath tenosynovitis showed the lowest kappa value, probably due to the fact that there were no cases of biceps tendon sheath tenosynovitis seen on MRI. Furthermore, the intra- and interobserver agreement for the power Doppler signal in both the posterior recess and the subacromial/subdeltoid bursitis of the shoulder was excellent, suggesting that power Doppler signal may be used in multicenter studies as a parameter for active shoulder joint synovitis. Power Doppler US for synovitis of the biceps tendon sheath and the subacromial/subdeltoid bursitis showed moderate to good and good to excellent intra- and interobserver agreement, respectively.

The poor and negative mean kappa values do not necessarily denote poor agreement but may have a technical explanation. When the prevalence of an abnormality in one item is close to 0 or 1, the agreement by chance is very high in such a way that the kappa, which represents the agreement exceeding the agreement by chance, may become artificially low or even negative. Since it is impossible to know in advance the prevalence of abnormalities in

each structure examined, especially in small sample sizes as used in reliability studies, prevalence biases cannot be ruled out by study design. Therefore, statistics alternative to the conventional Cohen's kappa have been used in order to minimize the effect of prevalence and biases (18,22).

Agreement for rotator cuff tear was good to excellent, confirming earlier studies (30,31). There was also a good agreement between both adult shoulder function tests and US for the presence of rotator cuff tear, whereas all of the other associations were not significant (Table 6). This suggests that the DASH and SPADI cannot be used to diagnose shoulder synovitis in RA.

Overall, the lowest agreements between US and MRI were found for gray-scale US of the long biceps tenosynovitis and bursitis of the subacromial/subdeltoid bursa. The lower agreement values could have been due to several reasons. Although biceps tendon sheath tenosynovitis is said to be present in a large percentage of patients with RA in one study, this could not be confirmed (32). One of the possible reasons may be that the cutoff value for a fluid rim around the biceps tendon sheath of 3 mm is too high and should be set at a lower level. Since no images were stored for reasons of time management, this presumption could not be checked. Some investigators were not familiar with the equipment and the scanner settings. The level of experience was also different for each sonographer. Perhaps a 10-minute investigation was for some investigators too short to perform a thorough examination. Three other elements of which the impact is unknown are the effect of the subsequent maneuvers on the localization of joint fluid; the difference in positioning of the patient during the MRI examination compared with the US examination, i.e., supine versus upright; and the lapse of a couple of days between the MRI and US examination.

In summary, this study shows that US is reliable in detecting synovitis of the axillary and posterior recess of the glenohumeral joint and subdeltoid/subacromial bursitis. We were able to detect these changes with a moderate to good interobserver reproducibility and similar intraobserver reproducibility. This implies that US can be launched in longitudinal shoulder studies by either an individual reader or multiple readers, although more studies are warranted for improvement of diagnosing long biceps tendon sheath tenosynovitis. We were not able to show a significant association between adult shoulder function tests and shoulder synovitis found on US or MRI. We recommend not only more standardization of the scanning technique, but also consistent criteria for diagnostic interpretation, especially of long biceps tendon sheath tenosynovitis.

ACKNOWLEDGMENTS

The authors would like to thank Siemens, Erlangen, Germany, for providing 5 US machines.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Bruyn had full access

to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Bruyn, Pineda, Moya, Filippucci, D'Agostino, Naredo.

Acquisition of data. Bruyn, Pineda, Hernandez-Diaz, Ventura-Rios, Moya, Pena, Espinosa, Möller, Filippucci, Iagnocco, Balint, Kane, Angulo, Ponte, Fernandez-Gallardo, Naredo.

Analysis and interpretation of data. Bruyn, Moya, Garrido, Groen, Kane, D'Agostino, Naredo.

REFERENCES

- Petersson CJ. Painful shoulders in patients with rheumatoid arthritis: prevalence, clinical and radiological features. *Scand J Rheumatol* 1986;15:275-9.
- Alasaarela E, Suramo I, Tervonen O, Lahde S, Takalo R, Hakala M. Evaluation of humeral head erosions in rheumatoid arthritis: a comparison of ultrasonography, magnetic resonance imaging, computed tomography and plain radiography. *Br J Rheumatol* 1998;37:1152-6.
- Kuper HH, van Leeuwen MA, van Riel PL, Prevoo ML, Houtman PM, Lolkema WF, et al. Radiographic damage in large joints in early rheumatoid arthritis: relationship with radiographic damage in hands and feet, disease activity, and physical disability. *Br J Rheumatol* 1997;36:855-60.
- Drossaers-Bakker KW, Kroon HM, Zwinderman AH, Breedveld FC, Hazes JM. Radiographic damage of large joints in long-term rheumatoid arthritis and its relation to function. *Rheumatology (Oxford)* 2000;39:998-1003.
- 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. *J Rheumatol* 2007;34:839-47.
- 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.
- Balint PV, Sturrock RD. Intraobserver repeatability and interobserver reproducibility in musculoskeletal ultrasound imaging measurements. *Clin Exp Rheumatol* 2001;19:89-92.
- Qvistgaard E, Torp-Pedersen S, Christensen R, Bliddal H. Reproducibility and inter-reader agreement of a scoring system for ultrasound evaluation of hip osteoarthritis. *Ann Rheum Dis* 2006;65:1613-9.
- Karim Z, Wakefield RJ, Quinn M, Conaghan PG, Brown AK, Veale DJ, et al. Validation and reproducibility of ultrasonography in the detection of synovitis in the knee: a comparison with arthroscopy and clinical examination. *Arthritis Rheum* 2004;50:387-94.
- Bruyn GA, Naredo E, Moller I, Moragues C, Garrido J, de Bock GH, et al. Reliability of ultrasonography in detecting shoulder disease in patients with rheumatoid arthritis. *Ann Rheum Dis* 2009;68:357-61.
- Hameed B, Pilcher J, Heron C, Kiely PD. The relation between composite ultrasound measures and the DAS28 score, its components and acute phase markers in adult RA. *Rheumatology (Oxford)* 2008;47:476-80.
- Williams JW, Holleman DR, Simel DL. Measuring shoulder function with shoulder pain and disability index. *J Rheumatol* 1995;22:727-32.
- Roach KE, Budiman-Mak E, Songsiridej N, Lestranakul Y. Development of a shoulder pain and disability index. *Arthritis Care Res* 1991;4:143-9.
- 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.
- Rubin JM, Adler RS, Fowlkes JB, Spratt S, Pallister JE, Chen JF, et al. Fractional moving blood volume: estimation with power Doppler US. *Radiology* 1995;197:183-90.
- Wakefield RJ, Balint PV, Szkudlarek M, Filippucci E, Backhaus M, D'Agostino M, et al. Musculoskeletal ultrasound

- including definitions for ultrasonic pathology. *J Rheumatol* 2005;32:2485–7.
17. Craig TG. Generalization of Scott's Index of intercoder agreement. *Public Opin Q* 1981;45:260–4.
 18. Cicchetti DV, Feinstein AR. High agreement but low κ : II. Resolving the paradoxes. *J Clin Epidemiol* 1990;43:551–8.
 19. Mak HK, Yau KK, Chan BP. Prevalence-adjusted bias-adjusted κ values as additional indicators to measure observer agreement [letter]. *Radiology* 2004;232:302–3.
 20. Byrt T, Bishop J, Carlin JB. Bias, prevalence and κ . *J Clin Epidemiol* 1993;46:423–9.
 21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
 22. Feinstein AR, Cicchetti DV. High agreement but low κ : I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43:543–9.
 23. Kim HA, Kim SH, Seo Y. Ultrasonographic findings of the shoulder in patients with rheumatoid arthritis and comparison with physical examination. *J Korean Med Sci* 2007;22:660–6.
 24. Naredo E, Aguado P, De Miguel E, Uson J, Mayordomo L, Gijon-Banos J, et al. Painful shoulder: comparison of physical examination and ultrasonographic findings. *Ann Rheum Dis* 2002;61:132–6.
 25. Luukkainen R, Sanila MT, Luukkainen P. Poor relationship between joint swelling detected on physical examination and effusion diagnosed by ultrasonography in glenohumeral joints in patients with rheumatoid arthritis. *Clin Rheumatol* 2007;26:865–7.
 26. Hermann KG, Backhaus M, Schneider U, Labs K, Loreck D, Zuhlsdorf S, et al. Rheumatoid arthritis of the shoulder joint: comparison of conventional radiography, ultrasound, and dynamic contrast-enhanced magnetic resonance imaging. *Arthritis Rheum* 2003;48:3338–49.
 27. Bruyn GA, Iagnocco A, Naredo E, Wakefield RJ, Schmidt WA. Ultrasonography of the shoulder in patients with rheumatoid arthritis: comment on the article by Hermann et al [letter]. *Arthritis Rheum* 2004;50:3054.
 28. Scheel AK, Schmidt WA, Hermann KG, Bruyn GA, D'Agostino MA, Grassi W, et al. Interobserver reliability of rheumatologists performing musculoskeletal ultrasonography: results from a EULAR "Train the trainers" course. *Ann Rheum Dis* 2005;64:1043–9.
 29. Naredo E, Moller I, Moragues 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.
 30. Burk DL, Karasick D, Kurtz AB, Mitchell DG, Rifkin MD, Miller CL, et al. Rotator cuff tears: prospective comparison of MR imaging with arthrography, sonography, and surgery. *AJR Am J Roentgenol* 1989;153:87–92.
 31. Swen WA, Jacobs JW, Algra PR, Manoliu RA, Rijksmans J, Willems WJ, et al. Sonography and magnetic resonance imaging equivalent for the assessment of full-thickness rotator cuff tears. *Arthritis Rheum* 1999;42:2231–8.
 32. Strunk J, Lange U, Kurten B, Schmidt KL, Neeck G. Doppler sonographic findings in the long bicipital sheath in patients with rheumatoid arthritis as compared with patients with degenerative diseases of the shoulder. *Arthritis Rheum* 2003;48:1828–32.