RHEUMATOLOGY

Rheumatology 2020;59:129–136
doi:10.1093/rheumatology/kez231
Advance Access publication 25 June 2019

Original article

Flares in rheumatoid arthritis: do patient-reported swollen and tender joints match clinical and ultrasonography findings?

Dorota Kuettel [1,2], Lene Terslev[3], Ulrich Weber [1,2], Mikkel Østergaard [3], Jette Primdahl [1,2,4], Randi Petersen [1], Mads Ammitzbøll-Danielsen[3], Sören Möller [5] and Kim Hørslev-Petersen [1,2]

Abstract

Objectives. To investigate how patient-reported flares in RA are related to clinical joint examination and inflammation detected by US.

Methods. Eighty RA patients with DAS28-CRP <3.2 and no swollen joints at baseline were followed for 1 year. In case of patient-reported hand flare with swollen and tender joints (SJ and TJ, respectively), patients underwent clinical examination for SJ/TJ and US of bilateral wrists, MCP and PIP 1st–5th, six extensor tendon compartments and wrist flexor tendons for synovitis/tenosynovitis. Percentage agreement and kappa were calculated between patient-reported SJ and TJ, clinical examination for SJ/TJ and US findings indicative of inflammation. With US as reference, sensitivity, specificity, positive/negative predictive value and accuracy of patient-reported and clinically examined joints were determined.

Results. Hand flare was reported by 36% (29/80) of patients. At time of flare, all clinical and ultrasonographic measures of disease activity deteriorated compared with baseline. Agreement between patient-reported SJ/TJ, clinically examined SJ/TJ and US was slight (kappa = 0.02–0.20). Patients and clinicians agreed in 79–93% of joints, more frequently on SJ than TJ. With US as reference, specificities were 86–100% and 88–100%, and sensitivities 12–34% and 4–32% for patient-reported SJ/TJ and clinically examined SJ/TJ, respectively.

Conclusion. Over 12 months of follow-up, hand flare was reported by every third RA patient. Self-reported flares were associated with increased disease activity as determined by clinical examination and US. Patient-reported joint assessment may aid in capturing flares between routine clinical visits.

Key words: rheumatoid arthritis, patient-reported outcomes, flares, ultrasonography

Rheumatology key messages

- Self-reported flares in RA patients were accompanied by worsening of clinical, laboratory and US parameters.
- RA patient self-assessments of joints may aid in capturing flares between routine clinical visits.

Introduction

In RA, clinical assessment of joint swelling and tenderness is a cornerstone of routine monitoring of synovitis, a potential predictor of joint damage [1]. Joint counts are important elements of composite disease activity indices in RA, such as the DAS or the Simplified Disease Activity Index [2–4]. However, there are several limitations in the evaluation of joint involvement, including reliability of clinical examination and inconsistencies in joint examination across consultations [5, 6]. Patients' self-assessment of disease activity may facilitate clinical decision-making, and patient-reported outcomes are already integrated in clinical trials and daily practice [7–9]. Shared decision-making and patient involvement in RA management are

[1]Danish Hospital for Rheumatic Diseases, University Hospital of Southern Denmark, Sønderborg, [2]Department of Regional Health Research, University of Southern Denmark, Odense, [3]Copenhagen Center for Arthritis Research (COPECARE), Center for Rheumatology and Spine Diseases, Rigshospitalet – Glostrup, Glostrup, [4]Sygehus Sønderjylland, University Hospital of Southern Denmark, Aabenraa and [5]OPEN – Odense Patient data Explorative Network, Odense University Hospital, Department of Clinical Research, University of Southern Denmark, Odense, Denmark

Submitted 6 March 2019; accepted 15 May 2019

Correspondence to: Dorota Kuettel, Danish Hospital for Rheumatic Diseases, University Hospital of Southern Denmark, Engelshøjgade 9A, DK-6400 Sønderborg, Denmark.
E-mail: Dkuttel@gigtforeningen.dkdkuettel@health.sdu.dk

CLINICAL SCIENCE

recommended by EULAR and are preferred by most patients [8, 9]. This may improve self-management abilities and lead to sustainable health benefits in people with arthritis [10].

Several studies have evaluated the reliability of patient assessment of disease activity by exploring agreement between patient-reported joint counts and clinical examination [11–14]. Patient-reported tender joints correlate better with the evaluation by trained assessors than patient-reported swollen joints [14]. However, both patient-reported and health professional joint assessment have shown varying reliability, especially regarding joint swelling [5, 15–18].

RA flares are common and often occur between consultations [19]. There is limited evidence for whether patient self-assessment of joint inflammation is a reliable approach to capture transient flares that would otherwise go undetected between routine clinical visits. The clinical relevance of detecting flares to modify treatment in RA patients is underlined by more radiographic progression in those experiencing a flare than in those without flares [20]. Moreover, self-reported flares are independent predictors of joint damage and are associated with worse functional outcome [19, 21].

Our objectives were to assess the agreement between patient-reported swollen and tender joints, clinical examination and inflammation detected by US in RA patients at the time of patient-reported flare.

Methods

Study design and participants

For this prospective, observational study with a 1-year follow-up, consecutive RA patients were recruited from the outpatient clinic at the Danish Hospital for Rheumatic Diseases, Sønderborg, Denmark, when attending routine appointments.

Patients $\geqslant$18 years were eligible if they fulfilled the ACR 1987 or ACR/EULAR 2010 criteria for RA, were RF and/or anti-CCP positive, had a DAS based on CRP (DAS28-CRP) $<$3.2 at baseline and no clinically swollen joints. Tender joints were not an inclusion criterion as tender joints in contrast to swollen joints are not necessarily related to inflammatory joint involvement [22]. Further requirements were stable DMARD treatment and no intra-articular glucocorticoid injections 4 weeks prior to study entry.

Patients were requested to contact the hospital in case of a flare in a hand/wrist defined as worsening of RA accompanied by at least one swollen and tender joint, as perceived by the patient. Only the first flare was registered. Information about when the flare began was recorded and a flare visit was scheduled within 3 days upon patient's contact to the hospital.

The FLAre-in-RA study was approved by the local ethics committee (The Regional Scientific Ethical Committees for Southern Denmark, S-20160027) and was conducted according to the Declaration of Helsinki 2008. Written informed consent was obtained from all participants.

Clinical, laboratory and patient-reported outcomes

Age, gender, weight, height, disease duration and ongoing therapy for RA were recorded at baseline.

At baseline and at flare visits, a senior rheumatologist or a trained research nurse carried out a clinical examination for swollen and tender joints and DAS28-CRP was calculated. Both evaluators were blinded to US results and patient reports. Likewise, the patients indicated swollen and tender joints on a mannequin sketch representing the same 28 joints. At the baseline visit patients received 10 min instruction by a trained research nurse and written information on how to perform joint assessment for swelling. Visual analogue scales were used to report the patient's and evaluator's global assessments. The Danish version of the HAQ was used to assess physical function. Patients completed two flare questionnaires: The Flare Instrument and the Outcome Measures in Rheumatology (OMERACT) flare questionnaire [23, 24]. Routine laboratory tests included CRP, RF and anti-CCP level.

US examination at baseline and at flare

A General Electric Logiq E9 US machine with a multifrequency linear array transducer 6–15 was used for all examinations. Standardized Colour Doppler (CD) settings were kept unchanged throughout the study with Doppler frequency 7.5 MHz, pulse repetition frequency of 0.4 and Doppler gain just below the noise threshold [25]. CD was chosen over power Doppler because it is the most sensitive modality on this US machine [26]. One experienced musculoskeletal ultrasonographer performed all US examinations blinded to clinical, laboratory and patient-reported data. Patients were asked not to talk about their symptoms during scanning and all examinations were performed in the same darkened room, on the same day as the clinical examination. The US protocol included multiplanar scanning of 22 joints/regions at each visit: bilateral wrists (radiocarpal, midcarpal and distal radioulnar joints, dorsal recesses), the 1st–5th MCP, dorsal recesses, the 1st interphalangeal joint, the 2nd–5th PIP, dorsal and volar recesses, bilateral I–VI extensor tendon compartments and bilateral three flexor tendons/groups: flexor carpi radialis, flexor pollicis longus and combined flexor digitorum superficialis and profundus. Synovitis and tenosynovitis were defined according to the OMERACT US definitions, assessed by CD and grey scale (GS) and graded semi quantitatively 0–3 according to the EULAR-OMERACT scoring systems [27–30]. US-positive joint/tendon sheath was defined as an EULAR-OMERACT combined score $\geqslant$1. Grade 1 EULAR-OMERACT combined score is defined as hypoechoic synovial/tenosynovial hypertrophy grade 1 and CD grade $\leqslant$1 [27, 28]. The wrist was considered positive for US tenosynovitis if the combined US score was $\geqslant$1 in one or more tendon.

To assess the reliability of the US examiner (D.K.), inter-rater reliability was tested against an US expert (L.T.), and

intra-rater reliability was assessed by scoring and re-scoring 40 randomly selected images from the study with a 2-month interval.

Statistical analysis

For normally distributed data, parametric analyses (paired sample *t*-test) and for non-normally distributed data non-parametric analyses (Wilcoxon matched-pairs signed rank test) were applied.

Cohen's kappa and percentage agreement were calculated between US inflammation (synovitis and/or tenosynovitis in wrist, synovitis in MCP and PIP) and 22 self-assessed, respectively, clinically examined joints in wrists and hands (separately for tenderness and swelling). Likewise, agreement between the patients' and the evaluators' assessment of tenderness and swelling was examined. The following thresholds for kappa agreement were applied: <0.00 poor, 0.00–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial and 0.81–1.00 almost perfect [31]. Furthermore, we determined sensitivity, specificity, positive/negative predictive value and accuracy of patient-report and clinical examination for joint swelling and tenderness using US-detected inflammation as the reference method. *P*-values $\leqslant 0.05$ were considered statistically significant. Statistical analyses were performed using Stata version 15.0 (StataCorp., College Station, TX, USA).

Results

Patients' characteristics, clinical findings and patient-reported outcomes

Eighty patients with established RA were included. At baseline, none of the patients reported swollen joints and no swollen joints were detected at clinical examination. Mean (S.D.), [range] numbers of patient-reported and clinically examined tender joints were 1.0 (1.9) [0–8] and 0.8 (1.5) [0–7], respectively. Mean DAS28-CRP was 2.0 (0.7) [1.2–3.1]. Demographic and clinical characteristics at baseline and at the flare visit are shown in Table 1.

Thirty-six percent (29/80) reported a hand flare and contacted the clinic for a flare visit during the 1-year follow-up. At the flare visit, all measures of disease activity increased as compared with baseline (Table 1). Average (S.D.) [range] DAS28-CRP increased to 3.5 (1.0) [1.2–5.3] ($P < 0.001$) and 65.5% (19/29) of patients fulfilled the DAS flare definition (an increase in DAS28 >1.2) at the time of patient-reported flare [32]. Furthermore, all patient-reported outcomes deteriorated significantly, for both flare questionnaires, and patients' global assessment increased from 22.1 (23.1) [0–84] to 43.4 (26.2) [0–85] (Table 1).

Onset of patient-reported flare and patients' delay

Mean (S.D.) [range] number of days between baseline and flare visit was 100.2 (92.6) [9–279]. The mean (S.D.) delay between patient reported flare onset and contact to the hospital was 3 (8.7) [0–48] days. Most patients (16/29) were scheduled for the flare visit the same day they contacted the outpatient clinic, while 10, 2 and 1 patients were seen within 1, 2 and 3 days, respectively.

Reliability assessment of the US examiner

For the US examiner, the inter-rater reliability expressed by kappa was 0.90 for both GS and CD synovitis, and 0.85 and 1.0 for GS and CD tenosynovitis, respectively. Intra-rater reliability was 0.85 and 1.0 for GS and CD, respectively, both for synovitis and tenosynovitis.

US findings at baseline and at flare

At baseline, US detected synovitis in 63.8% (37/58) and tenosynovitis in 19% (11/58) of wrists. In MCP and PIP joints synovitis was found in 25.2% (73/290) and 12.1% (35/289) of joints, respectively.

At flare, US showed synovitis in 84.5% (49/58) and tenosynovitis in 41% (24/58) of wrists. In MCP and PIP joints synovitis was detected in 40.0% (116/290) and 17.6% (51/289), respectively.

Agreement between patient-reported and clinically examined swollen and tender joints at flare

For swelling and tenderness, respectively, agreement between patient-report and clinical examination showed the highest concordance in PIP (93% and 83%), followed by MCP (88% and 82%) and wrist (79% and 79%) (Table 2). Kappa ranged from fair to moderate (0.23–0.50).

Agreement between patient-reported and clinically determined swollen and tender joints at flare and inflammation by US

The agreement for joint swelling and tenderness by clinical examination and signs of inflammation (synovitis and/or tenosynovitis) by US was slightly better than the agreement between patient-reported joint involvement and US (Table 3). The highest percentage agreement was found for PIP (74–81%), followed by MCP (60–65%) and wrist (41%). Overall, agreement expressed by kappa was slight, ranging from 0.02–0.20 for clinical examination and 0.03–0.12 for patient report.

Sensitivity, specificity, positive and negative predictive value, and accuracy of patient-reported and clinically examined swollen and tender joints with US inflammation as the reference method at flare

Using US as a reference, clinically swollen joints showed a high specificity of 100% (95% CI 63–100), 98% (95% CI 95–100) and 97% (95% CI 95–99) in wrist, MCP and PIP, respectively (Table 4). Likewise, specificity for patient-reported joint swelling was high, though slightly lower than clinical swelling: 87% (95% CI 47–100), 91% (95% CI 86–95) and 95% (95% CI 91–97) in wrist, MCP and PIP, respectively. The specificities of clinically assessed and patient-reported tender joints were good but lower than those of swollen joints (Table 4). The sensitivities of clinical examination and patient-reported joint involvement and inflammation by US were low, ranging from 4 to 34% (Table 4). The highest positive predictive value of 100% (95% CI 79–100) was observed for clinically examined

TABLE 1 Baseline and flare visit characteristics of RA patients reporting flare during 1 year of follow-up

Characteristics	BL	FV	P-value BL *vs* FV
Age, years	64.8 (9.7) [47–82]	N/A	N/A
Female/male (%)	20/9 (69/31)	N/A	N/A
Disease duration, years	10.3 (5.5) [2–21]	N/A	N/A
Anti-CCP positive (%)	27 (93)	N/A	N/A
IgM RF positive (%)	26 (90)	N/A	N/A
cDMARD (%)	27 (93)	27 (93)	N/A
cDMARD + bDMARD (%)	4 (14)	4 (14)	N/A
No ongoing treatment (%)	2 (7)	2 (7)	N/A
Glucocorticoids (%)	0 (0)	0 (0)	N/A
BMI	27.7 (5.2) [20.1–42.3]	N/A	N/A
DAS28-CRP	2.0 (0.7) [1.2–3.1]	3.5 (1.0) [1.2–5.3]	<0.0001
CRP, mg/l	5.7 (8.6) [0.0–33]	10.1 (12.7) [0.4–48]	<0.001
ESR, mm/h	12.8 (9.1) [2–40]	17.8 (13.9) [1–71]	<0.001
TJC28 [0–28]	0.8 (1.5) [0–7]	4.0 (3.5) [0–12]	<0.0001
SJC28 [0–28]	0 (0) [0–0]	1.5 (1.0) [0–4]	<0.0001
Patient-reported TJC28 [0–28]	1 (1.9) [0–8]	4.0 (3.0) [1–11]	<0.0001
Patient-reported SJC28 [0–28]	0 (0) [0–0]	2.8 (2.9) [1–11]	<0.0001
VAS_patient_global [0–100]	22.1 (23.1) [0–84]	43.4 (26.2) [0–85]	<0.0001
VAS_evaluator [1–100]	4.1 (3.2) [0–11]	16.2 (13.1) [2–74]	<0.0001
HAQ	0.4 (0.4) [0–1.5]	0.6 (0.5) [0–1.5]	<0.001
Flare instrument total score [0–120]	22.7 (23.2) [0–92]	47.4 (27.4) [0–93]	<0.0001
OMERACT flare questionnaire [range 0–50]	10.1 (9.6) [0–33]	22.9 (12.7) [1–49]	<0.0001
US joint count [0–22]	5.2 (3.4) [0–14]	8.1 (3.7) [1–18]	<0.001

Mean (S.D.) [range] scores or counts (%) of demographic, clinical, patient-reported and US characteristics at baseline and flare visit. BL: baseline; bDMARD: biological DMARD; cDMARD: conventional DMARD; FV: flare visit; OMERACT: Outcome Measures in Rheumatology; SJC: swollen joint count; TJC: tender joint count; VAS: visual analogue scale. *P*-value for paired data comparison between two time points.

wrist, for both swelling and tenderness, followed by patient-reported wrist. Positive predictive values were lower in MCP, and lowest in PIP.

Discussion

In this observational 1-year follow-up study, 36% of the RA patients in remission or low disease activity with clinically no swollen joints at inclusion experienced self-reported flare in the hand/wrist. Flares were consistently characterized by worsening of clinical and laboratory disease activity parameters and in parallel by increased ultrasonographic inflammation in joints and tendons, indicating inflammation as key trigger to flares. We found a fair to moderate agreement between patient-reported joint involvement and clinical joint examination for swollen and tender joints. However, joint swelling and tenderness, as determined by clinical examination and patients' self-report, did not correspond with US signs of inflammation at joint/site level, showing only slight kappa values. Both clinical examination and patients' self-report were highly specific for inflammation as detected by US, but their sensitivities were limited.

There are different approaches to defining a flare in RA. The two main definitions are a patient-reported worsening, and definitions based on clinical features and/or laboratory tests [33]. In this study, flares were defined by the patient's perception of disease worsening along with patient's perception of at least one swollen and tender joint. We do not presume that this definition covers the whole spectrum of flare experiences from the patients' perspective [34]. Nonetheless, these two domains have been identified as key features in describing worsening of disease activity by patients and clinicians [23, 34]. Flares, seen from the patients' position, do not always translate into increased inflammation [35]. However, all the patients who contacted our outpatient clinic because of perceived worsening of their RA displayed signs of increased inflammatory activity by clinical evaluation and/or US assessment.

Consistent with our findings, several studies have shown limited agreement at joint level between clinical examination and US [15, 36, 37]. Potential explanations for this discrepancy include factors such as joint deformity or OA. Clinical examination has inferior capacity to discriminate active from structural joint lesions compared with imaging modalities. OA or joint deformities such as subluxation may hamper clinical joint assessment, and joint tenderness may be induced mechanically rather than by immune-mediated inflammation. These factors may contribute to the overdiagnosis of synovitis by clinical examination/patient report.

Furthermore, it has been observed that the degree and direction of discrepancy between clinically and US-detected synovitis may be related to joint region and size:

Table 2 Agreement between patient-reported and clinically examined swollen and tender joints at flare

Joint/ site	No of joints/sites examined	Comparison between	+/+	+/–	–/+	–/–	% agreement	Kappa (95% CI)
Wrist	*n* = 58	Clinically examined swelling *vs* patient-reported swelling	11	5	7	35	79	0.50 (0.26–0.75)
		Clinically examined tenderness *vs* patient-reported tenderness	10	6	6	36	79	0.48 (0.23–0.73)
MCP	*n* = 290	Clinically examined swelling *vs* patient-reported swelling	10	8	27	245	88	0.30 (0.14–0.47)
		Clinically examined tenderness *vs* patient-reported tenderness	25	29	24	212	82	0.37 (0.24–0.51)
PIP	*n* = 289	Clinically examined swelling *vs* patient-reported swelling	4	4	15	266	93	0.27 (0.04–0.50)
		Clinically examined tenderness *vs* patient-reported tenderness	12	21	27	229	83	0.23 (0.09–0.39)

PIP *n* = 289, as one patient had an amputation.

Table 3 Agreement between patient-reported/clinically examined joints *vs* US inflammation[a] at flare

Joint/ site	No of joints/ sites examined	Comparison between	+/+	+/–	–/+	–/–	% agreement	Kappa (95% CI)
Wrist	*n* = 58	Clinically examined swelling *vs* US inflammation	16	0	34	8	41	0.12 (0.03–0.20)
		Patient-reported swelling *vs* US inflammation	17	1	33	7	41	0.08 (–0.03–0.19)
		Clinically examined tenderness *vs* US inflammation	16	0	34	8	41	0.12 (0.03–0.20)
		Patient-reported tenderness *vs* US inflammation	16	0	34	8	41	0.12 (0.03–0.20)
MCP	*n* = 290	Clinically examined swelling *vs* US inflammation	15	3	101	171	64	0.13 (0.06–0.20)
		Patient-reported swelling *vs* US inflammation	21	16	95	158	62	0.10 (0.01–0.19)
		Clinically examined tenderness *vs* US inflammation	34	20	82	154	65	0.20 (0.09–0.3)
		Patient-reported tenderness *vs* US inflammation	25	24	91	150	60	0.09 (–0.01–0.19)
PIP	*n* = 289	Clinically examined swelling *vs* US inflammation	2	6	49	232	81	0.02 (–0.06–0.11)
		Patient-reported swelling *vs* US inflammation	6	13	45	225	80	0.08 (–0.04–0.20)
		Clinically examined tenderness *vs* US inflammation	13	20	38	218	80	0.20 (0.06–0.34)
		Patient-reported tenderness *vs* US inflammation	8	31	43	207	74	0.03 (–0.09–0.15)

[a]US inflammation: synovitis and/or tenosynovitis in wrist and synovitis in MCP and PIP was present if EULAR-OMERACT combined score was $\geqslant 1$. Grade 1 EULAR-OMERACT combined score is defined as hypoechoic synovial/tenosynovial hypertrophy grade 1 and colour Doppler grade $\leqslant 1$. OMERACT: Outcomes Measures in Rheumatology.

superior concordance has been observed for larger joints (shoulder, elbow, knee and ankle) compared with smaller joints (PIP, MCP and wrist) [38].

US is more sensitive than clinical examination for detecting synovitis and tenosynovitis and our observations are consistent with previous reports showing low sensitivity and high specificity of clinical examination with US as reference [39–41]. One study assessed the concordance of clinical joint swelling and US synovitis in 28 joints of seven active RA patients and concluded that kappa agreement was poor, with percentage agreement of 73% [15]. In our study, the overall agreement between clinical examination and US ranged from 41 to 81%, while kappa agreement was slight.

To date, only a few studies have assessed the relationship between patients' self-assessment of joints and US signs of inflammation [18, 37, 38, 42]. Direct comparisons across studies are challenging because different US scoring systems are used in different sets of joints/sites, and different measures of agreement are used. Keeping these

TABLE 4 Sensitivity, specificity, PPV, NPV and accuracy of patient-reported/clinically examined, joints, with ultrasonography[a] as reference

Joint/site	Examination	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy
Wrist	Clinically examined swelling	0.32 (0.20–0.47)	1.00 (0.63–1.0)	1.00 (0.79–1.0)	0.19 (0.09–0.34)	0.43
	Patient-reported swelling	0.34 (0.21–0.49)	0.87 (0.47–1.0)	0.94 (0.73–1.0)	0.17 (0.07–0.33)	0.40
	Clinically examined tenderness	0.32 (0.20–0.47)	1.00 (0.63–1.0)	1.00 (0.79–1.0)	0.19 (0.09–0.34)	0.43
	Patient-reported tenderness	0.32 (0.20–0.47)	1.00 (0.63–1.0)	1.00 (0.79–1.0)	0.19 (0.09–0.34)	0.40
MCP	Clinically examined swelling	0.13 (0.07–0.20)	0.98 (0.95–1.0)	0.83 (0.59–0.96)	0.63 (0.57–0.69)	0.64
	Patient-reported swelling	0.18 (0.12–0.26)	0.91 (0.86–0.95)	0.57 (0.39–0.73)	0.62 (0.56–0.68)	0.62
	Clinically examined tenderness	0.29 (0.21–0.38)	0.89 (0.83–0.93)	0.63 (0.49–0.76)	0.65 (0.59–0.71)	0.65
	Patient-reported tenderness	0.22 (0.14–0.30)	0.86 (0.80–0.91)	0.51 (0.36–0.66)	0.62 (0.56–0.68)	0.60
PIP	Clinically examined swelling	0.04 (0.00–0.13)	0.97 (0.95–0.99)	0.25 (0.03–0.65)	0.83 (0.78–0.87)	0.81
	Patient-reported swelling	0.12 (0.04–0.24)	0.95 (0.91–0.97)	0.32 (0.13–0.57)	0.83 (0.78–0.88)	0.80
	Clinically examined tenderness	0.25 (0.14–0.40)	0.92 (0.87–0.95)	0.39 (0.23–0.58)	0.85 (0.80–0.89)	0.80
	Patient-reported tenderness	0.16 (0.07–0.29)	0.87 (0.82–0.91)	0.21 (0.09–0.36)	0.83 (0.78–0.87)	0.74

[a]US inflammation was defined as synovitis and/or tenosynovitis in wrist, synovitis in MCP and PIP, EULAR-OMERACT combined score $\geqslant 1$. Grade 1 EULAR-OMERACT combined score is defined as hypoechoic synovial/tenosynovial hypertrophy grade 1 and colour Doppler grade $\leqslant 1$. NPV: negative predictive value; OMERACT: Outcomes Measures in Rheumatology; PPV: positive predictive value.

limitations in mind, our results are in line with previous studies showing limited concordance between patient reports and US when different cut-offs for US positive signs of synovitis are applied [18, 42]. In our study, the agreement between patient-reported swollen joints and US inflammation varied between 41 and 80%, dependent on joint size, and for patient-reported tenderness ranged from 41% to 74%.

Despite being a frequent finding in RA patients, tenosynovitis still appears to be an underestimated manifestation of RA and widely used clinical indices do not include tendon involvement [30, 35]. In the current study, we looked for both synovitis and tenosynovitis at the time of flare, as the ability to differentiate between these two conditions can be clinically challenging. The Sonographic Tenosynovitis/arthritis Assessment in Rheumatoid Arthritis Patients in Remission (STARTER) study highlighted the importance of incorporating tenosynovitis-targeted US assessment of RA patients in clinical remission [35, 43]. Our study showed tenosynovitis both at baseline, despite required low disease activity and no clinically swollen joints, and during self-reported flares, which calls for inclusion of ultrasonographic tendon assessment at the time of flare.

Several studies comparing clinical examination and patients' self-assessment showed stronger associations between patients and physicians with respect to tender than to swollen joints [11, 13, 14]. In our study, we observed that patients and clinicians agreed more frequently on swelling than on tenderness. Even though we analysed data at joint level and not at patient level, our results support the notion that patients may reliably report swelling by self-evaluating joints at the time of flare. A short training has been shown to be sufficient to enhance patients' abilities to accurately assess joint swelling [44]. Although we did not carry out an extensive patient training program on joint evaluation, we may assume that the good performance in evaluating joints for swelling was partly due to the instructions the patients received at study entry. Together with the observation that patients performed only slightly differently from the clinical assessors when compared with US, this may have implications for clinical practice by making patients candidates for self-assessment of disease activity when a flare occurs between scheduled appointments.

Our study has various strengths: a homogeneous cohort of RA patients, the same rheumatologist performing all US assessments with high inter- and intra-observer reliability and the same research nurse instructing all the patients. Furthermore, 55% of the patients were evaluated the same day they contacted the outpatient clinic, and the remainder within 72 h. Clinical and US examinations and patient-reported outcomes, including joint assessment, were all performed on the same day, which strengthens the associations since flare is a dynamic phenomenon.

Limitations are the one-centre design and the relatively small sample size. Another potential limitation may be a selected spectrum of RA, since we only included RF and/or anti-CCP-positive patients. The presence of autoantibodies is indicative of a more severe disease course and is associated with bone damage [45]. However, it remains unknown whether patients who do not develop immune response differ from patients in whom autoantibodies are present regarding different flare phenotypes.

In the current study only tendons at the wrist level were included as distinction between synovitis and tenosynovitis may be clinically challenging in this region. Another limitation may be the applied definition of a positive joint/tendon sheath on US, as grade 1 synovial hypertrophy has been observed in normal subjects [46, 47]. However, the most common abnormality in healthy joints is effusion, which is not scored by the OMERACT-EULAR composite US score [27, 28, 47]. Grade 1 synovial hypertrophy has

previously been utilized as the cut-off [43, 48] and although it is a strict definition, recent papers have shown that both grade 1 synovial and tenosynovial hypertrophy are responsive to treatment, and thus reflect true inflammation [49, 50]. The optimal cut-off between pathology and normality in arthritic joints has yet to be determined.

In conclusion, our findings suggest that patient-reported flares reflect inflammatory disease activity. Patient report, clinical examination and US may capture different aspects of patient-reported flares; however, this assumption needs to be further investigated. Patient self-assessment of joints, which was highly concordant with clinical examination, may be a potential amendment to traditional monitoring of RA patients to capture flares between routine clinical visits.

Acknowledgements

We thank research secretary Kirsten Frøhlich and research radiographer Henning Jakobsen, Danish Hospital for Rheumatic Diseases in Sønderborg, for logistic support.

Funding: This work was supported by research grants from the Danish Rheumatism Association, University of Southern Denmark, The Region of Southern Denmark, Knud og Edith Eriksens Mindefond.

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

References

1 Welsing PM, van Gestel AM, Swinkels HL, Kiemeney LA, van Riel PL. The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis. Arthritis Rheum 2001;44:2009–17.

2 Van der Heijde DM, van't Hof MA, van Riel PL *et al*. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. Ann Rheum Dis 1990;49:916–20.

3 Prevoo MLL, van't Hof MA, Kuper HH *et al*. 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.

4 Smolen JS, Breedveld FC, Schiff MH *et al*. A simplified disease activity index for rheumatoid arthritis for use in clinical practice. Rheumatology (Oxford) 2003;42:244–57.

5 Pincus T. Limitations of a quantitative swollen and tender joint count to assess and monitor patients with rheumatoid arthritis. Bull NYU Hosp Jt Dis 2008;66:216–23.

6 Pincus T, Segurado OG. Most visits of most patients with rheumatoid arthritis to most rheumatologists do not include a formal quantitative joint count. Ann Rheum Dis 2006;65:820–2.

7 El M, El GM, Youssef SS, Palmer D. Incorporating patient-reported outcome measures in clinical practice: development and validation of a questionnaire for inflammatory arthritis. Clin Exp Rheumatol 2010;28:734–44.

8 Smolen JS, Landewé R, Bijlsma J *et al*. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. Ann Rheum Dis 2017;76:960–77.

9 Nota I, Drossaert CH, Taal E, Vonkeman HE, van de Laar MA. Patient participation in decisions about disease modifying anti-rheumatic drugs: a cross-sectional survey. BMC Musculoskelet Disord 2014;15:333.

10 Osborne RH, Wilson T, Lorig KR, McColl GJ. Does self-management lead to sustainable health benefits in people with arthritis? A 2-year transition study of 452 Australians. J Rheumatol 2007;34:1112–7.

11 Kavanaugh A, Lee SJ, Weng HH *et al*. Patient-derived joint counts are a potential alternative for determining disease activity score. J Rheumatol 2010;37:1035–41.

12 Radner H, Grisar J, Smolen JS, Stamm T, Aletaha D. Value of self-performed joint counts in rheumatoid arthritis patients near remission. Arthritis Res Ther 2012;14:R61.

13 Barton JL, Criswell LA, Kaiser R, Chen YH, Schillinger D. Systematic review and metaanalysis of patient self-report versus trained assessor joint counts in rheumatoid arthritis. J Rheumatol 2009;36:2635–41.

14 Figueroa F, Braun-Moscovici Y, Khanna D *et al*. Patient self-administered joint tenderness counts in rheumatoid arthritis are reliable and responsive to changes in disease activity. J Rheumatol 2007;34:54–6.

15 Marhadour T, Jousse-Joulin S, Chalès G *et al*. Reproducibility of joint swelling assessments in long-lasting rheumatoid arthritis: influence on Disease Activity Score-28 values (SEA-Repro study part I). J Rheumatol 2010;37:932–7.

16 Naredo E, Bonilla G, Gamero F *et al*. 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.

17 Salaffi F, Filippucci E, Carotti M *et al*. Inter-observer agreement of standard joint counts in early rheumatoid arthritis: a comparison with grey scale ultrasonography—a preliminary study. Rheumatology (Oxford) 2008;47:54–8.

18 Cheung PP, Ruyssen-Witrand A, Gossec L *et al*. Reliability of patient self-evaluation of swollen and tender joints in rheumatoid arthritis: a comparison study with ultrasonography, physician, and nurse assessments. Arthritis Care Res 2010;62:1112–9.

19 Kuettel D, Primdahl J, Christensen R, Ørnbjerg LM, Hørslev-Petersen K. Impact of patient-reported flares on radiographic progression and functional impairment in patients with rheumatoid arthritis: a cohort study based on the AMBRA trial. Scand J Rheumatol 2018;47:87–94.

20 Saleem B, Brown AK, Quinn M *et al*. Can flare be predicted in DMARD treated RA patients in remission, and is it important? A cohort study. Ann Rheum Dis 2012;71:1316–21.

21 Ometto F, Raffeiner B, Bernardi L *et al*. Self-reported flares are predictors of radiographic progression in rheumatoid arthritis patients in 28-joint disease activity score remission: a 24-month observational study. Arthritis Res Ther 2016;18:89.

22 Hammer HB, Michelsen B, Provan SA *et al*. Tender joint count may not reflect inflammatory activity in established

rheumatoid arthritis patients - results from a longitudinal study. Arthritis Care Res (Hoboken) 2018; doi: 10.1002/acr.23815.

23 Berthelot JM, De Bandt M, Morel J *et al*. A tool to identify recent or present rheumatoid arthritis flare from both patient and physician perspectives: the 'FLARE' instrument. Ann Rheum Dis 2012;71:1110–6.

24 Bartlett SJ, Barbic SP, Bykerk V *et al*. Content and construct validity, reliability, and responsiveness of the rheumatoid arthritis flare questionnaire: OMERACT 2016 Workshop Report. J Rheumatol 2017;44:1536–43.

25 Torp-Pedersen S, Christensen R, Szkudlarek M *et al*. Power and color Doppler ultrasound settings for inflammatory flow: impact on scoring of disease activity in patients with rheumatoid arthritis. Arthritis Rheumatol 2015;67:386–95.

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

27 D'Agostino MA, Terslev L, Aegerter P *et al*. Scoring ultrasound synovitis in rheumatoid arthritis: a EULAR OMERACT ultrasound taskforce-Part 1: definition and development of a standardised, consensus-based scoring system. RMD Open 2017;3:e000428.

28 Terslev L, Naredo E, Aegerter P *et al*. Scoring ultrasound synovitis in rheumatoid arthritis: a EULAR-OMERACT ultrasound taskforce-Part 2: reliability and application to multiple joints of a standardised consensus-based scoring system. RMD Open 2017;3:e000427.

29 D'Agostino M-A, Boers M, Wakefield RJ *et al*. Exploring a new ultrasound score as a clinical predictive tool in patients with rheumatoid arthritis starting abatacept: results from the APPRAISE study. RMD Open 2016;2:e000237.

30 Ammitzbøll-Danielsen M, Østergaard M, Naredo E, Terslev L. Validity and sensitivity to change of the semiquantitative OMERACT ultrasound scoring system for tenosynovitis in patients with rheumatoid arthritis. Rheumatology (Oxford) 2016;55:2156–66.

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

32 van der Maas A, Lie E, Christensen R *et al*. Construct and criterion validity of several proposed DAS28-based rheumatoid arthritis flare criteria: an OMERACT cohort validation study. Ann Rheum Dis 2013;72:1800–5.

33 van der Mass A, den Broeder AA, Measuring flares in rheumatoid arthritis. (Why) Do we need validated criteria? J Rheumatol 2014;41:189–91.

34 Bykerk VP, Lie E, Bartlett SJ *et al*. Establishing a core domain set to measure rheumatoid arthritis flares: report of the OMERACT 11 RA flare Workshop. J Rheumatol 2014;41:799–809.

35 Filippou G, Sakellariou G, Scirè CA *et al*. The predictive role of ultrasound-detected tenosynovitis and joint synovitis for flare in patients with rheumatoid arthritis in stable remission. Results of an Italian multicentre study of the Italian Society for Rheumatology Group for Ultrasound: the STARTER study. Ann Rheum Dis 2018;77:1283–9.

36 Rees JD, Pilcher J, Heron C, Kiely P. A comparison of clinical vs ultrasound determined synovitis in rheumatoid arthritis utilizing gray-scale, power Doppler and the intravenous microbubble contrast agent 'Sono-Vue'. Rheumatology (Oxford) 2007;46:454–9.

37 Janta I, Naredo E, Martinez-Estupinan L *et al*. Patient self-assessment and physician's assessment of rheumatoid arthritis activity: which is more realistic in remission status? A comparison with ultrasonography. Rheumatology (Oxford) 2013;52:2243–50.

38 Hirata A, Ogura T, Hayashi N *et al*. Concordance of patient-reported joint symptoms, physician-examined arthritic signs, and ultrasound-detected synovitis in rheumatoid arthritis. Arthritis Care Res 2017;69:801–6.

39 Szkudlarek M, Court-Payen M, Strandberg C *et al*. Power Doppler ultrasonography for assessment of synovitis in the metacarpophalangeal joints of patients with rheumatoid arthritis: a comparison with dynamic magnetic resonance imaging. Arthritis Rheum 2001;44:2018–23.

40 Hmamouchi I, Bahiri R, Srifi N *et al*. A comparison of ultrasound and clinical examination in the detection of flexor tenosynovitis in early arthritis. BMC Musculoskelet Disord 2011;12:91.

41 Cheung PP, Dougados M, Gossec L. Reliability of ultrasonography to detect synovitis in rheumatoid arthritis: a systematic literature review of 35 studies (1, 415 patients). Arthritis Care Res 2010;62:323–34.

42 Cheung PP, Gossec L, Ruyssen-Witrand A *et al*. The relationship of patient reported joints with active synovitis detected by power Doppler ultrasonography in rheumatoid arthritis. Clin Exp Rheumatol 2013;31:490–7.

43 Bellis E, Scirè CA, Carrara G *et al*. Ultrasound-detected tenosynovitis independently associates with patient-reported flare in patients with rheumatoid arthritis in clinical remission: results from the observational study STARTER of the Italian Society for Rheumatology. Rheumatology (Oxford) 2016;55:1826–36.

44 Levy G, Cheetham C, Cheatwood A, Burchette R. Validation of patient-reported joint counts in rheumatoid arthritis and the role of training. J Rheumatol 2007;34:1261–5.

45 Hecht C, Englbrecht M, Rech J *et al*. Additive effect of anti-citrullinated protein antibodies and rheumatoid factor on bone erosions in patients with RA. Ann Rheum Dis 2015;74:2151–6.

46 Witt M, Mueller F, Nigg A *et al*. Relevance of grade 1 gray-scale ultrasound findings in wrists and small joints to the assessment of subclinical synovitis in rheumatoid arthritis. Arthritis Rheum 2013;65:1694–701.

47 Padovano I, Costantino F, Breban M, D'Agostino MA. Prevalence of ultrasound synovial inflammatory findings in healthy subjects. Ann Rheum Dis 2016;75:1819–23.

48 Garrigues F, Jousse-Joulin S, Bouttier R *et al*. Concordance between clinical and ultrasound findings in rheumatoid arthritis. Joint Bone Spine 2013;80:597–603.

49 Terslev L, Østergaard M, Sexton J, Hammer HB. Is synovial hypertrophy without Doppler activity sensitive to change? Post-hoc analysis from a rheumatoid arthritis ultrasound study. Arthritis Res Ther 2018;20:224.

50 Ammitzbøll-Danielsen M, Østergaard M, Naredo E *et al*. The use of the OMERACT ultrasound tenosynovitis scoring system in multicenter clinical trials. J Rheumatol 2018;45:165–9.