RHEUMATOLOGY

Rheumatology 2019;58:1353–1360
doi:10.1093/rheumatology/key422
Advance Access publication 3 January 2019

Original article

Responsiveness of clinical and ultrasound outcome measures in musculoskeletal systemic lupus erythematosus

Khaled Mahmoud[1,2], Ahmed S. Zayat[1,2,3], Yuzaiful Yusof[1,2], Elizabeth Hensor[1,2], Philip G. Conaghan[1,2], Paul Emery[1,2] and Edward M. Vital[1,2]

Abstract

Objective. To assess the responsiveness of clinical outcome measures in musculoskeletal SLE compared with US.

Methods. A prospective pilot study was conducted in consecutive SLE patients with inflammatory musculoskeletal symptoms. Clinical assessments including SLEDAI, BILAG, 28 tender and swollen joint counts, physician and patient visual analogue scales (VAS), and US were performed at 0, 2 and 4 weeks following 120 mg i.m. methylprednisolone acetate. Responsiveness was analysed using changes and effect sizes using Cohen's criteria.

Results. Twenty patients were recruited. Fifteen out of 20 had clinical swelling at baseline. All clinical and US parameters were significantly improved at week 4 (all $P \leqslant 0.01$). Musculoskeletal-BILAG score improved in 16/20. Musculoskeletal-SLEDAI improved in 7/20. SLE responder index 4 criteria were assessed in 19 patients with SLEDAI $\geqslant 4$ at baseline and were met in 9/19 at 4 weeks. Effect sizes at 4 weeks were large (>0.5) for US, physician VAS and BILAG, and medium (>0.3) for joint counts and SLEDAI. Large effect sizes for improvement in US grey-scale and power Doppler were observed in both SLE responder index 4 responders ($r = -0.51$ and -0.56, respectively) and non-responders ($r = -0.62$ and -0.59, respectively) at 4 weeks.

Conclusion. This is the first study to measure the responsiveness of clinical outcome measures in musculoskeletal SLE against an objective inflammation measure. BILAG and physician VAS were the most responsive clinical instruments. US was highly responsive in musculoskeletal SLE, while SLEDAI and joint counts appeared suboptimal for detection of improvement. These results suggest that clinical trials based on the SLEDAI and SLE responder index 4 may underestimate the efficacy of therapy in SLE.

Key words: systematic lupus erythematosus, lupus arthritis, BILAG, SLEDAI, synovium, tendons and ligaments, outcome measures, ultrasonography

Rheumatology key messages

- US was highly responsive for the musculoskeletal manifestations of SLE.
- Most clinical outcome measures were less responsive than US; SLEDAI and SLE responder index 4 may underestimate response.
- BILAG-2004 and physician visual analogue scales appeared more responsive than SLEDAI-2K and SLE responder index 4 for musculoskeletal SLE.

Introduction

Inflammatory musculoskeletal features are common in SLE, being the first presenting symptom in around 50% of cases and affecting up to 95% of patients at some time [1, 2]. Joint pain in SLE has a significant impact on quality of life and results in loss of function [3–5]. Accordingly, musculoskeletal disease is a common reason for inclusion into clinical trials.

Recent phase III trials of many putative treatments in non-renal SLE have been negative (with the exception of

[1]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, [2]NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds and [3]Department of Rheumatology, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK

Submitted 20 August 2018; accepted 25 November 2018

Correspondence to: Edward M. Vital, Chapel Allerton Hospital, Chapeltown Road, Leeds LS7 4SA, UK. E-mail: e.m.vital@leeds.ac.uk

CLINICAL SCIENCE

belimumab [6]). This has led to questions over the most appropriate outcome measures to use in SLE trials. Also, in clinical practice, it is equally important to differentiate patients with good or incomplete responses to therapy for treat-to-target approaches and to minimize glucocorticoid use [7].

While non-renal SLE trials included many different types of organ involvement, musculoskeletal disease was most common. For example, in the pooled data from the study of belimumab in subjects with SLE, BLISS52 and BLISS76 trials, 1008/1684 (60%) patients had musculoskeletal (MSK)-BILAG A or B at baseline; 991/1684 (59%) had mucocutaneous BILAG A or B; and 272/1684 (16%) had haematology A or B; with lower percentages for other organ systems. In the phase III Efficacy and safety of subcutaneous tabalumab in patients with SLE (ILLUMINATE) study, at baseline, 81% of patients had musculoskeletal activity on the SLEDAI [8].

In SLE, outcome measures must account for disease activity in many different organs. For this reason, less detail is included for each organ compared with more organ-specific instruments such as the 28 joint count used in RA. The SLEDAI and BILAG, and composite endpoints derived from them such as the SLE responder index 4 (SRI-4) and BILAG-based composite lupus assessment, are commonly used in trials.

For musculoskeletal involvement, the SLEDAI-2K [9] is binary, scoring 4 points for tenderness with swelling, effusion, warmth or erythema in two or more joints in the past 30 days, and none for lesser degrees of arthritis. This scoring means that patients with a high level of disease activity at baseline who have a substantial improvement may be considered non-responders. The BILAG-2004 index [10] is semi-quantitative with 4 grades for each active organ system assessed. For the musculoskeletal domain, BILAG A (the highest score) requires observed active synovitis in more than two joints with marked loss of functional range of movements. BILAG B is scored for tendonitis/tenosynovitis or active synovitis in more than one joint (observed or through history) with some loss of functional range of movement (or improving BILAG A disease). BILAG C is scored for inflammatory pain (e.g. with morning stiffness) without synovitis (or improving BILAG B disease). Pain without inflammatory symptoms (e.g. pain that clinically appears to be because of OA) is scored as BILAG D or E, as are patients with previous joint inflammation but no current symptoms. Assigning these grades is dependent on the skill of the assessor, and in both these indices, the assessor must only score symptoms that are deemed to be due to active SLE rather than other pathologies, which is known to be a difficult distinction for arthralgia in many inflammatory arthritides.

Joint counts and visual analogue scales (VAS) have also been used in many SLE trials, but with limited independent validation [11]. Musculoskeletal US provides an objective measure of synovitis that has already been shown to have face and construct validity in SLE [12]. We recently showed that the BILAG and SLEDAI are specific but not sensitive for the detection of synovitis that is US-confirmed and associated with worse symptoms and serological abnormality [13]. These various instruments have never been compared longitudinally.

The objective of this study was therefore to compare the internal responsiveness of a range of clinical outcome measures and US in SLE patients receiving a therapy of known efficacy (glucocorticoids).

Methods

Patients

Twenty patients fulfilling the SLICC 2012 diagnostic criteria [14] for SLE were recruited in Leeds if they had been prescribed 120 mg i.m. methylprednisolone acetate for active musculoskeletal disease that day as part of routine care. This dose and method of administration is commonly used for musculoskeletal flares in the UK. Briefly, other eligibility criteria included: stable doses of NSAID, DMARDs and glucocorticoids (up to prednisolone $\leqslant$5 mg/day or equivalent) for at least 6 weeks prior to entry visit. CCP antibody-positive patients and those with improving disease were excluded. Clinical assessment and US were performed on the day of i.m. glucocorticoid treatment and repeated after 2 and 4 weeks to assess responsiveness. The study was approved by the local ethics committee and informed written consent was obtained from all patients (Leeds East Research Ethics Committee 10/H1306/88). We included all referred patients on an intent-to-treat basis (i.e. we did not withdraw patients based on their baseline clinical and US assessment).

Clinical and laboratory assessment

The clinical assessments were performed by trained rheumatologists who were blinded to the US assessment and were independent of the glucocorticoid treatment decision. SLE was assessed using BILAG-2004 [10] and SLEDAI-2K 30 days [15]. Joint disease was assessed using 28 tender joint counts (TJCs) and swollen joint counts (SJCs), painful joint count, physician musculoskeletal VAS, patient musculoskeletal disease activity VAS and minutes of early morning stiffness (EMS). BILAG-2004 numerical scores were calculated using the formula A = 12, B = 8, C = 1, D/E = 0 [16]. The BILAG-2004 is assessed over the previous 28 days. The SLEDAI-2K and SRI-4 have been validated measuring symptoms over the previous 10 or 30 days [17, 18]. Response to depomedrone is typically seen within a few days. For the purposes of this study we allowed a 5-day window for follow-up study visits and for patients who reported a rapid improvement in symptoms within a few days of the injection and for the majority of the period since the baseline visit to have a 4 week response at the last assessment.

Patients were tested at baseline for routine inflammatory and serological markers. SRI-4 was calculated as previously described [19]. SRI-4 response criteria were met if the patient had: at least a 4-point reduction in the SLEDAI-2K, no worsening in physician VAS and no worsening in BILAG.

US assessment

US [grey-scale (GS) and power Doppler (PD)] was performed using a General Electric Logiq E9 with multi-linear

Table 1 Summaries of clinical and US assessments at weeks 0, 2 and 4

Outcome measure	Week 0	Week 2	Week 4	Change week 2	Change week 4
MSK-BILAG, *n* (%)					
A	7/20 (35)	N/A	1/20 (5)	N/A	
B	8/20 (40)		2/20 (10)		Improved 16/20 (80%)
C	5/20 (25)		9/20 (45)		Same 4/20 (20%)
D			7/20 (35)		Worse 0/20 (0%)
BILAG-MSK (A = 12, B = 8, C = 1, D = 0)	8 (3, 12)	N/A	1 (0, 1)	N/A	−7 (−8, −1)
SLEDAI arthritis present, *n* (%)	19/20 (95)	N/A	10/20 (50)	N/A	Improved 7/20 (35%) Same 13/20 (65%)
SLEDAI arthritis	4 (4, 4)	N/A	2 (0, 4)	N/A	0 (−4, 0)
TJC (0–28)	8 (4, 12)	4 (1, 14)	2 (1, 11)	−3 (−4, 3)	−4 (−6, −1)
SJC (0–28)	2 (0, 5)	0 (0, 1)	0 (0–0)	−1 (−3, 0)	−2 (−3, 0)
Symptomatic joint count	15 (6, 22)	2 (0, 13)	4 (1, 15)	−7 (−19, 0)	−6 (−14, −1)
EMS (min)	25 (0, 60)	5 (0, 45)	3 (0, 41)	0 (−21, 0)	0 (−24, 0)
Patient VAS (mm)	57 (30, 79)	30 (9, 40)	33 (8, 49)	−23 (−29, −10)	−22 (−52, 2)
Physician VAS (mm)	55 (35, 68)	23 (5, 50)	15 (5, 35)	−24 (−45, −15)	−31 (−45, −15)
US—total PD	8 (2, 26)	1 (0, 6)	1 (0, 1)	−8 (−27, −2)	−8 (−10, −2)
US—total GS	19 (9, 43)	13 (5, 24)	8 (2, 13)	−12 (−23, −4)	−10 (−21, −3)
Joints with US synovitis	5.5 (1, 9)	3 (1, 8)	1 (0, 4)	−7 (−10, −3)	−5 (−12, −2)

All values presented are median (interquartile range) unless otherwise stated. MSK-BILAG: musculoskeletal element of BILAG; N/A: not applicable; TJC: tender joint count in 28 joints; SJC: swollen joint count in 28 joints; symptomatic joint count: number of joints indicated as painful or stiff by patients on a graphical questionnaire; EMS: early morning stiffness; PD: total US power Doppler score; GS: total US greyscale score; joints with US synovitis: number of joints scoring either GS >1 or PD >0; VAS: visual analogue scales.

6–15 MHz transducer. Two sonographers were trained and experienced in musculoskeletal US and were blinded to patients' clinical evaluation and also independent of the glucocorticoid treatment decision. PD was assessed with the highest gain level without background noise, pulse repetition frequency of 750 Hz and medium wall filter.

Bilateral hands and wrists were scanned. All joints in the hand and wrists were examined using the standard approach of examining the following: radio-carpal, inter-carpal and ulnar-carpal joints, first to fifth MCP joints and first to fifth PIP joints. Bilateral tendon sheaths including the extensor carpi ulnaris and second to fifth flexor digitorium tendon sheaths were assessed for the presence of tenosynovitis. The synovitis GS and PD were scored using the OMERACT definitions and proposed semiquantitative 0–3 scale [20–22]. The GS scoring was: 0 = no synovial hypertrophy, 1 = mild hypertrophy, 2 = moderate hypertrophy and 3 = severe hypertrophy. The PD scoring was: 0 = absence of signal, no intra-articular flow, 1 = mild hyperaemia, one or two vessels signal (including one confluent vessel), 2 = moderate hyperaemia, (>grade 1) and <50% of GS area and 3 = marked hyperaemia, vessels signal in more than half of the synovial area. Tenosynovitis was defined according to the OMERACT criteria [22] and the GS and PD signal scored using semi-quantitative 0–3 scale system (0 = normal, 1 = mild, 2 = moderate and 3 = severe) [23]. US abnormalities (62 areas) were summarized as total GS, PD, erosions and tenosynovitis, as well as numbers of joints with abnormal GS (⩾2) or PD (⩾1), erosions or tendons with tenosynovitis (as any GS and/or PD abnormality in the tendon sheath).

Statistical analysis

Overall clinical characteristics (demographics, therapies, clinical joint assessments and immunological parameters) and US characteristics were summarized for each group using proportions of patients or median and interquartile range as appropriate.

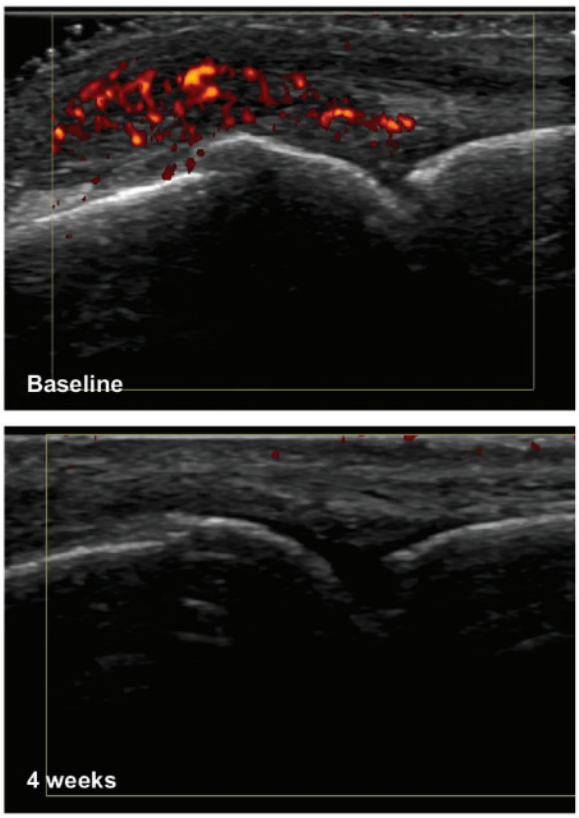
A variety of methods have been used to calculate effect sizes to measure internal responsiveness. Standardized response means may be used for parametric variables. The candidate outcome measures in this study included parametric, ordinal and categorical variables. We therefore used effect sizes calculated from a paired non-parametric test instead of paired t-tests as usually used to calculate effect size statistics [24]. Change in continuous variables was assessed using Wilcoxon Signed Ranks test. Effect size was calculated using standardized test statistic, Z, using the formula $r = Z/\mathrm{sqrt}(n1 + n2)$. Effect sizes were judged using Cohen's Criteria as large (>−0.5), medium (>−0.3) or small (>−0.1) [25].

Results

Baseline characteristics

All 20 patients were female and ANA positive. Mean (S.D.) age was 49.7 years (14.1) and mean (S.D.) disease duration 85 months (22). Eleven of 20 (55%) were receiving NSAID therapy. Fourteen were on stable-dose hydroxychloroquine, of whom three were also on stable-dose MTX or MMF and one was on epratuzumab. Three patients received MTX or MMF without HCQ. Three patients

Fig. 1 Example US images



US images of MCP joint in an SLE patient at baseline and 4 weeks. Baseline image shows grade 3 power Doppler, which has completely resolved at 4 weeks.

were not on HCQ or oral immunosuppressants. Three patients received stable-dose prednisolone $\leqslant 5$ mg/day. Fifteen of 20 patients had clinical joint swelling at baseline. The others all had either US synovitis (GS in 18/20, PD in 17/20) or >60 min of EMS, or new activity in other organ systems coincident with the onset of joint pain.

Changes in outcome measures

At 4 weeks there was a substantive and significant improvement in all clinical and US parameters measured (all $P < 0.025$, Table 1). However, 65% of patients still had symptoms with BILAG A–C. Sixteen of 20 patients had improvement by at least one MSK-BILAG grade, but only 7/20 had improvement in the musculoskeletal SLEDAI component. Residual symptoms were confirmed by TJC and symptomatic joint count, morning stiffness, and patient and physician VAS. On 4-week US there was a large reduction in PD. PD was present in nine patients at 4 weeks, but with a total score of <2 in eight of these (Fig. 1). GS scores were significantly reduced but higher than PD post-treatment, being present at $\geqslant 2$ in 13/20 patients.

Changes in these parameters at 2 weeks were more variable. TJCs, SJCs and EMS minutes had numerically, but not statistically, significantly improved. VAS showed a partial but significant improvement. US parameters had all significantly improved at 2 weeks, although to lesser degree than at 4 weeks.

Effect sizes ordered according to magnitude are shown in Table 2. At both 2 and 4 weeks, physician VAS had the largest effect size, although it must be noted that this assessment was not blinded to time point and may be more susceptible to observer bias than the other variables. Other than physician VAS, at 2 weeks only changes in US showed large effect sizes. Changes in clinical variables were only small–medium.

At week 4, effect sizes remained large for US and physician VAS. They were medium for other clinical variables (joint counts, EMS, patient VAS). Effect sizes for musculoskeletal components of BILAG and SLEDAI differed: the effect for MSK-BILAG was of a similar magnitude to US. Although the MSK-SLEDAI significantly improved, its effect size was substantially smaller than for BILAG, US and physician VAS.

Comparison of SLEDAI responders and non-responders

The 19 patients with an MSK-SLEDAI score of at least 4 points at baseline were grouped into SRI-4 responders ($n = 9$) and SRI-4 non-responders ($n = 10$). SRI-4 and change in MSK-SLEDAI were generally equivalent in this patient group. All SRI-4 responders also had improvement in the musculoskeletal component of the SLEDAI except for one who improved in other organ domains and had a mixed response in musculoskeletal variables. All SRI-4 non-responders did not have improvement in the musculoskeletal component of the SLEDAI. Full data are show in supplementary Table S1, available at *Rheumatology* online.

We then compared change in TJCs and SJCs, and US GS and PD in each of these groups (Fig. 2). For TJCs and SJCs there were large effect sizes in SRI-4 responders ($r = -0.505$ and -0.492, and $P = 0.024$ and 0.028, respectively) and medium effect sizes in SRI-4 non-responders ($r = -0.365$ and -0.331, and $P = 0.122$ and 0.160, respectively). For US, large effect sizes for improvements in both GS and PD were observed in both SRI-4 responders ($r = -0.517$ and -0.564, and $P = 0.021$ and 0.021, respectively) and SRI-4 non-responders ($r = -0.629$ and -0.596, and $P = 0.008$ and 0.012, respectively). In many cases the size of the improvements in SRI-4 non-responders was large. For example, a 70% improvement was seen in 30, 60, 40 and 70% of patients for TJCs, SJC, US GS and US PD, respectively (supplementary Table S2, available at *Rheumatology* online).

Discussion

In this study, we compared the internal responsiveness of clinical outcome measures and US in SLE in patients receiving a known efficacious therapy. All commonly used clinical variables significantly improved by week 4 but there was variation in responsiveness between them. BILAG-2004 and physician VAS had similar responsiveness to

TABLE 2 Effect sizes for change at 2 and 4 weeks according to magnitude

Outcome measure	No. pairs	*P*	*Z*	Effect size	Cohen criteria
Week 0–week 2					
Physician VAS	16	0.001	−3.409	−0.603	Large
GS score	16	0.002	−3.13	−0.571	Large
No. joints with US synovitis	16	0.011	−3.160	−0.559	Large
PD score	16	0.002	−3.099	−0.548	Large
Symptomatic joint count	10	0.047	−1.988	−0.445	Medium
Patient VAS	16	0.016	−2.409	−0.426	Medium
EMS (min)	16	0.046	−1.997	−0.353	Medium
SJC	16	0.059	−1.889	−0.334	Medium
TJC	15	0.274	−1.093	−0.200	Small
Week 0–week 4					
Physician VAS	20	<0.001	−3.388	−0.593	Large
MSK-BILAG numeric	20	0.008	−3.643	−0.576	Large
PD score	20	<0.001	−3.627	−0.573	Large
No. joints with US synovitis	20	0.001	−3.627	−0.573	Large
GS score	20	<0.001	−3.503	−0.554	Large
Symptomatic joint count	14	0.010	−2.576	−0.487	Medium
MSK-SLEDAI score	20	0.003	−3.000	−0.474	Medium
TJC	20	0.007	−2.683	−0.424	Medium
EMS (min)	20	0.012	−2.527	−0.400	Medium
SJC	20	0.007	−2.425	−0.383	Medium
Patient VAS	20	0.020	−2.331	−0.369	Medium

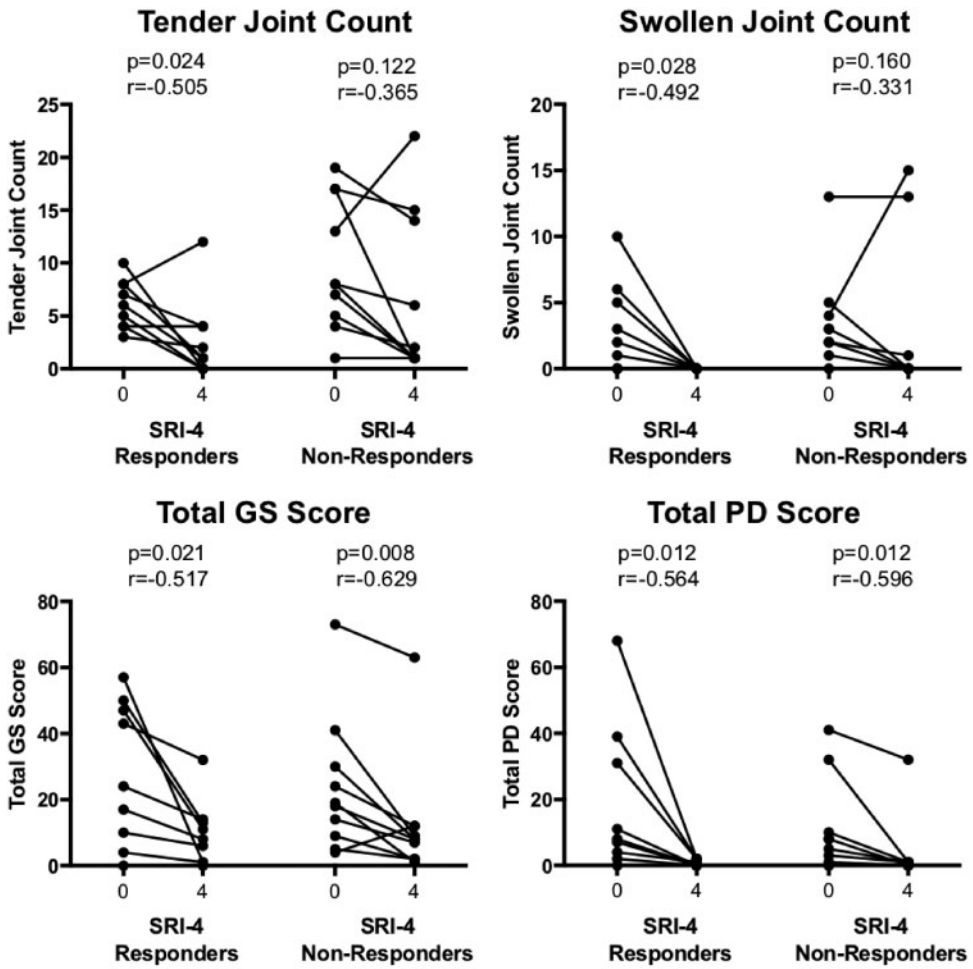
MSK-BILAG numeric calculated using A = 12, B = 8, C = 1, D = 0. MSK-SLEDAI score calculated using arthritis present in previous 30 days = 4, arthritis absent = 0. *P*-values are results of Wilcoxon signed ranks test, *Z*: standardized test statistic, effect size *r* calculated as *r* = Z/sqrt(2N). MSK-BILAG: musculoskeletal element of BILAG; MSK-SLEDAI: musculoskeletal element of SLEDAI; EMS: early morning stiffness; GS score: total ultrasound grey scale score; PD score: total US power Doppler score; VAS: visual analogue scales; SJC: swollen joint count; TJC: tender joint count.

US, but are more susceptible to observer bias. SRI-4 underestimated response, with substantial objective improvements in synovitis in SRI-4 non-responders. If replicated in larger studies these results may have implications for the design of clinical trials in SLE as well as routine clinical practice.

A dilemma in clinical trials in SLE has been that many therapies that appear to be effective in other contexts have produced negative Randomized Control Trials. There are many possible reasons for this, including the recruitment of some ANA-negative patients and use of active comparator arms. However, there are reasons to believe that choice of outcome measures is at least partly responsible for these discrepancies in the evidence base. In the belimumab programme, phase II data using the Safety of Estrogen in Systemic Lupus Erythematosus National Assessment (SELENA)-SLEDAI were negative [26]. The SRI was derived from these phase II data and used to design a phase III trial, which produced the opposite result [18]. The rituximab Efficacy and safety of rituximab in moderately-to-severely active SLE (EXPLORER) study was negative for its BILAG-based primary and secondary endpoints, but had positive results in *post hoc* analyses such as BILAG A flare rate [27, 28]. While the SRI has been highly successful in several clinical trials, the response rates in the two phase III trials of belimumab were rather low at 43–58% *vs* 34–44% for belimumab and placebo, respectively [29]. The data we report here show that the SRI-4 underestimates clinical improvement in patients with arthritis and therefore may suggest that clinical trials would show higher response rates and greater differentiation of active and placebo arms if imaging outcome measures, or more reponsive clinical outcome measures, were used.

For effective treatment of SLE in the clinic it is essential to be able to measure disease activity accurately, especially when using biologic therapies. An international task force recently recommended treating to a target of low disease activity in SLE, as well as minimizing glucocorticoid exposure [30]. The low disease activity target was recommended to use a validated lupus activity index and/or organ-specific markers. Our results suggest that choice of definition of disease activity could alter treatment decisions, although this needs to be confirmed in longitudinal studies. For example, the UK National Institute of Health and Clinical Excellence criteria for belimumab mandate that treatment should only be continued if there is at least a 4-point reduction in the SLEDAI [31]. Our data indicate that patients with musculoskeletal disease not achieving this 4-point reduction may still have clinically meaningful improvement, and physician VAS data suggest that overall physician judgement may be a better guide to response. Nevertheless, many other studies show that patients with musculoskeletal symptoms but not clinical joint swelling (not meeting BILAG A/B or SLEDAI criteria) may have subclinical synovitis [12].

Fig. 2 Change in joint counts and US and SRI-4 response



Patients who had a MSK-SLEDAI score of 4 points at baseline were grouped according to whether they met the SRI-4 response criterion at the 4 week follow-up. *P*-values show the results of a Wilcoxon matched pairs test within each group and effect size *r*. PD: total US power Doppler score; GS: total US grey-scale score; SRI-4: SLE responder index 4.

Hence in forming their judgement of response physicians may wish to consider US in patients with ongoing inflammatory symptoms despite a degree of improvement.

Physician VAS appeared to be highly responsive in this study. It must be noted that assessors were not blinded to time point and this may affect subjective outcome measures due to observer bias. Observer bias may also affect the BILAG 'improving' score, wherein synovitis that is still present but determined to be improving results in a lower BILAG score than if it is deemed to be stable or worsening. Furthermore, the BILAG is affected by the skill and experience of the assessor. All our assessments were performed by trained assessors experienced in SLE clinical trials. An advantage of US is that it is more objective. However, it is operator-dependent and may be more difficult to standardize in multicentre studies. Joint counts were not as responsive as other instruments here, but are easier to standardize in multicentre studies given their widespread use in other inflammatory arthritis.

When first developed, the BILAG and SLEDAI were validated against the physician's intention-to-treat and judgement of overall disease activity. In our cross-sectional study we noted that US synovitis is common in patients without joint swelling and no clinical instrument could detect this. This suggests that validation against an objective measure of disease activity would be more valuable. Although there is no other study focusing specifically on musculoskeletal disease, one previous study compared the sensitivity to change of five clinical instruments for overall disease activity (SLAM, SLEDAI, BILAG, ECLAM and Lupus Activity Index) [32]. Similar to our study, in that paper the SLEDAI was less responsive that the BILAG.

Our results suggest that an organ-specific outcome measure may be more valuable in this common manifestation. This has already been established in the other most common manifestation of SLE: cutaneous disease. The Cutaneous Lupus Activity and Severity Index (CLASI)

[33] provides an organ-specific, continuous measure of cutaneous disease activity. In recent clinical trials of sifalimumab and anifrolumab, the CLASI showed a high rate of responsiveness [34, 35]. In our study, physician VAS was more responsive than the musculoskeletal component of the SLEDAI. Tender, swollen and symptomatic joint counts had similar responsiveness to the SLEDAI, but may be advantageous in multicentre trials in being less dependent on the experience and opinion of the assessor, and less susceptible to observer bias. The data in this study and our previous larger cross-sectional study demonstrate that joint counts and US findings vary more than BILAG and SLEDAI grades. It is therefore likely that a composite outcome measure could be designed for musculoskeletal disease that offers similar advantages to the CLASI. This is being determined in our future research. One previous paper has also shown the potential advantages of specific musculoskeletal outcome measures in patients treated with belimumab [36]. The CLASI and joint counts have also revealed nuances of response in individual organ domains in patient subgroups after belimumab therapy [37].

Our study has some limitations. Patient numbers were relatively small. We used a single-centre design; this may be important for tools that require training (e.g. BILAG) or inter-observer standardization (US). Assessors were not blinded to therapy or time point, which may have affected some instruments. However, clinical and US assessors were blinded to each other's findings. Lastly, we have not yet assessed external responsiveness—i.e. responsiveness compared with some external anchor [24].

Despite these limitations, our results are unique in comparing responsiveness to an objective standard and indicate the limitations of existing tools for musculoskeletal lupus. Our results suggest that an organ-specific outcome measure for musculoskeletal disease would have advantages in both clinical trials and routine clinical practice. This is being definitively assessed in a larger longitudinal study currently in progress.

Acknowledgements

The views expressed are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health. K.M. was funded by the University of Benghazi. Y.Y. is an NIHR doctoral research fellow and E.M.V. is an NIHR Clinician Scientist.

Funding: This research was supported by the National Institute for Health Research Leeds Biomedical Research Centre based at Leeds Teaching Hospitals NHS Trust. E.M.V.'s research was supported by a grant from the National Institute of Health Research (CS-2013–13-032).

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

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

1 Merrill JT, Neuwelt CM, Wallace DJ *et al*. Efficacy and safety of rituximab in moderately-to-severely active systemic lupus erythematosus: the randomized, double-blind, phase II/III systemic lupus erythematosus evaluation of rituximab trial. Arthritis Rheum 2010;62:222–33.

2 Zoma A. Musculoskeletal involvement in systemic lupus erythematosus. Lupus 2004;13:851–3.

3 Drenkard C, Bao G, Dennis G *et al*. Burden of systemic lupus erythematosus on employment and work productivity: data from a large cohort in the southeastern United States. Arthritis Care Res (Hoboken) 2014;66:878–87.

4 Eilertsen GO, Nikolaisen C, Becker-Merok A, Nossent JC. Interleukin-6 promotes arthritis and joint deformation in patients with systemic lupus erythematosus. Lupus 2011;20:607–13.

5 Mahmoud K, Zayat A, Vital EM. Musculoskeletal manifestations of systemic lupus erythmatosus. Curr Opin Rheumatol 2017;29:486–92.

6 Furie R, Petri M, Zamani O *et al*. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. Arthritis Rheum 2011;63:3918–30.

7 van Vollenhoven RF, Mosca M, Bertsias G *et al*. Treat-to-target in systemic lupus erythematosus: recommendations from an international task force. Ann Rheum Dis 2014;73:958–67.

8 Isenberg DA, Petri M, Kalunian K *et al*. Efficacy and safety of subcutaneous tabalumab in patients with systemic lupus erythematosus: results from ILLUMINATE-1, a 52-week, phase III, multicentre, randomised, double-blind, placebo-controlled study. Ann Rheum Dis 2016;75:323–31.

9 Gladman DD, Ibanez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. J Rheumatol 2002;29:288–91.

10 Isenberg D, Rahman A, Allen E *et al*. BILAG 2004. Development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. Rheumatology (Oxford) 2005;44:902–6.

11 Cipriano E, Ceccarelli F, Massaro L *et al*. Joint involvement in patients affected by systemic lupus erythematosus: application of the swollen to tender joint count ratio. Reumatismo 2015;67:62–7.

12 Zayat AS, Md Yusof MY, Wakefield RJ *et al*. The role of ultrasound in assessing musculoskeletal symptoms of systemic lupus erythematosus: a systematic literature review. Rheumatology (Oxford) 2015;55:485–94.

13 Zayat AS, Mahmoud K, Md Yusof MY *et al*. Defining inflammatory musculoskeletal manifestations in systemic lupus erythematosus. Rheumatology (Oxford) 2019;58:304–12.

14 Petri M, Orbai AM, Alarcon GS *et al*. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64:2677–86.

15 Bombardier C, Gladman DD, Urowitz MB *et al*. Derivation of the SLEDAI. A disease activity index for lupus patients.

The Committee on Prognosis Studies in SLE. Arthritis Rheum 1992;35:630–40.

16 Yee C-S, Bruce I, Cresswell L *et al*. Numerical scoring for the BILAG-2004 index. Rheumatology (Oxford) 2010;49:1665–9.

17 Touma Z, Urowitz MB, Ibañez D, Gladman DD. SLEDAI-2K 10 days versus SLEDAI-2K 30 days in a longitudinal evaluation. Lupus 2011;20:67–70.

18 Furie RA, Petri MA, Wallace DJ *et al*. Novel evidence-based systemic lupus erythematosus responder index. Arthritis Rheum 2009;61:1143–51.

19 Luijten KM, Tekstra J, Bijlsma JW, Bijl M. The Systemic Lupus Erythematosus Responder Index (SRI); a new SLE disease activity assessment. Autoimmun Rev 2012;11:326–9.

20 Naredo E, Collado P, Cruz A *et al*. Longitudinal power Doppler ultrasonographic assessment of joint inflammatory activity in early rheumatoid arthritis: predictive value in disease activity and radiologic progression. Arthritis Rheum 2007;57:116–24.

21 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.

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

23 Naredo E, D'Agostino MA, Wakefield RJ *et al*. Reliability of a consensus-based ultrasound score for tenosynovitis in rheumatoid arthritis. Ann Rheum Dis 2013;72:1328–34.

24 Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000;53:459–68.

25 Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum, 1988.

26 Wallace DJ, Stohl W, Furie RA *et al*. A phase II, randomized, double–blind, placebo–controlled, dose–ranging study of belimumab in patients with active systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2009;61:1168–78.

27 Merrill J, Buyon J, Furie R *et al*. Assessment of flares in lupus patients enrolled in a phase II/III study of rituximab (EXPLORER). Lupus 2011;20:709–16.

28 Merrill JT, Neuwelt CM, Wallace DJ *et al*. Efficacy and safety of rituximab in moderately–to–severely active systemic lupus erythematosus: the randomized, double-blind, phase II/III systemic lupus erythematosus evaluation of rituximab trial. Arthritis Rheumatol 2010;62:222–33.

29 Furie R, Petri M, Zamani O *et al*. A phase III, randomized, placebo–controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. Arthritis Rheumatol 2011;63:3918–30.

30 van Vollenhoven RF, Mosca M, Bertsias G *et al*. Treat-to-target in systemic lupus erythematosus: recommendations from an international task force. Ann Rheum Dis 2014.

31 NICE. Belimumab for treating active autoantibody-positive systemic lupus erythematosus. 2016. Technology appraisal guidance [TA397], https://www.nice.org.uk/guidance/ta397.

32 Ward MM, Marx AS, Barry NN. Comparison of the validity and sensitivity to change of 5 activity indices in systemic lupus erythematosus. J Rheumatol 2000;27:664–70.

33 Klein R, Moghadam-Kia S, LoMonico J *et al*. Development of the CLASI as a tool to measure disease severity and responsiveness to therapy in cutaneous lupus erythematosus. Arch Dermatol 2011;147:203–8.

34 Furie R, Khamashta M, Merrill JT *et al*. Anifrolumab, an anti-interferon–α receptor monoclonal antibody, in moderate-to-severe systemic lupus erythematosus. Arthritis Rheumatol 2017;69:376–86.

35 Khamashta M, Merrill JT, Werth VP *et al*. Sifalimumab, an anti-interferon-α monoclonal antibody, in moderate to severe systemic lupus erythematosus: a randomised, double-blind, placebo-controlled study. Ann Rheum Dis 2016.

36 Iaccarino L, Andreoli L, Bocci EB *et al*. Clinical predictors of response and discontinuation of belimumab in patients with systemic lupus erythematosus in real life setting. Results of a large, multicentric, nationwide study. J Autoimmun 2018;86:1–8.

37 Parodis I, Gomez A, Frodlund M *et al*. Smoking reduces the efficacy of belimumab in mucocutaneous lupus. Expert Opin Biol Ther 2018;18:911–20.