A study of flare assessment in systemic lupus erythematosus (SLE) based on paper patients

Isenberg D[1](MD), Sturgess J[2](PhD), Allen E[2](PhD), Aranow C[3](MD), Askanase A[4](MD), Sang-Cheol B[5](MD), Bernatsky S[6](MD), Bruce I[7](PhD), Buyon J[8](MD), Cervera R[9](MD), Clarke A[10](MD), Dooley Mary Anne[11](MD), Fortin P[12](MD), Ginzler E[13](MD), Gladman D[14](MD), Hanly J[15](MD), Inanc M[16](MD), Jacobsen S[17](MD), Kamen D[18](MD), Khamashta M[19](FRCP), Lim S[20](MD), Manzi S[21](MD), Nived O[22](MD), Peschken C[23](MD), Petri M[24](MD), Kalunian K[25](MD), Rahman A[1](PhD), Ramsey-Goldman R[26](MD), Ruiz-Irastorza G[27](MD), Sanchez-Guerrero J[28](MD), Steinsson K[29](MD), Sturfelt G[22](MD), Urowitz M[14](MD), van Vollenhoven R[30](MD), Wallace DJ[31](MD), Zoma A[32](PhD), Merrill J[33](MD), Gordon C[34](PhD).

Affiliations:

[1] Centre for Rheumatology, University College London, London, UK

[2]Department of Statistics, The Hospital For Tropical Diseases, London, UK

[3]Feinstein Institute for Medical Research, Manhasset, New York, UK

[4]Rheumatology, Columbia University, New York, US

[5]Hanyang University Hospital for Rheumatic Diseases, Seoul, South Korea

[6]Department of Medicine, McGill University, Quebec, Canada

[7] Arthritis Research UK Epidemiology Unit, Institute of Inflammation and Repair, The University of Manchester, and NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

[8]New York School of Medicine, New York, US

[9] Department of Medicine, Universitat de Barcelona. Barcelona, Catalonia, Spain

[10] Division of Rheumatology, Cumming School of Medicine, University of Calgary, Canada

[11] Rheumatology, University of North Carolina, US12 Infectious and immune diseases

[12] Centre Hospitalier de l'Université Laval (CHUL), Québec, Canada

[13] Downstate Medical Center Rheumatology, Brooklyn, New York, US

[14] Lupus Program, Centre for Prognosis Studies in The Rheumatic Disease and Krembil Research Institute, Toronto Western Hospital, University of Toronto, Toronto, Ontario, Canada

[15] Division of Rheumatology, Nova Scotia Rehabiliation Center, Halifax, Canada

[16] Department of Internal Medicine, Instanbul University, Istanbul, Turkey

[17] Copenhagen Lupus and Vasculitis Clinic, Centre For Rheumatology and Spine Diseases, Rigshospitalet, Copenhagen, Denmark

[18] Division of Rheumatology and Immunology, Medical University of South Carolina, Charleston, US

[19] Division of Women's Health, King's College London, UK

[20] Department of Medicine, Emory University, Atlanta, US

[21] Allegheny Health Network, Pittsburgh, US

[22] Department of Rheumatology, Lund University, Lund, Sweden

[23] Department of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada

[24] Johns Hopkins Lupus Centre, Johns Hopkins University, Baltimore, US

[25] University of California, San Diego, La Jolla, US

[26] Northwestern University, Feinberg School of Medicine, Chicago, US

[27] Autoimmune Diseases Research Unit, Department of Internal Medicine, BioCruces Health Research Institute. Hospital Universitario Cruces, University of the Basque Country, Barakaldo, Bizkaia (Spain)

[28] Mount Sinai Hospital and University Health Network, University of Toronto, Canada

[29] Department of Rheumatology, Landspitali University Hospital, Reykjavik, Iceland

[30] Rheumatology Unit, Department of Medicine, Karolinska University Hospital, Solna, Sweden

[31] Department of Medicine, UCLA, California, US

[32] Dept of Rheumatology Hairmyres Hospital, East Kilbride, Scotland, UK

[33] Clinical Pharmacology, Oklahoma Medical Research Foundation, Oklahoma City, US

[34]Rheumatology Research Group, Institute of Inflammation and Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

Correspondence to: Prof David Isenberg, Centre for Rheumatology, University College London, London, UK

Email: d.isenberg@ucl.ac.uk

Keywords: flare, lupus, BILAG 2004, SELENA,

ABSTRACT

OBJECTIVE:

To determine the level of agreement of disease flare severity (distinguishing severe, moderate and mild flare and persistent disease activity) in a large paper patient exercise involving 988 individual cases of systemic lupus erythematosus.

METHODS:

988 individual lupus case histories were assessed by three individual physicians. Complete agreement about the degree of flare (or persistent disease activity) was obtained in 451 cases (46%) and these provided the reference standard for the second part of the study. This component utilised three flare activity instruments (BILAG 2004, SELENA flare index (SFI) and the revised SELENA flare index (rSFI)). The 451 patient case histories were distributed to 18 pairs of physicians being carefully randomised in a manner designed to ensure a fair case mix and equal distribution of flare according to severity.

RESULTS:

The three physician assessment of flare matched the level of flare using the three indices thus 67% for BILAG 2004, 72% for SFI and 70% for rSFI. The corresponding weighted kappas for each instrument were 0.82, 0.59 and 0.74 respectively. We undertook a detailed analysis of the discrepant cases and several factors emerged including a tendency to score moderate flares as severe and persistent activity as flare especially when the SFI and rSFI instruments were used. Overscoring was also driven by scoring treatment change as flare even if there were no new or worsening clinical features.

CONCLUSIONS:

Given the complexity of assessing lupus flare, we were encouraged by the overall results reported. However the problem of capturing lupus flare accurately is not completely solved.

SIGNIFICANCE AND INNOVATION

This study addresses the ongoing dilemma of how best to capture flare in patients with SLE. In our previous attempt to do this using 16 “live” patients we could only assess a modest number of lupus manifestations. In the current study we have utilised nearly 1,000 paper-based case histories in order to determine the capacity of three flare activity instruments (BILAG 2004, SELENA flare index (SFI) and the revised SELENA flare index (rSFI) to capture flare. We show that all three instruments are able to do this in many cases but there is an ongoing need to do even better.

Introduction:

In the past 30 years methods of assessing disease activity in patients with lupus have improved considerably. Both global score systems, such as the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI), Systemic Lupus Activity Measures (SLAM) and the European Community Lupus Activity Measure (ECLAM), and the British Isles Lupus Assessment Group (BILAG) system, which focuses on individual organs or systems, have emerged as viable and effective activity assessment instruments. SLEDAI and BILAG have been revised to SLEDAI-2K [1] and BILAG-2004 index [2], respectively, with large-scale studies undertaken to demonstrate their validity, reliability and sensitivity to change [reviewed in 3].

Although both the SLEDAI and BILAG activity assessments are widely used in large-scale international trials of new biologic drugs [e.g. 4, 5, 6], there are few data assessing their usefulness in determining clinical flares in patients with lupus [7]. The Lupus Foundation of America held two investigator meetings in 2007 and 2008 which sought to agree the definition of flare in lupus patients [8]. They concluded that flare is 'a measurable increase in disease activity in one or more organ systems involving new or worse clinical signs and symptoms and/or laboratory measurements. It must be considered clinically significant by the assessor and usually there will be at least consideration of initiation or increase in treatment.'

A particular challenge in patients with lupus has been distinguishing mild, moderate and severe flares and distinguishing them from 'ongoing', persistent disease. This problem is in part related to difficulties in agreeing what constitutes such flares in different organs and systems, but also because when a patient is flaring there may be a difference in the degree of flare in different organs or systems.

A live patient study of 16 flaring patients with lupus [9] took place in London in May 2009. In this study three assessment instruments were utilised to assess flare. One was based on the BILAG 2004, a second was the classic SELENA Flare Index (SFI), and the third was the revised SELENA flare index (rSFI: an organ-based system that is based on, but not directly linked to, the SLEDAI).

In that live patient study [9], a panel of rheumatologists [one of whom was the patient's own clinician] determined the severity of flare in each patient and then different individual rheumatologists assessed each patient for flare with all 3 instruments. Intra-class correlation co-efficients (with 95% confidence interval) were calculated to indicate a measure of internal reliability. The results were 0.54 (0.32 to 0.78) for the BILAG 2004 flare assessment compared with 0.21 (0.08 to 0.48) for the revised SELENA flare instrument and 0.18 (0.06 to 0.45) for the Physicians Global Assessment of flare. Severe flare was associated with good agreement between the three instruments, but the mild to moderate flares were less consistent. Although assessing real patients offers the tangible advantage of testing potential flare instruments in a more realistic way, obviously the numbers of patients (and their clinical problems) that can be studied at any given time is restricted. In order to expand our experience of flare assessment in lupus, we have now undertaken a major paper patient-based exercise which has allowed us to review a much broader array of lupus symptomatology and to utilize a larger number of assessors to determine flare status using individual instruments on paper case reports based on real cases.

The objective of this study was to determine the level of agreement of flare severity (severe/moderate/mild/none: ie persistent disease activity) identified in paper patient cases using three flare instruments and physician defined flares determined by a panel of three physicians.

Methods:

Generation of Clinical Case scenarios

Participating physicians were given a standardised report form to complete which included four sections: previous lupus assessment and treatment; details of assessment at visit being evaluated for flare; results of relevant investigations (blood, urine, imaging, biopsies etc) influencing this assessment; treatment changes at this visit.

Thirty physicians submitted a total of 988 anonymous individual paper case reports based on the medical records of their patients. Each patient met the revised classification criteria of the American College of Rheumatology [10] and/or the 2012 Systemic Lupus International Collaborating Clinics (SLICC) criteria [11]. The submitting physician was also asked to provide their assessment of the flare category for the paper patient case (severe, moderate or mild flare or "persistent/ongoing" disease) and to submit roughly equal numbers of each category. The actual distribution of paper patient cases received was: 25% severe, 28% moderate, 22% mild, 25% persistent disease activity without flare.

Each patient case report was then assessed for flare category by a further two reviewing physicians who did not know the patient. The cases were allocated at random and randomisation was designed to ensure an equal distribution of submitted category type and to ensure no physician reviewed their own submitted cases or the same case twice. The flare category assigned by each of the three physicians (one submitting and two reviewing) was then compared.

All three physicians agreed on the flare category in 451 cases (46%). We refer to this flare assessment as three physician consensus (TPC) and these are the cases that were carried forward to be assessed by the flare instruments in the study. The TPC result for each case formed the reference standard for our later analyses. Flare category distribution in these cases was: 32% severe, 21% moderate, 24% mild, 23% persistent disease not flare.

TPC was not achieved in 535 cases (54%). In 376 cases (38%) there was partial agreement i.e. any two physicians agreed with each other (but not with the third physician), in 41 cases (4%) there was no agreement between any physician i.e. all three physicians recorded different levels of flare or persistent activity for these cases. One or both of the reviewing physicians rejected 118 cases (12%)

as “unable to code” on the information provided. Two of us (DAI and CG) reviewed these cases to assess whether there were any particular reasons why it has been hard to achieve TPC.

The 451 TPC patient case histories were carefully reviewed by one of us (DAI) and assigned into one of eight clinical groups thus: those with musculo-skeletal and/or skin disease only; joint and/or skin and renal disease; mainly serositis; mainly renal; mainly gastro-intestinal; mainly CNS; joints and/or skin plus serositis, and other which included those that were predominantly haematological and/or constitutional or other combinations. Two cases were excluded from further assessment as one was found to be a duplicate and the other case report form had been completed incorrectly.

Assessment of TPC paper case reports using standard instruments

The 451 TPC patient case histories were distributed to 18 pairs of physicians. The cases were randomised in a manner designed to ensure each pair received a fair case-mix, based on the clinical groups set out above and an equal distribution of flare by category type.

The 18 pairs of rheumatologists were each asked to agree on the level of flare in, between 20 and 26 individual paper cases. Each pair was assigned to use one of the three flare assessment instruments, these were the BILAG-2004 index [12] (6 pairs) or the SFI [13] (6 pairs) and the rSFI [9] (6 pairs). Full details of the flare instruments are shown as supplementary material. Analyses were done both including and excluding flare defined only by treatment change.

The pairs of rheumatologists were asked to undertake the assessments of flare using the instrument assigned to them separately and then to confer to achieve full agreement on the score and level of flare using that instrument or to record persistent activity without flare.

Statistical methods used:

For each flare instrument, the level of flare agreed by the assessing pair of physicians was compared to the TPC flare assessment (the reference standard) for each case. The level of agreement between the flare instrument used and the TPC flare assessment was determined in two ways; firstly by simply calculating the percentage of cases where there was complete agreement and secondly by calculating a weighted kappa with quadratic weights (equivalent to an intraclass correlation coefficient). The weighted Kappa give weights to the frequencies in each cell of the table according to their distance from the diagonal, thus allowing different levels of agreement to contribute.

Results:

Of the 451 cases 7 were rejected by the pairs of physicians for providing too little information to score using the assigned flare instrument (6 cases for BILAG 2004 index and 1 for SELENA-SLEDAI flare instrument).

Tables 1, 2 and 3 show the results for each type of flare assessed by the BILAG 2004, SFI and the rSFI respectively and compared with the TPC level of flare (note that different clinical cases were assessed by each instrument so the indices cannot be directly compared with each other).

For all three flare instruments, agreement on the level of flare as assessed by each instrument and by the TPC assessment occurred in a similar proportion of cases. The level of flare matched the TPC assessment of flare in the individual cases in a similar proportion of times using the original versions of these instruments: 67% for BILAG-2004 index, 72% for SFI and 70% for rSFI. The corresponding weighted kappas for each instrument were BILAG 2004 index 0.82, SFI instrument 0.59, rSFI 0.74.

An analysis of the discrepant cases where there was a difference in the assessment of flare between the TPC and the flare instrument was undertaken. There was a consistent pattern across all three instruments of a tendency to score moderate flares as severe and scoring some of the no flare/persistent activity cases as flare when the SFI and rSFI instruments were used. Close examination of the data in tables 2a and 3a suggested that this over-scoring was driven by scoring treatment change as flare even if there were no new or worsening clinical features. Tables 2b and 3b show the adjusted results. Only one out of the 17 cases assessed by the SELENA-SLEDAI flare instrument as a flare but as no flare by TPC had increased lupus activity without treatment change and in 2 cases the treatment change led to a higher level of flare than the clinical features alone would have scored. With the rSFI instrument only 1 out of 5 cases recorded as severe SELENA flare but no flare by TPC had clinical features of flare, the others were due to treatment change alone. In one case there was mild flare by TPC but severe SELENA flare due to treatment change._ Excluding flare defined primarily by treatment change when assessing flares (tables 2b and 3 b) using the SFI and rSFI led to an increase in the proportion of cases with agreement between the TPC and the instruments to 79% and 78% respectively and improved the weighted kappa scores to 0.82 and 0.73 respectively.*

For the cases assessed by the BILAG 2004 index in Table 1, there were 2 cases with severe BILAG flare that were appropriately scored and it was not clear why the TPC assessment was of no flare. For the 2 cases with severe flare by TPC assessed as mild by the BILAG assessors, review suggested that they had not been scored correctly and that the criteria for severe flare (BILAG A scores) were present in both cases. There were also 2 out of 4 cases recorded by TPC as severe flare with a moderate flare reported by the BILAG assessors but the criteria for severe flare were present. In 3 out of 4 of these cases scored at a lower level by the BILAG assessors, neurological items were underscored as B instead of A for reasons that are not clear. The patients with severe flares recorded by BILAG, but moderate flares by TPC, were correctly scored for BILAG in 9 out of 11 cases and were mostly mucocutaneous or musculoskeletal flares (8 out of 11 cases , with 2 cardio-respiratory flares and one renal flare).

Discussion:

This paper-patient exercise has allowed the assessment of a much wider range of clinical problems compared to the previous live patient exercise (9). By utilizing over 40 rheumatologists with a major interest in SLE we were able to collect a large number of cases for flare assessment and involved a much larger number of assessors. The disposition of case review assignments was arranged to ensure that a similar range of cases was reviewed by all participating physicians. In the first part of the study, the 988 case histories were reviewed independently by three rheumatologists and full agreement as to whether the patient was suffering a mild, moderate or severe flare or experiencing persistent/ongoing disease was achieved in 451 cases (46%). Review of discrepant cases suggested that there were some errors in the scoring of the BILAG 2004 index without which the levels of agreement would have been higher, emphasising the importance of training in the use of the glossary and the scoring manual. Although there was some evidence that single system severe A level flares by BILAG 2004 index in mucocutaneous and musculoskeletal systems were sometimes perceived to be moderate level flares by the TPC, the case descriptions varied in the amount of detail provided and it is difficult to attach much significance to the small number of cases that this applied to given that the presence of a significant flare was clear to all assessors.

The case histories with full agreement were then utilized, in effect, as our 'gold standard' cases. The three lupus flare instruments ie BILAG-2004, SFI index and rSFI were used to assess flare in different cases in the second part of the study and performed equally well, particularly if treatment change was excluded as a component of the flare score. In practice, the scores from the pairs of rheumatologists that completed the BILAG-2004 or SFI index or the rSFI index demonstrated a high level of agreement overall with the pre-determined TPC level of flare, and for the specific types of flare, especially for severe flare (levels of agreement of 85% or higher). Distinguishing mild and moderate flares from persistent activity/no flare and severe flare as assessed by the TPC method was more consistently achieved when the SFI and rSFI instruments were used without the rules that required all treatment change to determine flare. This issue has previously been reported in a smaller retrospective real patient cohort study [14]. It is apparent that treatment change by itself will quite frequently fail to indicate worsening disease as adverse events, or treatment intolerance, may complicate matters. As a 'treat to target' philosophy gains traction in the management of SLE this increases the likelihood that treatment change will be widely used to define flare when it may simply reflect 'fine tuning' of managing persistent disease. Nevertheless an 'intent to treat principle' remains useful to reflect broadly clinically important levels of disease activity.

It should be noted that a few cases (seven) were not scored but were included in the analysis despite insufficient information, predominately (6/7 cases) those for BILAG flare assessment as this is more demanding in terms of the information required for accurate and comprehensive completion of case report forms. Where there was initial discord in flare scoring for the other cases reported, the pairs reported that by discussion (usually via telephone conference) agreement could be reached relatively easily in the vast majority of cases even though the quality of the case scenarios was rather variable. The pairs of assessors only used one instrument and each of the three of types of flare instrument were not used on the same cases by the same people.

In 535 (54%) of the 988 cases originally submitted for flare assessment, consensus on the level of flare without using a flare instrument could not be agreed by three physicians. Although in some cases there was insufficient information for some of the physicians to rate the case, for example because the severity was not clear enough or the time of onset of the deterioration was not adequately defined (to distinguish flare from persistent activity or damage), there were other cases where consensus agreement could not be achieved despite reasonable scenarios . These were mostly those cases with multiple systems involved, presumably because different physicians ranked the components differently, particularly if some systems changed severity to different extents or not consistently with some symptoms worse and others stable or improved. This is reminiscent of a problem noted previously in a study evaluating responsiveness using BILAG and SLEDAI compared with a physician visual analogue scale [15] and emphasises the advantage in obtaining consistency of scoring in flare instruments with defined glossaries rather than relying solely on physician opinion. However, limitations of disease definitions that rely on predefined thresholds of severity cannot be excluded.

Although the capacity of paper patient exercises greatly increases the range of possible combinations of lupus clinical features for assessment in studies of this kind, nothing captures the dilemma of lupus assessment as much as a real patient in front of a clinician. Nevertheless the practicalities of getting large numbers of patients who are actually flaring into a clinical assessment study of the type that we reported previously (9) are considerable. In retrospect we would probably have obtained greater agreement amongst the raters/assessors and a larger panel of cases for flare assessment with much tighter requirements/standardization for the writing of the case histories. An advantage of the case histories is that they were devised from real clinic patients. We are aware that there may have been biases in that some of the assessors may have been more or less experienced in the assessment instruments used.

It is notable that it is not just in SLE that challenges are evident in attempting to capture flare adequately, and distinguish it from ongoing disease. Thus the OMERACT RA Flare Group have been involved in similar studies for several years (16). Their work is ongoing and the fact that they are dealing with a single organ/system highlights the added complexity when dealing with lupus.

Although it would not be true to say that the problem of capturing flare accurately in patients with lupus is ended, the relatively high levels of agreement obtained with the flare instruments and weighted kappas ranging from 0.59 (SFI instrument with treatment) to 0.82 (for SELENA without treatment and BILAG-2004) was encouraging especially given some problems with inadequate histories and the great diversity of rheumatologists from many different countries involved in the study. All of the instruments used in this study have construct and content validity based on their use with these paper patient scenarios. The choice of instrument to be used in future flare studies will depend on the types of patients to be assessed, the need to distinguish different types of flares and the training and experience of the likely investigators.

Acknowledgements:

The SLICC group gratefully acknowledges the support of many colleagues who helped us to perform the study – in particular:-

Dr M Aringer [Dresden]

Dr S Chambers [Cayman Islands]

Dr Costedoat-Chalumeau [Paris]

Dr S Croca [Lisbon]

Dr I Giles [London]

Dr B Griffiths [Newcastle]

Dr P Lanyon [Nottingham]

Dr E Moreland [Melbourne]

Dr M Mosca [Pisa]

Dr JM Pego-Reigosa [Vigo]

Dr M Schneider [Dusseldorf]

Dr C Vasconcelos [Porto]

References

1) Gladman DD, Ibanez D, Urowitz MB. Systemic lupus erythematosus activity index 2000. J Rheumatol 2000; 29; 288-91.

2) Yee Cs, Farewell V, Isenberg DA et al. The BILAG 2004 index is sensitive to change for assessment of SLE disease activity. Rheum 2009; 48; 691-5.

3) Nuttall A, Isenberg DA. Assessment of disease activity damage and quality of life in systemic lupus erythematosus. Best Practice and Res Clinical Rheumatology 2013; 27; 300–16.

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

5) 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.

6) Isenberg D, Gordon C, Licu D et al. Efficacy and safety of atacicept for prevention of flares in patients with moderate-to-severe systemic lupus erythematosus (SLE): 52-week data (APRIL-SLE randomised trial). Ann Rheum Dis 2015;74:2006-2015.

7) Yee CS, Farewell V, Isenberg DA et al. The use of the Systemic Lupus Erythematosus Disease Activity Index-2000 to define active disease and minimal clinically meaningful change based on data from a large cohort of systemic lupus erythematosus patients. Rheumatology 2011; 50: 982-8.

8) Ruperto N, Hanrahan LM, Alarcon GS, et al. International consensus for a definition of disease flare in lupus. *Lupus* 2011; **20**:453-462.

9) Isenberg DA, Allen E, Farewell V et al. An assessment of disease flare in patients with systemic lupus erythematosus: a comparison of BILAG-2004 and the flare version of SELENA. Am Rheum Dis 2011; 70; 54-9

10) Hochberg MC. Updating the American College of Rheumatology Criteria For the Classification of Systemic Lupus Erythematosus. Arthritis Rheum 1997; 40; 1725

11) Petri M, Orbai AM, Alarcon GS, Gordon C, Merrill J, Fortin P et al. Derivation and validation of the Systemic Lupus Erythematosus International Collaborating Clinics classification criteria for Systemic Lupus Erythematosus. Arthritis Rheum 2012; 64; 2677-80.

12) Yee CS, Cresswell L, Farewell V. Numerical scoring for the BILAG-2004 index *Rheumatology* 2010; 49**:** 1665-9.

13) Buyon JP, Petri MA, Kim MY et al. The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: a randomized trial. Ann Intern Med 2005; 142: 953-62.

14) Thanou A, Chakravarty E, James JA et al. How should lupus flares be measured? Deconstruction of the safety of estrogen in lupus erythematosus national assessment systemic lupus erythematosus disease activity index flare index. Rheumatology 2014; 53: 2175-81.

15) Wollaston SJ, Farewell JT, Isenberg DA et al. Defining response in systemic lupus erythematosus: a study by the Systemic Lupus Erythematosus International Collaborating Clinic Group. J Rheumatol 2004; 3: 2390-4.

16) Bartlett SJ, Bykerk VP, Cooksey R et al. Feasibility and domain validation of RA flare core domain set: A report of the Omeract 2014 RA flare group plenary. J Rheum 2015: 42; 2185-9.

ACKNOWLEDGEMENTS:

The authors gratefully acknowledge the support of a grant from a combined ACR/EULAR committee ($50,000) to facilitate the undertaking of this study.

Table 1: Number of cases by BILAG 2004 flare assessment and % agreement with TPC flare assessment

		TPC flare assessment				
		No flare	Mild flare	Moderate flare	Severe flare	Total
BILAG flare assessment n (% agreement)	No flare	**27 (75)**	11 (31)	2 (7)	0	40 (27)
	Mild flare	4 (11)	**23 (64)**	6 (21)	2 (4)	35 (24)
	Moderate flare	1 (3)	0	**9 (31)**	4 (8)	14 (9)
	Severe flare	2 (6)	0	11 (38)	**41 (85)**	54 (36)
	Insufficient information	2 (6)	2 (6)	1 (3)	1 (2)	6 (4)
Total		36	36	29	48	149

TPC : three physician consensus

Table 2a: Number of cases by SELENA – SLEDAI flare instrument assessment and % agreement with TPC flare assessment

		TPC flare assessment			
		No flare	Mild/moderate flare (combined)	Severe flare	Total
SELENA flare assessment n (% agreement)	No flare	**17 (49)**	5 (7)	0	22 (15)
	Mild/moderate flare	8 (23)	**49 (67)**	0	57 (38)
	Severe flare	9 (26)	19 (26)	**44 (100)**	72 (47)
	Insufficient information	1 (3)	0	0	1
Total		35	73	44	152

TPC : three physician consensus

Table 2b: Number of cases by SELENA – SLEDAI flare assessment and % agreement with TPC flare assessment – excluding flare defined by treatment change

		TPC flare assessment			
		No flare	Mild/moderate flare (combined)	Severe flare	Total
SELENAFlare Index without treatment flare assessment n (% agreement)	No flare	**22 (63)**	10 (14)	0	32 (21)
	Mild/moderate flare	8 (23)	**53 (73)**	0	61 (40)
	Severe flare	4 (11)	10 (14)	**44 (100)**	58 (38)
	Insufficient information	1 (3)	0	0	1
Total		35	73	44	152

TPC : three physician consensus

Table 3a: Number of cases by SELENA flare assessment and % agreement with TPC flare assessment

		TPC flare assessment				Total
		No flare	Mild flare	Moderate flare	Severe flare	
Revised SELENA flare assessment n (% agreement)	No flare	**18 (58)**	0	0	0	18 (12)
	Mild flare	4 (13)	**16 (44)**	0	0	20 (13)
	Moderate flare	4 (13)	19 (53)	**19 (63)**	1 (2)	43 (29)
	Severe flare	5 (16)	1 (3)	11 (37)	**52 (98)**	69 (46)
	Insufficient information	0	0	0	0	0
Total		31	36	30	53	150

TPC : three physician consensus

Table 3b: Number of cases by SELENA flare assessment and % agreement with TPC flare assessment-excluding flare defined by treatment change

		TPC flare assessment				Total
		No flare	Mild flare	Moderate flare	Severe flare	
Revised SELENA flare assessment n (% agreement)	No flare	**23 (74)**	0	0	2 (4)	25 (17)
	Mild flare	4 (13)	**25 (69)**	5 (16.7)	1 2)	35 (23)
	Moderate flare	3 (10)	11 (31)	**23 (77)**	3 (6)	40 (27)
	Severe flare	1 (3)	0 (0)	2 (7)	**47 (89)**	50 (33)
	Insufficient information	0	0	0	0	0
Total		31	36	30	53	150

TPC : three physician consensus