






SLE-DAS enables an accurate definition of severe lupus disease activity: derivation and validation in a post hoc study of anifrolumab phase II and III studies

Diogo Jesus ^{1,2}, Ana Matos ^{3,4}, Carla Henriques ^{3,5}, Andrea Doria ⁶, Luis Sousa Ines ^{2,7}

To cite: Jesus D, Matos A, Henriques C, *et al.* SLE-DAS enables an accurate definition of severe lupus disease activity: derivation and validation in a post hoc study of anifrolumab phase II and III studies. *Lupus Science & Medicine* 2025;**12**:e001499. doi:10.1136/lupus-2025-001499

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/lupus-2025-001499>).

The results were presented, in part, at the 14th European Lupus Meeting (2024) and published as a conference abstract (Jesus D, *et al.* *Lupus Sci Med* 2024;**11**[Suppl 1]:A90.2).

Received 3 January 2025
Accepted 25 April 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to

Professor Luis Sousa Ines;
luisines@gmail.com

ABSTRACT

Objectives This study aimed to derive and validate a cut-off for severe disease activity (SDA) using the SLE Disease Activity Score (SLE-DAS) and compare its accuracy and impact on health-related quality of life (HR-QoL) with the British Isles Lupus Assessment Group 2004 (BILAG-2004) and SLE Disease Activity Index 2000 (SLEDAI-2K).

Methods We performed a post hoc analysis of pooled placebo arm data from the MUSE (A Phase II, Randomized Study to Evaluate the Efficacy and Safety of MEDI-546 in Subjects with Systemic Lupus Erythematosus), TULIP-1 and TULIP-2 (Treatment of Uncontrolled Lupus via the Interferon Pathway) trials, including 438 patients with moderate-to-severe SLE. SLE-DAS was scored retrospectively, and a cut-off for SDA was derived using receiver operating characteristic (ROC) curves against the BILAG-2004 numerical score >11 as gold standard. Multiple linear regression analysis and Cohen's d effect size were applied to evaluate the effectiveness of SLE-DAS, BILAG-2004 and SLEDAI-2K SDA classifications in capturing HR-QoL patient-reported outcomes (PROs).

Results The optimal SLE-DAS cut-off for SDA was >9.90 (area under the ROC curve=0.847, sensitivity=77.8%, specificity=79.6%). Patients classified as SDA by both SLE-DAS and BILAG-2004 or only by SLE-DAS exhibited similar disease activity, while those classified by BILAG-2004 alone had less severe disease and better HR-QoL. The SLE-DAS cut-off was associated with worse HR-QoL across multiple PROs more consistently than BILAG-2004 or SLEDAI-2K.

Conclusion The SLE-DAS cut-off for SDA provides an accurate definition of SDA in SLE, with good discriminative power and consistent associations with worse HR-QoL. This SLE-DAS definition enhances disease activity classification and offers a practical tool for guiding treatment decisions in clinical practice, as well as selecting patients with SDA for inclusion in clinical trials.

INTRODUCTION

Accurate assessment of SLE disease activity with a validated instrument is recommended for guiding treatment decisions, monitoring

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The SLE Disease Activity Score (SLE-DAS) is a validated tool for assessing global disease activity in SLE, supporting treat-to-target strategies with validated definitions for remission and low, mild and moderate-to-severe disease activity.

WHAT THIS STUDY ADDS

⇒ This study derives and validates the SLE-DAS cut-off (>9.90) for severe disease activity (SDA), demonstrating good discriminatory ability and a superior association with worse health-related quality of life compared with British Isles Lupus Assessment Group 2004 and SLE Disease Activity Index 2000.
⇒ It highlights discrepancies in SDA classification between SLE-DAS and other tools, emphasising SLE-DAS's superior association with clinical and patient-reported outcomes.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This SLE-DAS definition enhances disease activity classification and provides a practical tool for guiding treatment decisions in clinical practice, as well as selecting patients with SDA for inclusion in clinical trials.
⇒ Future research could adopt SLE-DAS as a continuous measure for efficacy endpoints in trials, refining therapeutic strategies and improving patient outcomes.

treatment response and predicting patient outcomes.^{1–3} According to the 2023 European Alliance of Associations for Rheumatology (EULAR) recommendations for the management of SLE, treatment should be tailored according to the intensity of SLE disease activity, categorised as mild, moderate and severe.³

However, there are no consensual definitions for categories of SLE disease activity.

This is a significant obstacle to implementing evidence-based, standard-of-care treatment recommendations in the clinical setting, therefore compromising the patient outcomes. **Discrimination of mild, moderate and severe SLE disease activity based on the SLE Disease Activity Index (SLEDAI) or the British Isles Lupus Assessment Group (BILAG) index has been proposed, but these instruments have limitations and were not validated for this purpose.**¹ The SLEDAI has limited measurement performance to discriminate disease activity because it dichotomously scores (as present/absent) each manifestation, thus failing to capture the extent of disease activity within each organ manifestation.^{2,4,5} In addition, it assigns inappropriate weights for some activity items (eg, severe thrombocytopenia scores lower than oral ulcers) and does not include several severe lupus features, such as haemolytic anaemia, and gastrointestinal and cardiopulmonary manifestations.^{6,7} The BILAG index, while offering a more comprehensive assessment across multiple organ systems and providing a graded severity for each manifestation, also has its drawbacks. It relies on subjective definitions of severity, which can lead to inconsistent scoring, and its transitional nature may cause disease activity to be categorised differently based on the physician's perception of changes over time.¹ Additionally, BILAG's severity scoring does not account for differences in the relative seriousness of organ involvement. For instance, an 'A' score for severe mucosal ulceration is graded equivalently to life-threatening complications like lupus nephritis, pneumonitis or enteritis.^{8,9} Hence, there is an important unmet need for more accurate and validated definitions of SLE disease activity categories for use in clinical trials and in daily practice.

The SLE Disease Activity Score (SLE-DAS) may provide a solution for accurate and practical discrimination of disease activity categories. The SLE-DAS is a validated, composite instrument for measuring global disease activity, which was developed to provide an accurate, sensitive-to-change and feasible measure in the clinical setting. It comprises 17 weighted clinical and laboratory parameters, including continuous measures for arthritis, proteinuria, thrombocytopenia and leucopenia.^{5, 10–12} The SLE-DAS provides a balanced weighting system of disease activity features and includes haemolytic anaemia, and cardiopulmonary and gastrointestinal involvement, making it a comprehensive tool. The SLE-DAS provides validated definitions for the remission and low disease activity treatment targets, which demonstrated high classification performance.^{13–16} Additionally, the SLE-DAS definitions of mild activity ($2.08 < \text{SLE-DAS} \leq 7.64$) and moderate-to-severe activity ($\text{SLE-DAS} > 7.64$) presented in validation a high performance, as compared with either the expert assessment or the BILAG.¹³ These SLE-DAS categories are accurate and feasible in the clinical setting, taking only 1–2 min to assess with its free-to-use online calculator (<http://sle-das.eu>).¹³

The aim of this study is to derive and validate the SLE-DAS definition for severe disease activity (SDA).

PATIENTS AND METHODS

Study population

We performed a post hoc analysis of the intention-to-treat pooled data from the placebo arms of the MUSE (A Phase II, Randomized Study to Evaluate the Efficacy and Safety of MEDI-546 in Subjects with Systemic Lupus Erythematosus) (NCT01438489),¹⁷ TULIP-1 (Treatment of Uncontrolled Lupus via the Interferon Pathway) (NCT02446912)¹⁸ and TULIP-2 (NCT02446899)¹⁹ placebo-controlled, randomised, double-blind, 52-week trials that assessed the efficacy and safety of anifrolumab in patients with moderate-to-severe disease activity who were receiving standard therapy. The study design and methods have been published in detail.^{17–19} Briefly, eligible patients were required to meet the American College of Rheumatology 1997 classification criteria for SLE²⁰ and exhibit active disease at baseline defined cumulatively as a SLE Disease Activity Index 2000 (SLEDAI-2K) score ≥ 6 , a clinical SLEDAI-2K score ≥ 4 , a BILAG-2004 organ domain score ≥ 1 A item or ≥ 2 B items and a Physician Global Assessment (PGA) score ≥ 1 , despite receiving at least one standard-of-care medication. Patients with severe active renal or central nervous system disease were excluded.^{16–18} The MUSE and TULIP trials were conducted according to the principles of the Declaration of Helsinki. The protocols were approved by institutional review boards, and all participants provided written informed consent. For this post hoc analysis, access to the individual patient data from the placebo arms of these randomised clinical trials (RCTs) was granted by Astra-Zeneca through the Vivli consortium.

Patient assessments

We analysed the prospectively assessed individual patient data from SLEDAI-2K, BILAG-2004, Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI), 28-swollen joint count (28-SJC), analytical results and the health-related quality of life (HR-QoL) patient-reported outcomes (PROs) ((Lupus Quality of Life (LupusQoL), EuroQoL 5 Dimension (EQ-5D), Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-F) and Patient Global Assessment (PtGA)) from the dataset. All data necessary to retrospectively score the SLE-DAS were available using the SLEDAI-2K, BILAG-2004, CLASI individual items, 28-SJC and analytical results from the dataset.

Statistical analysis

Derivation and validation of the SLE-DAS cut-off for SDA

Derivation and validation of the SLE-DAS cut-off for SDA were performed using data from week 12. At this visit, the proportion of patients with SDA versus non-SDA was expected to be balanced, based on the timepoint results from the MUSE and TULIP trials. To establish a gold standard for SDA, we used the numerical scoring scheme developed and validated by the BILAG group as comparator.²¹ Specifically, we employed a numerical threshold of BILAG-2004 > 11 based on prior applications

in disease stratification.²¹ Although the specific threshold of BILAG-2004 >11 for defining SDA has not been extensively validated, several studies confirmed the numerical transformation of BILAG-2004, showing that it is sensitive to change in disease activity and associated with treatment decisions, with other disease activity indices, flares and damage accrual.^{21–24} Derivation was performed using data from the MUSE and TULIP-2 trials and applying receiver operating characteristic (ROC) curves against the BILAG-2004 numerical definition of SDA (numerical BILAG-2004 >11). The area under the ROC curve (AUC) was considered as a measure of discriminatory ability. The optimal cut-off was determined by applying the Closest to (0,1) Criteria (ER), where the optimal threshold is the point closest to the point (0,1) on the ROC curve. Bootstrap CIs for the cut-off were calculated using 2000 bootstrap replicates.

Validation was assessed using data from the TULIP-1 trial. Sensitivity, specificity and geometric mean (G-mean) for the cut-off value of SLE-DAS SDA against BILAG-2004 were calculated. We performed additional analyses, assessing the discordant classification of SDA between SLE-DAS and BILAG-2004. To this purpose, we compared the physician-assessed disease activity features and the HR-QoL PROs of the group of patients fulfilling the SDA definition by both SLE-DAS and BILAG-2004 with the group classified as SDA only according to SLE-DAS (but not BILAG-2004) and the group as SDA only by BILAG-2004 (but not SLE-DAS), applying Fisher's exact test, χ^2 test or Mann-Whitney U test, using for these sensitivity analyses the aggregated data of MUSE and both TULIP trials.

Comparison of HR-QoL between patients classified as SDA and non-SDA by SLE-DAS, BILAG-2004 or SLEDAI-2K

We further analysed the classification of all patients at week 12 with SDA using the derived cut-off of SLE-DAS, the prespecified numerical BILAG-2004 >11 and SLEDAI-2K >12. The selection of SLEDAI-2K >12 was based on its adoption in the EULAR recommendations for the treatment of SLE and its use in the Safety of Estrogens in Lupus Erythematosus National Assessment (SELENA)-SLEDAI Flare Index to define severe flares.^{3 25 26} However, it is important to acknowledge that there is no SLEDAI cut-off extensively validated as an independent definition of SDA. The agreement between these different classifications was assessed using Fleiss' kappa, a generalisation of Cohen's kappa for three or more raters. Additionally, we assessed if patients classified as SDA as compared with non-SDA by each of the SLE-DAS, SLEDAI-2K or BILAG-2004 criteria had a worse HR-QoL using the Mann-Whitney U test. To quantify the magnitude of the differences, we calculated Cohen's d, which is a standardised effect size used to measure the difference between two means, allowing for the comparison of effect sizes across different variables.²⁷ Multiple linear regression analysis with backward elimination method was applied to evaluate the effectiveness of SLE-DAS, BILAG-2004

and SLEDAI-2K SDA classifications in capturing HR-QoL PROs. Only the SDA definitions retained in the final model were considered relevant for predicting HR-QoL PROs. This methodology ensured that the most pertinent predictors were identified, providing insights into which disease activity classifications best capture the patients' perspective in HR-QoL.

Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting or dissemination of this research.

RESULTS

Patient characteristics

We assessed a total of 438 patients with SLE (95, 174 and 169 from the MUSE, TULIP-1 and TULIP-2 RCTs, respectively). These were all the placebo arm participants made available by AstraZeneca through Vivli, and no additional selection criteria were applied. Baseline characteristics of the patients included are presented in online supplemental table 1 and in more detail in the original publications of MUSE, TULIP-1 and TULIP-2 trials.

Derivation and validation of the SLE-DAS cut-off for SDA

At week 12, 46.6% and 42.4% of 438 patients with SLE were classified as SDA by BILAG-2004 in the derivation and validation cohorts, respectively. In the derivation cohort, the best cut-off to identify patients with SDA was SLE-DAS >9.90 (bootstrap 95% CI 8.59 to 10.24), yielding an AUC=0.847 (95% CI 0.811 to 0.882) (online supplemental figure 1). When applied in the validation cohort, this cut-off showed a sensitivity of 77.8% and a specificity of 79.6% (online supplemental table 2). Online supplemental tables 3 and 4 present a cross-tabulation of SLE-DAS and BILAG-2004 disease activity categories, as well as the performance of the >9.90 SLE-DAS cut-off in patients with moderate-to-severe disease activity according to BILAG-2004. When this cut-off was reassessed within this subgroup, sensitivity remained unchanged in both the derivation and validation cohorts (76.8% and 77.8%, respectively), while the specificity decreased slightly, resulting in a modest reduction in G-mean from 78.7 to 73.5 in the validation cohort.

At week 12, from the 438 patients with SLE, 34% were classified as SDA by both SLE-DAS and BILAG-2004, 12.6% only by SLE-DAS and 10% only by BILAG-2004. Among the patients classified as SDA based on SLE-DAS, the most prevalent manifestations were arthritis (85.3%) and cutaneous involvement, including cutaneous rash (82.4%) and alopecia (76%) (online supplemental table 5). Mucosal ulcers were also frequently observed, affecting 33.8% of patients. Renal involvement was identified in 15.2% of patients, while serositis was present in 11.8%. Haematological abnormalities included leucopenia (8.8%) and thrombocytopenia (2%). Less frequent manifestations included neuropsychiatric symptoms (2.5%), systemic vasculitis (0.6%), cardiopulmonary involvement (0.5%) and myositis (3.9%). The

Table 1 Comparison of disease features of patients fulfilling the severe disease activity (SDA) definition both by SLE-DAS and BILAG-2004 with those in SDA only by SLE-DAS (but not BILAG-2004) or only by BILAG-2004 (but not SLE-DAS)

	SDA by SLE-DAS and BILAG-2004 n=149 (34.0%)	SDA only by SLE-DAS n=55 (12.6%)	SDA only by BILAG-2004 n=44 (10%)
Neuropsychiatric, n (%)	4 (2.68)	1 (1.82)	0 (0.00)
P value*		NA	NA
Systemic vasculitis, n (%)	1 (0.67)	0 (0.00)	0 (0.00)
P value*		NA	NA
Mucocutaneous vasculitis, n (%)	30 (20.13)	8 (14.55)	5 (11.36)
P value*		0.363	0.185
Cardiopulmonary, n (%)	1 (0.67)	0 (0.00)	0 (0.00)
P value*		NA	NA
Serositis, n (%)	19 (12.75)	5 (9.09)	1 (2.27)
P value*		0.471	0.049
Lupus nephritis, n (%)	24 (16.11)	7 (12.73)	0 (0.00)
P value*		0.551	0.002
Arthritis, n (%)	129 (86.58)	45 (81.82)	25 (56.82)
P value*		0.394	<0.0001
28-SJC, median, IQR	4 (2; 8)	3 (1; 7)	1 (0; 2)
P value*		0.2668	<0.0001
Myositis, n (%)	6 (4.03)	2 (3.64)	0 (0.00)
P value*		1	0.34
Mucocutaneous involvement, n (%)	142 (95.30)	51 (89.09)	41 (93.18)
P value*		0.117	0.698
Thrombocytopenia, n (%)	3 (2.01)	1 (1.82)	1 (2.27)
P value*		1	1
Leucopenia, n (%)	13 (8.72)	5 (9.09)	4 (9.09)
P value*		1	1
Hypocomplementaemia, n (%)	72 (48.32)	28 (50.91)	19 (43.18)
P value*		0.743	0.548
Increased anti-dsDNA, n (%)	77 (51.68)	31 (56.36)	13 (29.55)
P value*		0.552	0.01
Severe CLASI-A, n (%)	9 (6.04)	0 (0.00)	2 (4.55)
P value*		0.117	1
SLEDAI-2K, median (IQR)	12 (8; 14)	10 (8; 12)	8 (6; 9)
P value*		0.0996	<0.0001

*Comparison between patients classified as SDA by SLE-DAS+BILAG-2004 versus patients classified as SDA only by SLE-DAS or BILAG-2004 using χ^2 test, Fisher's exact test or Mann-Whitney U test as appropriate. Significant p values are presented in bold. SDA by BILAG-2004, British Isles Lupus Assessment Group 2004 index numerical score >11; SDA by SLE-DAS, SLE Disease Activity Score >9.90; Severe CLASI-A, Cutaneous Lupus Erythematosus Disease Area and Severity Index activity score ≥ 21 ; 28-SJC, number of swollen joints in a 28-joint count; SLEDAI-2K, SLE Disease Activity Index 2000.

groups classified as SDA according to both SLE-DAS and BILAG-2004 and those classified only by SLE-DAS did not present discernible differences in active disease features and HR-QoL PROs (table 1 and figures 1 and 2). In contrast, patients classified as SDA only by BILAG-2004 presented lower SLEDAI-2K score, less cases with active serositis, nephritis, arthritis and with anti-dsDNA

positivity as compared with patients classified as SDA by both SLE-DAS and BILAG-2004. Notably, patients classified as SDA by BILAG-2004 alone presented significantly less severe impact in HR-QoL PROs, including in five of eight domain scores of the LupusQoL, FACIT-F, PtGA, EQ-5D visual analogue scale (VAS) and EQ-5D index (figure 2).

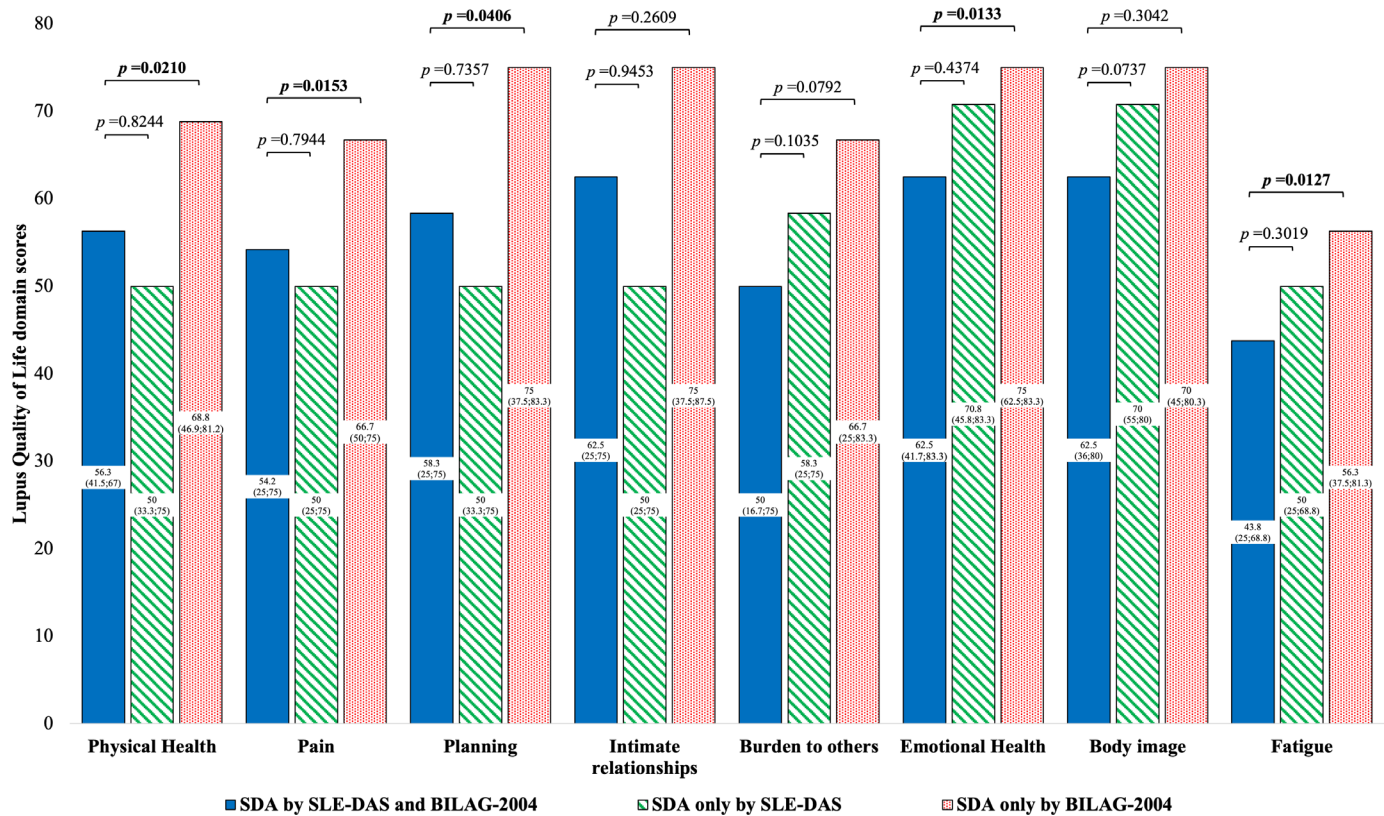


Figure 1 Comparison of Lupus Quality of Life domain scores of patients fulfilling the severe disease activity (SDA) definition by SLE Disease Activity Score (SLE-DAS) and British Isles Lupus Assessment Group 2004 (BILAG-2004) with those classified as SDA only by SLE-DAS (but not BILAG-2004) or only by BILAG-2004 (but not SLE-DAS) at week 12. The height of the bars represents the mean values, and the number in each bar of the graph represents the median and the IQRs are listed below. P values were calculated using the Mann-Whitney U test.

HR-QoL comparison between patients classified as SDA and non-SDA by SLE-DAS, BILAG-2004 or SLEDAI-2K

At week 12, 11.6% of the 438 patients with SLE were classified as SDA by all three definitions (SLEDAI-2K, SLE-DAS and BILAG-2004), 22.4% only by SLE-DAS and BILAG-2004, 3.0% only by SLE-DAS and SLEDAI-2K, 0.2% only by BILAG-2004 and SLEDAI-2K, 9.6% only by SLE-DAS, 9.8% only by BILAG-2004 and 0.7% only by SLEDAI-2K. The agreement between SLE-DAS, BILAG-2004 and SLEDAI-2K was fair (Fleiss' kappa=0.33, 95% CI 0.28 to 0.39, $p < 0.0001$).²⁸

Patients classified as SDA as compared with patients classified as non-SDA by SLE-DAS and BILAG-2004 presented significantly worse impact in HR-QoL in all domains of LupusQoL, FACIT-F, PtGA and the EQ-5D (figure 3 and online supplemental figure 2). In contrast, patients classified as SDA or non-SDA by SLEDAI-2K did not present significantly different impact in five of eight domains of LupusQoL, FACIT-F, PtGA and EQ-5D (online supplemental figure 3). Notably, the SLE-DAS SDA presented numerically higher effect sizes for all HR-QoL PROs compared with BILAG-2004 and SLEDAI-2K (figure 3 and online supplemental figures 2 and 3).

Multiple linear regression was applied to predict HR-QoL. The results from the full model including the SLE-DAS, BILAG-2004 and SLEDAI-2K SDA definitions

are presented in online supplemental table 6 and the results from the backward selection model in table 2. The SLE-DAS SDA was the only SDA definition retained in the final model from backward selection for PtGA, FACIT-F, five of the eight domains of LupusQoL and both the EQ-5D index and VAS (table 2). In contrast, the SDA definition of BILAG-2004 was the only SDA definition retained in the final model for the body image domain of LupusQoL. Meanwhile, the SLEDAI-2K SDA definition showed a positive partial correlation with the pain domain of LupusQoL and the EQ-5D index, when the SDA category according to SLE-DAS is held constant. Interestingly, this suggests that within the same SLE-DAS category (SDA or non-SDA), being classified as SDA by SLEDAI-2K is linked to lower reported pain levels and better overall quality of life by the EQ-5D index. Overall, the SLE-DAS SDA definition was associated with worse HR-QoL across multiple PROs more consistently than BILAG-2004 or SLEDAI-2K SDA.

DISCUSSION

In this study, we derived and validated the SLE-DAS cut-off value for SDA and evaluated its effectiveness in capturing HR-QoL in patients with SLE. Our findings demonstrate that the SLE-DAS is a trustworthy and accurate tool for

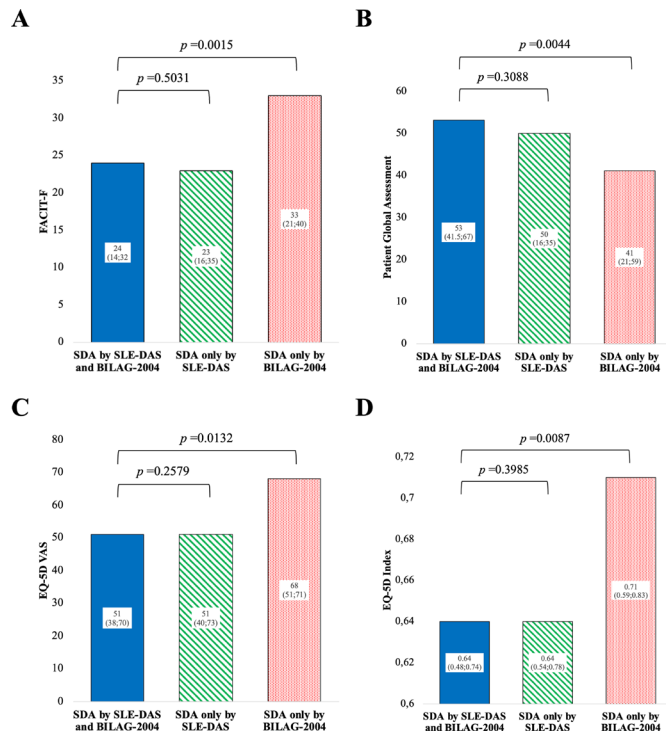


Figure 2 Health-related quality of life patient-reported outcomes among patients fulfilling the severe disease activity (SDA) definition by SLE Disease Activity Score (SLE-DAS) and British Isles Lupus Assessment Group 2004 (BILAG-2004) and those in SDA only by SLE-DAS (but not BILAG-2004) or only by BILAG-2004 (but not SLE-DAS) at week 12. (A) Comparison of Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-F). (B) Comparison of Patient Global Assessment. (C) Comparison of EuroQol 5 Dimension (EQ-5D) visual analogue scale (VAS). (D) Comparison of the EQ-5D index. The height of the bars represents the mean values, and the numbers inside the bars represent the median and the IQRs. P values were calculated using the Mann-Whitney U test.

identifying patients with SDA exhibiting good discriminatory ability (AUC=0.837). Additionally, it is associated with worse HR-QoL across multiple PROs, presenting high predictive value for worse HR-QoL compared with BILAG-2004 and SLEDAI-2K SDA.

The SLE-DAS SDA cut-off value of >9.90 exhibited good discriminatory ability, with sensitivity and specificity around 80% in the validation cohort, offering a robust alternative to the existing measures such as SLEDAI-2K and BILAG-2004. To further explore the performance within the moderate-to-severe range, we conducted an additional analysis restricted to patients with BILAG-defined moderate-to-severe activity. Sensitivity remained stable, while specificity decreased slightly, as expected due to the reduced contrast in this subset. These findings support the robustness of the >9.90 SLE-DAS cut-off even in a more challenging subset. Importantly, when we evaluated disease features and HR-QoL PROs in cases of discordant classification between SLE-DAS and BILAG-2004, patients classified as SDA by either SLE-DAS and BILAG-2004 or only by SLE-DAS showed similarly severe disease.

In contrast, those classified as SDA only by BILAG-2004 presented less severe disease, as assessed by both physician and patient evaluations. Moreover, the SLE-DAS SDA was consistently associated with worse HR-QoL across all PROs, including PtGA, FACIT-F, LupusQoL and EQ-5D, highlighting its utility in capturing the broader impact of SDA on patients' lives.

The previously validated SLE-DAS cut-off of >7.64 for identifying moderate-to-severe disease activity was originally validated against the BILAG.¹³ In the present study, we further refined disease activity classification by identifying a cut-off of >9.90 to define SDA, using the same gold standard. These results support the interpretation of the SLE-DAS range greater than 7.64 and lower than or equal to 9.90 as representing moderate disease activity and reinforce the clinical applicability of the SLE-DAS for disease activity stratification.

The multiple regression analysis showed that the SDA of BILAG-2004 was the most pertinent predictor of HR-QoL only in the body image domain of LupusQoL. This suggests that although BILAG-2004 SDA is associated with HR-QoL, the SLE-DAS SDA reflects better the overall burden of severe disease on HR-QoL. Furthermore, the non-significant differences between SDA and non-SDA according to SLEDAI-2K, with respect to overall quality of life, call into question the validity of SLEDAI-2K as an accurate measure of SDA and its impact on patients' quality of life.

It is important to note the limitation of the BILAG-2004 in categorising global disease activity. Unlike SLEDAI and SLE-DAS, which weight organ manifestations according to their relative severity, the BILAG assigns the same grading (from A to E in the original version or from 12 to 0 in the numerical version) to each of its nine organ domains. This approach does not consider the relative severity of SLE involvement in these systems. For example, an 'A' or '12' score could result from severe mucosal ulceration (mucocutaneous domain) or fever with anorexia and weight loss (constitutional domain), and it is graded the same as severe lupus nephritis, pneumonitis or enteritis. Therefore, the BILAG score is best regarded as intrinsic to each organ system rather than as a tool for assessing global disease activity. This limitation may explain why patients classified as SDA by either SLE-DAS and BILAG-2004 or only by SLE-DAS showed similarly severe disease, while those classified as SDA only by BILAG-2004 presented with less severe disease, as assessed by both physician and patient evaluations. Additionally, while increasing the BILAG-2004 cut-off could theoretically refine its classification of severe disease, such modifications would exclude patients with single-domain 'A' scores (eg, severe lupus nephritis) unless accompanied by other high-activity domains. Since no validation studies have assessed alternative numerical thresholds for defining SDA, further research is needed to evaluate their clinical applicability.

The PGA has been proposed as an alternative method to define SDA, serving as an independent measure of

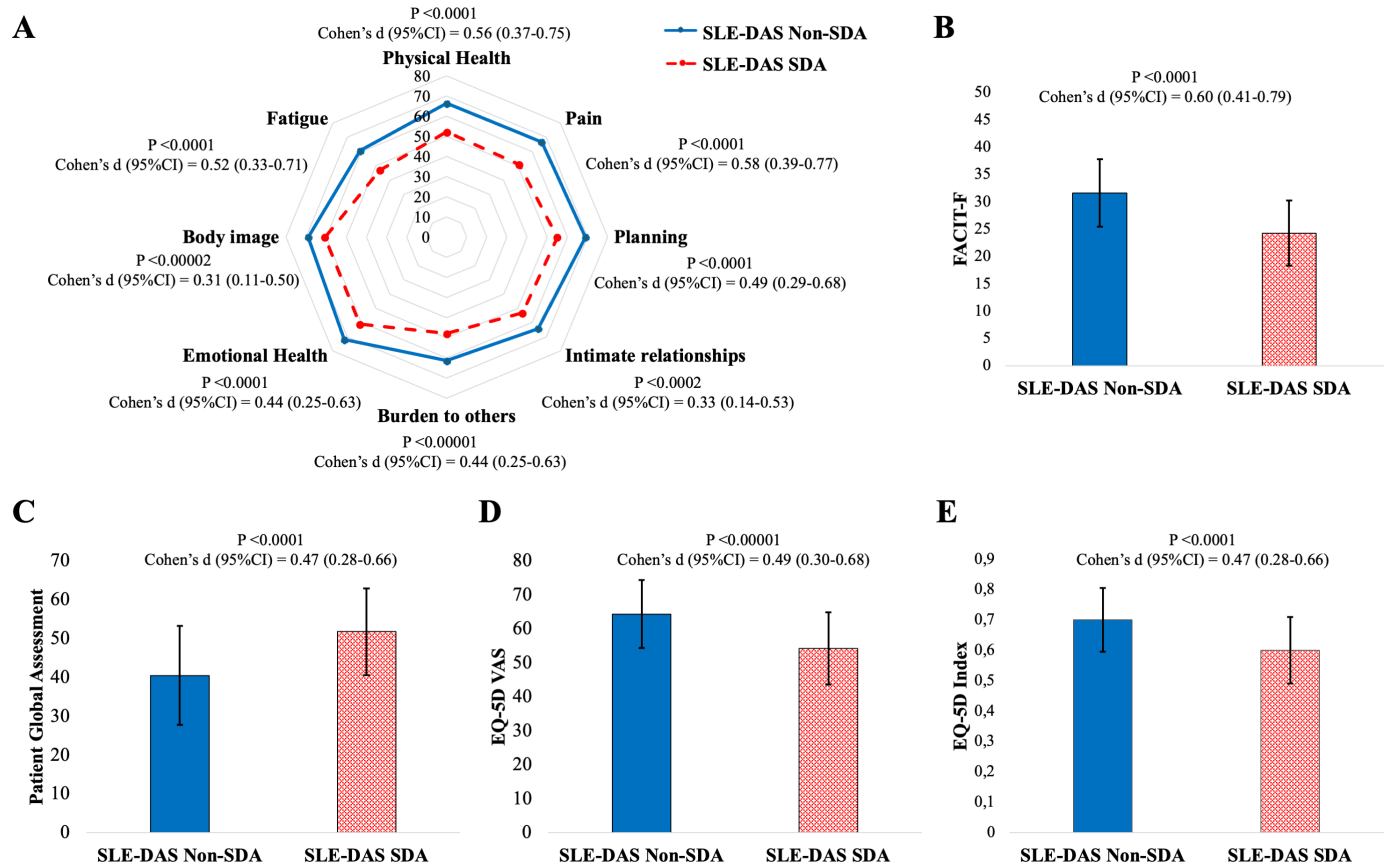


Figure 3 Health-related quality of life patient-reported outcomes among patients classified as severe disease activity (SDA) and non-SDA according to SLE Disease Activity Score (SLE-DAS) at week 12. The radial chart (A) illustrates the comparisons of Lupus Quality of Life (LupusQoL) mean domain scores; the bar charts (B), (C), (D) and (E) illustrate the comparisons of Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-F), Patient Global Assessment, EuroQol 5 Dimension (EQ-5D) visual analogue scale (VAS) and EQ-5D index, respectively. The height of the bars represents mean scores, while the whiskers represent SD. P values were calculated using the Mann-Whitney U test, and Cohen's d was assessed to compare the magnitude of these differences.

disease activity in SLE; however, its application as gold standard is limited. A systematic review highlighted a wide variability in inter-rater reliability, reducing its robustness for defining SDA.²⁹ Furthermore, as stated in the PGA International Standardisation Consensus in SLE study, the PGA should be rated by the same medical doctor at each visit, restricting its applicability in our study.³⁰ Given these limitations, we opted to use the numerical transformation of BILAG-2004, which provides a more reproducible and standardised method for defining SDA.

Unlike SLEDAI-2K, which dichotomously scores disease activity in each organ domain, and BILAG-2004, which is time consuming and provides organ-based categories, the SLE-DAS offers a global score that can be easily calculated during clinical visits. This feature enhances its feasibility and applicability in routine clinical practice, allowing for timely and accurate assessment of disease activity.

It is important to acknowledge that categorising a continuous score inevitably leads to some loss of information. However, the classification of disease activity into discrete categories plays a critical role for decision-making in clinical practice and research. This approach enables the tailoring of treatment strategies based on

disease intensity, aligning with EULAR recommendations that advocate for treatment decisions based on predefined disease activity categories.³ Additionally, categorical classifications of disease activity are essential for selecting patients in clinical trials, ensuring homogeneity in study populations and facilitating comparability across studies. Future research should further investigate the performance of these indices across their full spectrum of scores to enhance understanding of their relationship with clinical outcomes and disease burden.

The limitations of our study include the post hoc nature of the analysis and the use of pooled data from clinical trials that may limit the generalisability of our findings to broader SLE populations. However, the study is strengthened by using prospectively collected data of rigorously conducted, double-blind RCTs, in which all parameters needed to score the SLE-DAS were collected.

Despite these limitations, our study provides important insights into the assessment of SDA in SLE and underscores the potential of the SLE-DAS as a valuable tool for guiding treatment decisions and improving patient outcomes. Another key aspect of this study is the application of criterion validity in comparing disease activity

Table 2 Multiple linear regression for health-related quality of life comparison between patients classified as SDA versus non-SDA according to SLE-DAS (>9.90), BILAG-2004 (numerical score >11) and SLEDAI-2K (>12), based on backward selection

	Regression coefficient	SE	P value
PtGA			
SLE-DAS SDA	11.31	2.31	<0.001
FACIT-F			
SLE-DAS SDA	-7.31	1.17	<0.001
LupusQoL			
Physical health			
SLE-DAS SDA	-14.16	2.42	<0.001
Pain			
SLE-DAS SDA	-18.33	2.86	<0.001
SLEDAI-2K SDA	8.65	3.95	0.029
Planning			
SLE-DAS SDA	-14.11	2.79	<0.001
Intimate relationships			
SLE-DAS SDA	-10.91	3.29	0.001
Burden to others			
SLE-DAS SDA	-9.44	3.48	0.007
BILAG-2004 SDA	-7.36	3.50	0.036
Emotional health			
SLE-DAS SDA	-10.88	2.35	<0.001
Body image			
BILAG-2004 SDA	-10.46	2.64	<0.001
Fatigue			
SLE-DAS SDA	-13.68	2.54	<0.001
EQ-5D VAS			
SLE-DAS SDA	-10.15	1.98	<0.001
EQ-5D index			
SLE-DAS SDA	-0.12	0.02	<0.001
SLEDAI-2K SDA	0.07	0.03	0.019

BILAG-2004, British Isles Lupus Assessment Group 2004; EQ-5D, EuroQol 5 Dimension; FACIT-F, Functional Assessment of Chronic Illness Therapy–Fatigue; LupusQoL, Lupus Quality of Life; PtGA, Patient Global Assessment; SDA, severe disease activity; SLEDAI-2K, SLE Disease Activity Index 2000; SLE-DAS, SLE Disease Activity Score; VAS, visual analogue scale.

classifications. While BILAG-2004 was used to derive the SLE-DAS SDA cut-off, the comparison between SLE-DAS, SLEDAI-2K and BILAG-2004 SDA definitions was conducted through HR-QoL PROs, an external and clinically relevant criterion reflecting patients' burden. This approach ensures that SDA classification aligns with both physician-assessed disease activity and its impact on patients. By accurately identifying patients with SDA, the SLE-DAS can help clinicians tailor therapeutic strategies

more effectively, ultimately enhancing the quality of life for patients with SLE. Additionally, our methodology ensured a comprehensive validation of the SLE-DAS SDA threshold by including patients across the full spectrum of disease activity. This approach was necessary to avoid selection bias and ensure that the ROC analysis accurately defined the optimal cut-off for distinguishing severe from non-severe disease.

Future research should focus on further validating the SLE-DAS SDA cut-off in diverse patient populations and exploring its application in different clinical scenarios. Additionally, longitudinal studies are needed to assess the impact of using the SLE-DAS SDA on long-term patient outcomes and to evaluate its responsiveness to changes in disease activity over time.

In conclusion, the SLE-DAS SDA definition is a practical and accurate measure for identifying SDA in SLE, presenting high discriminative power and consistent associations with worse HR-QoL across multiple PROs. These findings highlight its potential to enhance clinical management and improve patient outcomes. The SLE-DAS offers a practical and comprehensive tool for assessing disease activity, addressing key limitations of existing measures and providing a valuable addition to the armamentarium for treat-to-target management of SLE.

Author affiliations

¹Rheumatology Department, Unidade Local de Saude da Regiao de Leiria, Leiria, Portugal

²Faculdade de Ciências da Saúde, Universidade da Beira Interior, Covilha, Portugal

³Polytechnic Institute of Viseu, Viseu, Portugal

⁴Research Centre in Digital Services (CISeD), Viseu, Portugal

⁵Centre for Mathematics, University of Coimbra, Coimbra, Portugal

⁶Division of Rheumatology, University of Padua, Padua, Italy

⁷Rheumatology Department, Unidade Local de Saúde de Coimbra, Coimbra, Portugal

Acknowledgements This publication is based on research using data from data contributor AstraZeneca that has been made available through Vivli.

Contributors DJ and LSI contributed to the conception and design of the project, analysis and interpretation of data, and drafting and critical revision of the manuscript. DJ contributed to data management. AM and CH contributed to the design of the project, statistical analysis, and interpretation of data, and critically revised the manuscript. AD contributed to the design of the project, interpretation of data, and drafting and critical revision of the manuscript. LSI takes the responsibility for the integrity of the data and the accuracy of the data analysis and is the guarantor. All the authors approved the final version of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Disclaimer Vivli has not contributed to or approved, and is not in any way responsible for, the contents of this publication.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The MUSE, TULIP-1 and TULIP-2 trials were conducted in accordance with the principles of the Declaration of Helsinki and the International Conference on Harmonization Guidelines for Good Clinical Practice, and all patients provided written informed consent in accordance with local requirements. As this was a post hoc analysis of anonymised data, no ethics committee or institutional review board approvals were required. All such approvals were obtained in the

original trials. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. This publication is based on research using data from data contributor AstraZeneca that has been made available through Vivli.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Diogo Jesus <http://orcid.org/0000-0003-3136-0722>

Ana Matos <http://orcid.org/0000-0001-6408-5857>

Carla Henriques <http://orcid.org/0000-0002-2142-2849>

Andrea Doria <http://orcid.org/0000-0003-0548-4983>

Luis Sousa Ines <http://orcid.org/0000-0003-3172-3570>

REFERENCES

- Inês LS, Fredi M, Jesus D, *et al*. What is the best instrument to measure disease activity in SLE? - SLE-DAS vs Easy BILAG. *Autoimmun Rev* 2024;23:103428.
- Jesus D, Rodrigues M, Matos A, *et al*. Performance of SLEDAI-2K to detect a clinically meaningful change in SLE disease activity: a 36-month prospective cohort study of 334 patients. *Lupus (Los Angel)* 2019;28:607–12.
- Fanouriakis A, Kostopoulou M, Andersen J, *et al*. EULAR recommendations for the management of systemic lupus erythematosus: 2023 update. *Ann Rheum Dis* 2024;83:15–29.
- Yee C-S, Farewell VT, Isenberg DA, *et al*. The use of 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 (Oxford)* 2011;50:982–8.
- Jesus D, Matos A, Henriques C, *et al*. Derivation and validation of the SLE Disease Activity Score (SLE-DAS): a new SLE continuous measure with high sensitivity for changes in disease activity. *Ann Rheum Dis* 2019;78:365–71.
- 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.
- Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. *J Rheumatol* 2002;29:288–91.
- Hay EM, Bacon PA, Gordon C, *et al*. The BILAG index: a reliable and valid instrument for measuring clinical disease activity in systemic lupus erythematosus. *Q J Med* 1993;86:447–58.
- Isenberg DA, 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.
- Jesus D, Matos A, Henriques C, *et al*. Response to: "SLE-DAS: ready for routine use" by Mathew *et al* *Ann Rheum Dis* 2020;79:e117.
- Jesus D, Matos A, Henriques C, *et al*. Response to: "Performance of the systemic lupus erythematosus disease activity score (SLE-DAS) in a Latin American population" by Rodríguez-González *et al* *Ann Rheum Dis* 2020;79:e159.
- Jesus D, Zen M, Doria A, *et al*. Response to: "Assessment of responsiveness of the musculoskeletal component of SLE-DAS in an independent cohort", by Hassan *et al* *Ann Rheum Dis* 2020;79:e52.
- Jesus D, Larosa M, Henriques C, *et al*. Systemic Lupus Erythematosus Disease Activity Score (SLE-DAS) enables accurate and user-friendly definitions of clinical remission and categories of disease activity. *Ann Rheum Dis* 2021;80:1568–74.
- Assunção H, Jesus D, Larosa M, *et al*. Definition of low disease activity state based on the SLE-DAS: derivation and validation in a multicentre real-life cohort. *Rheumatology (Oxford)* 2022;61:3309–16.
- van Vollenhoven RF, Bertsias G, Doria A, *et al*. 2021 DORIS definition of remission in SLE: final recommendations from an international task force. *Lupus Sci Med* 2021;8:e000538.
- Franklyn K, Lau CS, Navarra SV, *et al*. Definition and initial validation of a Lupus Low Disease Activity State (LLDAS). *Ann Rheum Dis* 2016;75:1615–21.
- 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.
- Furie RA, Morand EF, Bruce IN, *et al*. Type I interferon inhibitor anifrolumab in active systemic lupus erythematosus (TULIP-1): a randomised, controlled, phase 3 trial. *Lancet Rheumatol* 2019;1:e208–19.
- Morand EF, Furie R, Tanaka Y, *et al*. Trial of Anifrolumab in Active Systemic Lupus Erythematosus. *N Engl J Med* 2020;382:211–21.
- Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;40:1725.
- Yee C-S, Cresswell L, Farewell V, *et al*. Numerical scoring for the BILAG-2004 index. *Rheumatology (Oxford)* 2010;49:1665–9.
- Yee C-S, Gordon C, Isenberg DA, *et al*. The BILAG-2004 systems tally—a novel way of representing the BILAG-2004 index scores longitudinally. *Rheumatology (Oxford)* 2012;51:2099–105.
- Yee C-S, Gordon C, Isenberg DA, *et al*. Comparison of Responsiveness of British Isles Lupus Assessment Group 2004 Index, Systemic Lupus Erythematosus Disease Activity Index 2000, and British Isles Lupus Assessment Group 2004 Systems Tally. *Arthritis Care Res (Hoboken)* 2022;74:1623–30.
- Yee C-S, Gordon C, Akil M, *et al*. The BILAG-2004 index is associated with development of new damage in SLE. *Rheumatology (Oxford)* 2023;62:668–75.
- 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.
- Petri M, Buyon J, Kim M. Classification and definition of major flares in SLE clinical trials. *Lupus (Los Angel)* 1999;8:685–91.
- Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol* 2013;4:863.
- McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22:276–82.
- Chessa E, Piga M, Floris A, *et al*. Use of Physician Global Assessment in systemic lupus erythematosus: a systematic review of its psychometric properties. *Rheumatology (Oxford)* 2020;59:3622–32.
- Piga M, Chessa E, Morand EF, *et al*. Physician Global Assessment International Standardisation Consensus in Systemic Lupus Erythematosus: the PISCOS study. *Lancet Rheumatol* 2022;4:e441–9.