Ann Rheum Dis 85 (2026) 1993–2003

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Systemic lupus erythematosus

A data-driven, weighted flare score that accurately predicts organ damage in systemic lupus erythematosus

Dionysis Nikolopoulos[1,2], Sofia Pitsigavdaki[3], Danae-Mona Nöthling[4,5], Spyridon Katechis[6], Myrto Nikoloudaki[3], Christina Bergmann[4,5], Ettore Silvagni[7], Argyro Repa[3], Andreas Ramming[4,5], Antonio Marangoni[7], Marcello Govoni[7], Prodromos Sidiropoulos[3,8], Antonis Fanouriakis[6], Dimitrios Boumpas[6], Georg Schett[4,5], Alessandra Bortoluzzi[7], Panagiotis Garantziotis[4,5], George Bertsias[3,8,*]

[1] *Division of Rheumatology, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden*
[2] *Center for Molecular Medicine (CMM), Stockholm, Sweden*
[3] *Rheumatology and Clinical Immunology, University of Crete Medical School, Heraklion, Greece*
[4] *Department of Internal Medicine 3-Rheumatology and Immunology, Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg and Universitätsklinikum Erlangen, Erlangen, Germany*
[5] *Deutsches Zentrum Immuntherapie (DZI), FAU Erlangen-Nürnberg and Universitätsklinikum Erlangen, Erlangen, Germany*
[6] *4th Department of Internal Medicine, Attikon University Hospital, National and Kapodistrian University of Athens Medical School, Athens, Greece*
[7] *Rheumatology Unit, Department of Medical Sciences, University of Ferrara and Azienda Ospedaliero-Universitaria S.Anna, Ferrara, Italy*
[8] *Division of Immunity, Institute of Molecular Biology and Biotechnology-Foundation for Research and Technology – Hellas (FORTH), Heraklion, Greece*

ARTICLE INFO

Article history:
Received 30 January 2026
Received in revised form 23 May 2026
Accepted 13 June 2026

ABSTRACT

Objectives: Flares are key determinants of outcome in systemic lupus erythematosus (SLE), yet the conventional mild/moderate vs severe classification inadequately reflects prognosis. We sought to refine flare classification to improve damage prediction and develop a data-driven flare score.

Methods: We analysed a multicentre longitudinal cohort of 354 patients with active SLE requiring treatment at inclusion (median follow-up 66 months). Flares were adjudicated using a modified Safety of Estrogens in Lupus Erythematosus National Assessment-SLE Disease Activity Index Flare Index (SFI). Random Forests and modified flare definitions assessed the contribution of individual SFI variables to subsequent damage. We derived a flare score using bootstrap-validated mixed-effects models. This was validated in an independent cohort (n = 110).

Results: Both mild/moderate and severe SFI flares contributed to elevated 12-month damage risk (incidence rate ratio [IRR]:1.42 and 1.54, respectively). Among SFI subcriteria, major organ involvement requiring high-dose glucocorticoids (GCs) ranked highest for damage prediction

*Correspondence to Prof. George Bertsias, Rheumatology and Clinical Immunology, University of Crete Medical School and University Hospital of Heraklionm, 71008 Voutes, Heraklion, Greece

E-mail address: gbertsias@uoc.gr (G. Bertsias).

Handling editor Josef Smolen.

https://doi.org/10.1016/j.ard.2026.06.013

0003-4967/© 2026 The Author(s). Published by Elsevier B.V. on behalf of European Alliance of Associations for Rheumatology (EULAR). This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

(weight = 6), followed by the isolated GCs escalation criteria (weights = 1.5-3), whereas minor organ involvement and immunosuppressants/biologics addition carried lower weights (1 each). A flare score based on these subcriteria demonstrated good predictive performance (Brier score = 0.07). A threshold of ≥ 3.5 identified a subset of high-risk flares (40% of all episodes) that displayed the strongest association with subsequent damage (IRR: 2.49, 95% CI: 1.51-4.10) and outperformed the original SFI definition in the validation cohort (area under the curve −receiver operating characteristic: 0.893 vs 0.816). Although both Definitions of Remission in SLE (DORIS) and lupus low disease activity state correlated with lower risk of SFI-defined flares, only sustained DORIS (>50% of time) reduced high-risk flares.

Conclusions: We defined and validated a novel weighted SLE flare score that better predicts damage accrual and may support treat-to-target implementation and clinical trial design.

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Flares are major determinants of organ damage in systemic lupus erythematosus (SLE).
- The Safety of Estrogens in Lupus Erythematosus National Assessment-SLE Disease Activity Index (SELENA-SLEDAI) Flare Index (SFI) distinguishes mild/moderate from severe flares but does not adequately weight their long-term impact on damage.

WHAT THIS STUDY ADDS

- Using longitudinal multicentre data, both mild/moderate and severe SFI events were independently associated with subsequent organ damage.
- Data-driven deconvolution of the SFI identified key subcriteria spanning both mild/moderate and severe flares (new/worsened organ manifestations, glucocorticoid [GC] escalation, treatment intensification) as most strongly associated with damage.
- A weighted flare score (range 1-8) based on these features defined a subgroup of high-risk flares associated with subsequent organ damage—including damage classified as activity-related as well as GC-related—and outperformed conventional SFI flares in an independent cohort.
- Sustained Definition of Remission in SLE remission, but not intermittent low disease activity, was associated with reduced occurrence of such high-risk flares.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

- The new SLE flare definition based on clinical relevance provides a more accurate, outcome-based stratification of flare severity, distinguishing events linked to irreversible organ damage.
- Incorporation of this weighted score may improve treat-to-target monitoring and serve as an endpoint in SLE clinical studies.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic autoimmune disorder characterised by recurrent flares of disease activity that typically require therapeutic intervention, alongside periods of relative quiescence often sustained by ongoing immunosuppressive treatment [1]. Flares vary in frequency, severity, and organ distribution, reflecting the disease heterogeneity [2]. Their prevention represents a major goal in SLE management, which is reflected by the inclusion of flares as a key secondary endpoint in trials of novel therapeutic agents [3].

The clinical assessment of SLE flares remains challenging. The Safety of Estrogens in Lupus Erythematosus National Assessment-SLE Disease Activity Index Flare Index (SFI) is the most widely applied framework for their definition and stratification [2,4–6]. However, the index was originally designed for clinical trial standardisation rather than longitudinal risk stratification. Importantly, its dichotomous nature (mild/moderate vs severe) may not accurately reflect the prognostic heterogeneity of disease exacerbations. Furthermore, SFI integrates both clinical and therapeutic criteria without adequately weighting their relative contribution to relevant long-term outcomes, potentially leading to misclassification of flare severity in routine clinical practice. These limitations underscore the need for data-driven approaches capable of distinguishing flares that are clinically relevant and lead to irreversible damage from those with transient or limited clinical impact.

To address these unmet needs, we conducted a retrospective multicentre study exploiting a harmonised, multicentre longitudinal SLE cohort. We sought to dissect the spectrum of flare severity, quantify their temporal patterns and outcomes, and derive a composite flare score based on objective associations with organ damage accrual. We identified score thresholds linked to excess damage accumulation. Importantly, we validated the prognostic significance of these high-risk flares in an independent SLE cohort. By refining the conceptual and operational definition of flares, our study aims to enhance prognostic precision, facilitate treat-to-target implementation, and inform the design of future clinical trials in SLE.

METHODS

Study design, inclusion criteria, and participants

This was a retrospective longitudinal cohort study of patients who were consecutively enrolled and followed at 3 tertiary lupus clinics. All centres maintain harmonised SLE registries [7] and adhere to standardised protocols for clinical and laboratory data collection. Eligible participants were aged ≥ 16 years and fulfilled the 2012 Systemic Lupus International Collaborating Clinics (SLICC) [8] and/or the 2019 European Alliance of Associations for Rheumatology/American College of Rheumatology (EULAR/ACR) [9] classification criteria for SLE. To ensure inclusion of patients with clinically active disease, entry required a SLE disease activity index 2000 (SLEDAI-2K) [10] ≥ 6 and/or a Physician Global Assessment (PGA) [11] ≥ 1.5, accompanied by treatment intensification (Supplementary Methods). This strategy was chosen to avoid the mixture of long-standing quiescent and active cases.

Patients with concomitant systemic autoimmune/inflammatory diseases other than SLE were excluded. The baseline visit was defined as the first encounter fulfilling inclusion criteria within the registry period (January 2008-June 2018). Per protocol, follow-up visits were scheduled every 4 to 6 months during the first year and at least annually thereafter. In practice,

observed intervals were even shorter during the first year, yielding a median interval of 3 months (IQR: 3); this aligned with our protocol and reflected the close monitoring required for early, active disease. Data were recorded prospectively using structured electronic forms and retrospectively verified for completeness and accuracy.

Clinical assessment, flares adjudication, and variables collection

Baseline assessments included demographic characteristics, disease duration, prior treatments, and major comorbidities. At each visit, disease activity was evaluated with SLEDAI-2K and PGA (0-3 scale), scored prior to review of serology. Flares were defined using the SELENA-SLEDAI Flare Index (SFI) [12], modified to include contemporary therapies under the definition of severe flare (Supplementary Table S1). The SFI was selected as the primary flare definition because it is a composite instrument integrating SLEDAI changes, organ manifestations (beyond those captured by the SLEDAI), treatment intensification, and physician judgement (PGA), and has been associated with damage accrual in prior cohorts [13,14].

All individual subcriteria from both the mild/moderate and severe SFI definitions were recorded at each visit, irrespective of whether the corresponding flare category was ultimately met. Although the dose-escalation rules for the glucocorticoid (GC) −related severe subcriteria 8 and 9 differ subtly (Supplementary Table S1), in our dataset, all events fulfilling subcriterion 8 also reached ≥ 30 mg/d and would therefore have met the subcriterion 9 dose threshold. To avoid double counting, subcriterion 9 was not scored separately when subcriterion 8 was fulfilled; consequently, the 2 subcriteria differed in practice only by the presence of major organ involvement.

Organ damage was assessed using the SLICC/ACR Damage Index (SDI) [15]. Treatment exposure, including GCs (dose and route), hydroxychloroquine, and other immunosuppressive/biologic agents, was recorded at each visit. Definition of Remission in SLE (DORIS) [16] and lupus low disease activity state (LLDAS) [17] was determined according to established definitions. Laboratory parameters required for SLEDAI-2K scoring were obtained within 30 days of clinical assessment. All definitions and scoring procedures were harmonised across centres and followed the methodology previously described [18,19]. Data were stored in a secure, pseudonymised registry with ongoing quality control and cross-validation between centres.

Construction of the flare score

Random Forest models were trained to evaluate the relative importance of individual SFI subcriteria for SDI progression. Models used an 80/20 train-test split, 5-fold cross-validation, and 1000 trees optimised for the Receiver operating characteristic (ROC) area under the curve (AUC); permutation-based importance quantified the influence of features. In addition, mixed-effects Poisson models with 1000 bootstrap replications were fitted to implement a set of sensitivity analyses in which modified flare definitions were constructed by systematically altering the SFI structure: 'ignore-one' models omitted each subcriterion from the full definition; 'absent-one' models were restricted to flare episodes that did not fulfil that subcriterion; and 'present-one' models were restricted to flare episodes in which the subcriterion was fulfilled. Log-likelihood differences between each modified and the original SFI model were then used to quantify the influence of each subcriterion on the flare-damage association. Top-ranked predictors from both approaches were combined to construct reduced flare definitions, and retained subcriteria were jointly entered into a final model to derive a unified weighted flare score. Regression coefficients were standardised and converted into integer weights, and model calibration and accuracy were assessed using the Hosmer-Lemeshow test and Brier score.

Statistical analysis

Incidence rates (IRs) of flares were calculated per 100 person-years with exact Poisson 95% CIs, and temporal patterns of flare occurrence were examined using Kaplan-Meier curves, cumulative incidence estimates at specific timepoints and yearly IRs. Visits were classified as flare-free, mild/moderate flare, or severe flare, and the incidence of new damage within the subsequent 12 months was calculated. In addition, mixed-effects Poisson regression with a random intercept for each patient was used to account for repeated observations and to estimate adjusted incidence rate ratios (IRRs) for damage following each flare type. Biological sex (female or male), baseline disease activity, organ damage, and SLE duration were included as covariates. The prognostic performance of high-risk flares was further examined in an independent SLE cohort (Erlangen) using Poisson regression and ROC-AUC. To determine whether early treatment−target attainment protected against future flares, a 12-month landmark design was used: for each patient, the cumulative proportion of time spent in DORIS or LLDAS during the first 12 months was grouped into predefined strata, and Cox models estimated the hazard of severe or high-risk flares occurring beyond 12 months. All analyses were performed using Stata v18.0 (StataCorp LLC, College Station, TX, USA) and R v4.5.1 (R Foundation for Statistical Computing, Vienna, Austria).

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.

RESULTS

High flare burden in SLE during long-term follow-up

We included 354 patients (92.5% female) with a median (IQR) age of 45.7 (21.3) years and disease duration of 8.7 (66.4) months (Table 1). At baseline, all patients had active disease with median (IQR) PGA, SLEDAI-2K, and clinical SLEDAI-2K of 2.0 (0.5), 8 (4), and 6 (4), respectively. Over a median follow-up of 66 (IQR: 60) months and excluding the baseline visit, 290 patients (81.9%) experienced at least 1 flare, including 242 (68.4%) with mild/moderate and 210 (59.3%) with severe flares adjudicated according to the SFI (Table 1). The estimated IRs (per 100 person-year) were 47.7 (95% CI: 44.8-50.8), 27.0 (95% CI: 24.8-29.3), and 20.7 (95% CI: 18.9-22.8) for any, mild/moderate and severe flares, respectively.

To characterise the dynamics of flare occurrence, Kaplan-Meier survival curves and IR plots were generated over the entire follow-up period. The cumulative probability of mild/moderate flares was 23.8%, 43.6%, and 55.6% at 12, 24, and 36 months (Fig 1A), respectively, whereas that of severe flares was 18.3%, 28.2%, and 34.6% at the same time points (Fig 1B). The corresponding IR of mild/moderate flares remained relatively stable across follow-up (Fig 1C). By contrast, severe flares occurred most frequently during the first year (45.4 per 100 person-years) and declined thereafter (IRRs: 0.37-0.64 across

Table 1
Demographic and clinical characteristics of patients with SLE at inclusion visit

	N (%) or median (IQR)	
Inclusion visit data		
No. patients	354	
Sex (female)	328 (92.6%)	
Ethnicity (White)	340 (96.1%)	
Age (y)	45.7 (21.3)	
Disease duration (mo)	8.7 (66.4)	
Organ damage (SDI > 0)	94 (26.6%)	
PGA (0-3)	2.0 (0.5)	
SLEDAI-2K	8 (4)	
Clinical SLEDAI-2K	6 (4)	
Follow-up data		
No. visits per patient	12 (9)	
Follow-up, per patient (mo)	66 (60)	
Flare (SFI) events (excl. baseline)		
Any (at least 1; no. events)	290 (81.9%); 3 (4)	
Mild/moderate	242 (68.4%); 1 (3)	
Severe	210 (59.3%); 1 (2)	
Flares incidence rate (per 100 person-year)		
Any	47.7 (95% CI: 44.8-50.8)	
Mild/moderate	27.0 (95% CI: 24.8-29.3)	
Severe	20.7 (95% CI: 18.8-22.8)	
Use of glucocorticoids (inclusion)		
Oral route	219 (61.9%)	
Dose (mg prednisone-equivalent; among users)	15.0 (22.0)	
Intramuscular route	42 (11.9%)	
Dose (mg methylprednisolone acetate; among users)	160 (40)	
Intravenous route	91 (25.7%)	
Dose (mg methylprednisolone; among users)	1000 (500)	
Use of other treatments	At inclusion, *n* (%)	Ever, *n* (%)
Hydroxychloroquine	282 (79.7)	319 (90.1)
Methotrexate	80 (22.6)	163 (46.0)
Leflunomide	8 (2.3)	42 (11.9)
Azathioprine	83 (23.4)	169 (47.7)
Calcineurin inhibitor	13 (3.7)	27 (7.6)
Thalidomide	0 (0.0)	5 (1.4)
Mycophenolate	32 (9.0)	108 (30.5)
Cyclophosphamide	21 (5.9)	44 (12.4)
Belimumab	29 (8.2)	112 (31.6)
Rituximab	8 (2.3)	47 (13.3)
IVIG	2 (0.6)	7 (2.0)
Other biologics	0 (0.0)	14 (4.0)

IVIG, intravenous immunoglobulin; PGA, SELENA-SLEDAI Physician Global Assessment; SDI, SLICC/ACR damage index; SFI, Safety of Estrogens in Lupus Erythematosus National Assessment- systemic lupus erythematosus Disease Activity Index Flare Index; SLE, systemic lupus erythematosus; SLEDAI-2K, SLE disease activity index 2000.

subsequent years, all *P*-values < .02) (Fig 1D). Notably, even after 10 years of observation, the IR of severe flares remained notable at 18.2 (95% CI: 13.2-24.6) per 100 person-years. Collectively, these findings demonstrate that the overall flare burden peaked within the first year after inclusion and progressively stabilised over time, particularly for severe flares.

Flares are particularly common in younger and female patients

We next stratified the occurrence of SLE flares according to baseline demographic and disease-related parameters (Supplementary Table S2). Males had significantly fewer mild/moderate flares than females (12.9 vs 28.1 per 100 person-years; IRR: 0.46, 95% CI: 0.29-0.72), whereas the frequency of severe events was comparable between sexes. Compared with younger patients (< 45 lt; 45 years), those aged ≥ 45 years experienced fewer severe flares (17.2 vs 24.4 per 100 person-years; IRR: 0.71, 95% CI: 0.58-0.85), whereas mild/moderate flare rates did not differ across age groups. Disease duration (> 1 gt; 1 vs ≤ 1 year) showed no significant associations. Although clinical disease activity at inclusion was not strongly related to subsequent flares, active serology (high anti-dsDNA and/or low complement) was linked to reduced flare occurrence, particularly of mild/moderate intensity (IRR: 0.66; 95% CI: 0.55-0.78).

Both mild/moderate and severe SFI flares contribute to damage accrual

Flares are recognised drivers of adverse long-term outcomes in SLE; however, the relative prognostic impact of mild/moderate vs severe events remain uncertain. To address this, we classified all patient visits according to flare status—no flare, mild/moderate flare, or severe flare—and then assessed the incidence of new organ damage occurring within the subsequent 12 months. Following flare-free visits, new damage developed at a rate of 12.9 per 100 person-years (95% CI: 11.3-14.7). This rate increased to 19.3 (95% CI: 14.6-25.0) and 25.4 (95% CI: 19.6-32.4) per 100 person-years after visits with mild/moderate and severe flares, respectively (Supplementary Table S3). In a more robust analysis, mixed-effects Poisson regression was applied to account for repeated observations within individuals. Both mild/moderate and severe SFI flares were associated with increases in SDI documented at subsequent visits, with a tendency towards stronger effects for severe flares (Table 2). Similar results were obtained when both flare types were included simultaneously in the model. By computing a modified SLE-DAS [20] (mSLE-DAS; Supplementary Methods) at each visit and defining flares as an increase of ≥ 1.72 since the previous visit [21], we observed that SFI-defined flares performed at least as well as mSLE-DAS–defined flares in predicting subsequent damage (Supplementary Tables S4 and S5). Collectively, these findings suggest that irrespective of their classification, flares contribute to damage accrual.

Identification and weighting of flare subcriteria with prognostic impact

The SFI is a composite measure integrating disease activity scores, clinical manifestations, and treatment changes. We sought to delineate the relative contribution of its individual subcriteria to the risk of organ damage, as this could facilitate patient risk stratification. Because moderate-to-strong co-segregation was observed among several subcriteria (Supplementary Fig S1), we employed 2 complementary strategies to rank their prognostic importance. First, Random Forest models were applied, as they accommodate possible nonlinear relationships and interdependencies between variables while minimising overfitting. Second, we evaluated the dependency of the original mild/moderate and severe SFI definitions on each subcriterion using 3 sensitivity analyses: (1) 'ignore-one' models excluding each subcriterion sequentially from the full definition; (2) 'absent-one' models restricted to flare episodes not fulfilling the subcriterion; and (3) 'present-one' models restricted to flares in which the subcriterion was present (Supplementary Table S6).

Both approaches yielded convergent results. Among the severe flare components, subcriterion 8 (new/worsened major organ manifestations requiring GCs intensification) and subcriterion 10 (immunosuppressants/biologics addition) emerged as

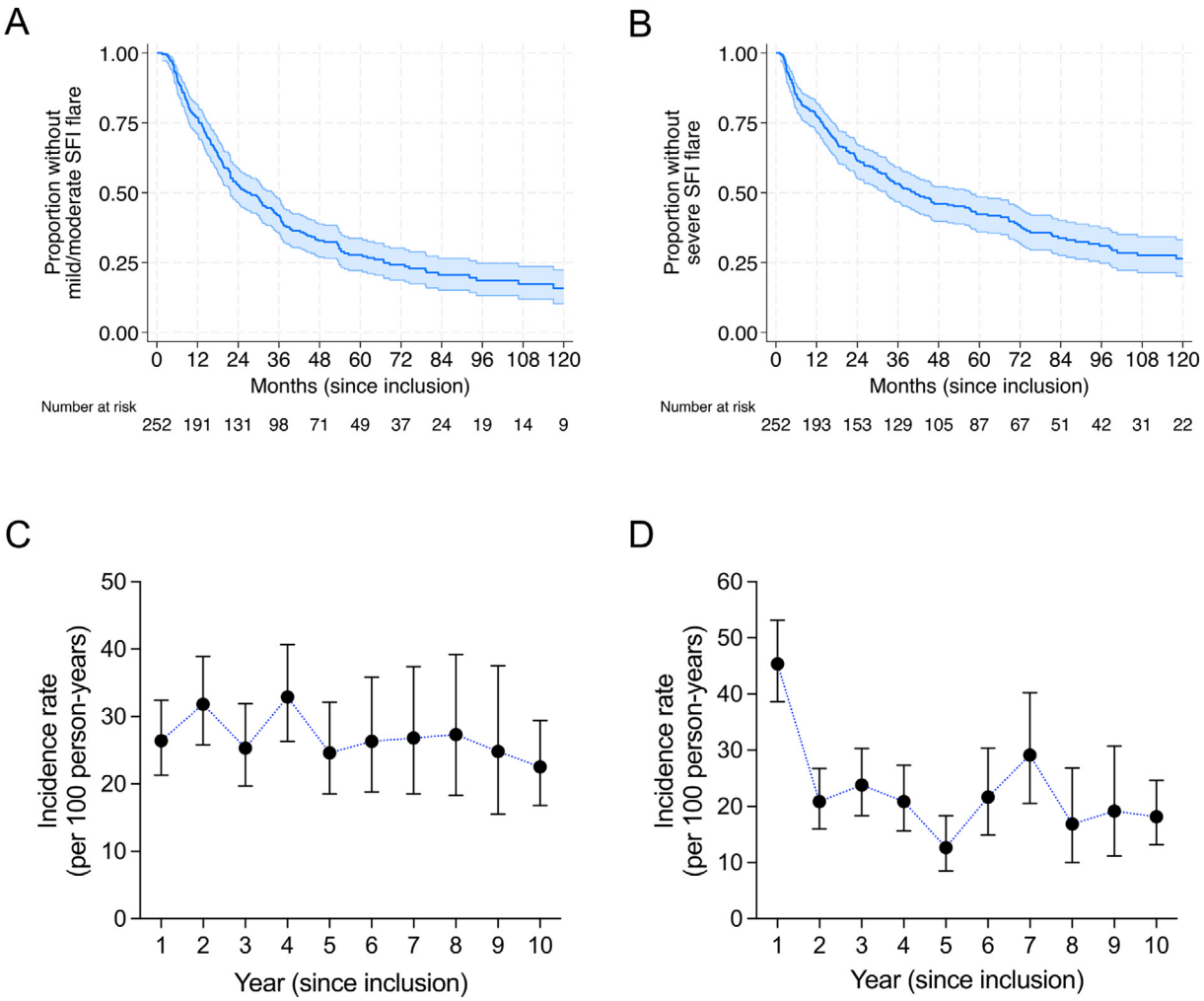


Figure 1. Frequency, incidence, and timing of SLE flares during follow-up. Kaplan-Meier curves show the time to first flare (cumulative probability) of mild/moderate flares (A) and severe flares (B) from study inclusion onwards. Numbers at risk are shown below each panel. Yearly incidence rates (per 100 person-years) of mild/moderate flares (C) and severe flares (D) during follow-up are shown with 95% CIs. Flares were assessed according to the modified SFI. SFI, Safety of Estrogens in Lupus Erythematosus National Assessment—systemic lupus erythematosus Disease Activity Index Flare Index; SLE, systemic lupus erythematosus.

Table 2

Association of mild/moderate and severe SFI flares with the risk for subsequent organ damage accrual in patients with SLE

	Mild/moderate SFI: yes vs no			Severe SFI: yes vs no		
SDI increase at subsequent visit(s)	IRR	95% CI	*P* value	IRR	95% CI	*P* value
Within 6 mo	1.39	0.89-2.18	.154	1.74	1.09-2.79	.021
Within 12 mo	1.42	1.07-1.88	.014	1.54	1.10-2.16	.012
Within 18 mo	1.20	0.96-1.52	.114	1.52	1.15-2.01	.004
Within 24 mo	1.20	0.97-1.49	.089	1.51	1.19-1.94	.001
	Mild/moderate SFI vs no-flare visit			Severe SFI vs no-flare visit		
SDI increase at subsequent visit(s)	IRR	95% CI	*P* value	IRR	95% CI	*P* value
Within 6 mo	1.52	0.94-2.48	.089	1.89	1.12-3.21	.017
Within 12 mo	1.53	1.13-2.05	.005	1.66	1.15-2.39	.007
Within 18 mo	1.30	1.02-1.65	.033	1.58	1.16-2.14	.003
Within 24 mo	1.30	1.04-1.62	.022	1.55	1.19-2.02	.001

IRR, incidence rate ratio; SDI, SLICC/ACR damage index; SFI, SELENA-SLEDAI Flare Index.
Mixed-effect Poisson regression with a random intercept for each patient was used to account for repeated observations and to estimate adjusted IRRs for damage increase within the subsequent 12 months following each flare type. Sex, baseline disease activity, organ damage and SLE duration were included as covariates.

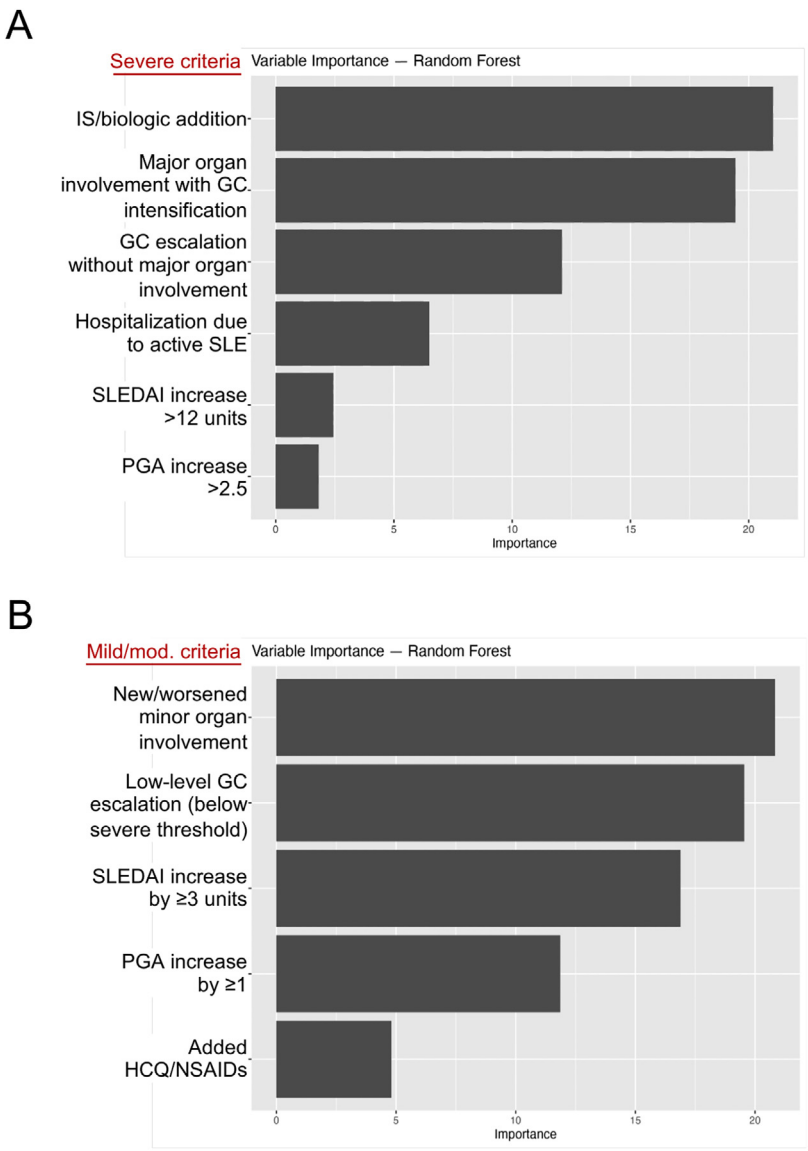


Figure 2. Data-driven prioritisation of SFI subcriteria according to their prognostic relevance for organ damage. Permutation-based variable importance rankings from Random Forest models predicting subsequent organ damage accrual are presented for severe SFI subcriteria (A) and mild/moderate SFI subcriteria (B). Importance reflects the relative influence of each individual subcriterion on subsequent organ damage accrual as determined by permutation-based measures. GC, glucocorticoid; HCQ, hydroxychloroquine; IS, immunosuppressant; NSAIDs, nonsteroidal anti-inflammatory drugs; SFI, Safety of Estrogens in Lupus Erythematosus National Assessment- systemic lupus erythematosus Disease Activity Index Flare Index; SLE, systemic lupus erythematosus; SLEDAI, SLE Disease Activity Index PGA; Physician Global Assessment.

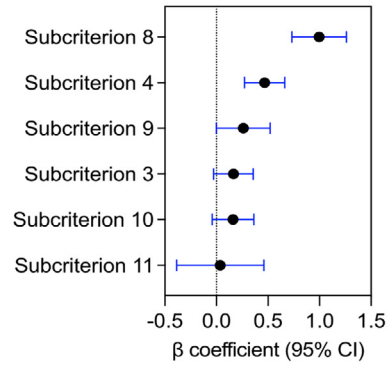
the most influential, followed by subcriterion 9 (GCs addition/escalation) and subcriterion 11 (hospitalisation due to active SLE) (Fig 2A, Supplementary Table S6). For mild/moderate flares, subcriterion 3 (new/worse clinical manifestations) and subcriterion 4 (increased or added GCs) ranked highest, followed by subcriterion 1 (increase in SLEDAI) and subcriterion 2 (increase in PGA) (Fig 2B, Supplementary Table S6).

Guided by these findings, we next constructed simplified flare definitions by iteratively combining top-ranked subcriteria (Supplementary Table S7). For severe flares, a Poisson model restricted to subcriteria 8 and 10 retained most of the prognostic information of the original SFI, with slightly higher overall accuracy (83.1% vs 81.0%). Including subcriteria 9 and 11 further improved the model β-coefficient and overall fit, yielding the best balance between sensitivity and specificity (Supplementary Table S7). For mild/moderate flares, a simplified definition comprising subcriteria 3 and 4 retained the prognostic value of the original SFI while modestly improving overall model performance (accuracy 76.2% vs 75.1%). When the remaining subcriteria 1 and 2 were added to this basic model, they showed negative β-coefficients and no improvement in model fit, and were therefore excluded (Supplementary Table S7). Collectively, reduced models containing only 4 of 6 SFI subcriteria for severe flares and 2 of 5 SFI subcriteria for mild/moderate flares preserve the prognostic performance of the full definitions while improving simplicity.

Development of a weighted SLE flare score

Building on our findings that both mild/moderate and severe flares are associated with an increased likelihood of organ damage accrual, and having identified the most impactful SFI subcriteria underlying this relationship, we sought to develop a unified weighted flare score. Retained subcriteria from both the mild/moderate and severe SFI definitions were jointly entered into a Poisson regression model with 1000 bootstrap replications to ensure coefficient stability (Fig 3A). Subcriterion 11 (hospitalisation due to active SLE) had no measurable effect and was excluded from further weighting. Based on the relative β-coefficients, weighting factors were assigned proportionally (Fig 3B), providing the basis for the new flare score, which ranged from 1 to 8. Decomposition of subcriterion 9 into oral-only and intravenous methylprednisolone (IV-MP)-containing events showed a stronger association with subsequent damage for the former (β = 0.36 vs 0.17; Supplementary Table S8). However, because 81% of episodes were IV-MP–containing, the overall subcriterion was dominated by this lower-risk

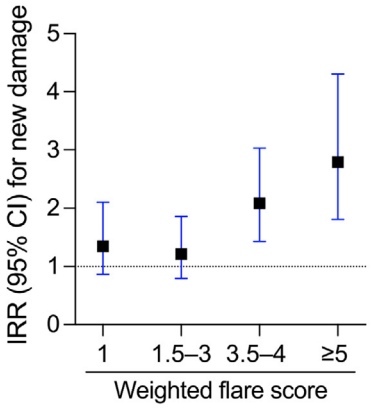
A



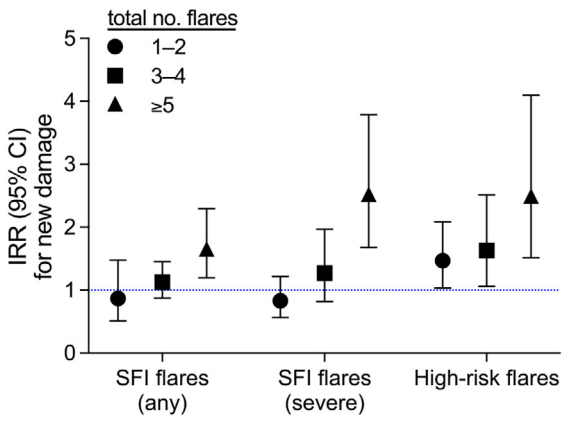
B

Subcriterion	β co-efficient	Relative weight	Rounded factor
Major organ involvement with GC intensification	0.9962	6.2	6
Low-level GC escalation (below severe threshold)	0.4677	2.9	3
GC escalation without major organ involvement	0.2604	1.6	1.5
New/worsened minor organ involvement	0.1645	1.0	1
Immunosuppressant/biologic addition	0.1608	1.0	1
Hospitalization due to active SLE	0.0376	-	-

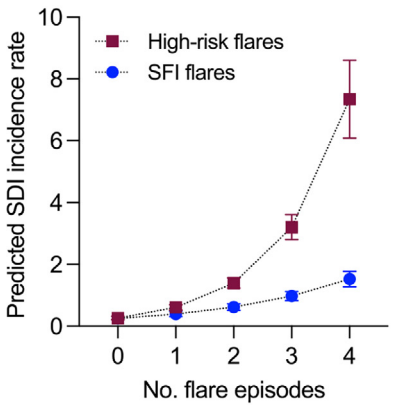
C



D



E



F

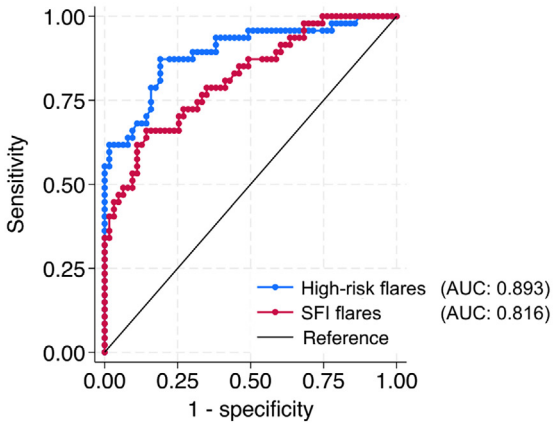


Figure 3. Derivation and validation of the weighted SLE flare score and high-risk flares. Standardised regression coefficients (95% CIs) from the bootstrap-validated mixed-effects Poisson model including the retained SFI subcriteria associated with subsequent damage accrual (A). Relative contribution of each retained SFI subcriterion, expressed as proportional weights and converted into rounded integer factors for construction of the final SLE flare score. Because severe SFI subcriterion 8 (major organ involvement requiring GC intensification) and subcriterion 9 (standalone GC escalation) were mutually exclusive in our dataset, the 2 subcriteria differ only by the presence of major organ involvement (B). Incidence rate ratios (IRRs) for new organ damage across predefined SLE flare score categories (1, 1.5-3, 3.5-4, and ≥ 5) in the derivation cohort (C). IRRs for new organ damage according to the number (1-2, 3-4, or ≥ 5) of any SFI flares, severe SFI flares, and high-risk flares (D). Model-predicted incidence rates of damage accrual (increase in SDI by ≥ 1) according to increasing numbers of high-risk flares compared with severe SFI flares in the external validation cohort (E). Receiver operating characteristic curves (ROC-AUC) in the external validation cohort comparing high-risk flares and severe SFI flares for discrimination of any increase in SDI (≥ 1) (F). GC, glucocorticoid; IRR, incidence rate ratio; ROC-AUC, receiver operating characteristic—area under the curve; SDI, SLICC ACR Damage Index; SFI, Safety of Estrogens in Lupus Erythematosus National Assessment-systemic lupus erythematosus Disease Activity Index Flare Index; SLE, systemic lupus erythematosus.

phenotype. As decomposition did not improve model fit, we retained the original SFI structure for simplicity. The Hosmer-Lemeshow goodness-of-fit test indicated no evidence of model miscalibration ($\chi^2[7] = 4.72$, $P = .69$), and the Brier score was 0.07, well within the acceptable range (0-0.25) and close to 0, indicating good overall predictive accuracy.

Operational use of the SLE flare score and definition of high-risk flares

Examination of the distribution of the new flare score across all longitudinal visits led to the formation of 4 approximately even categories: A (score = 1), B (scores 1.5−3), C (scores 3.5-4), and D (scores ≥ 5). Compared with flare-free intervals, the risk of new organ damage within a rolling window of 12 months increased progressively across higher flare score categories, reaching statistical significance for categories C ($\beta = 0.73$, $P < .001$) and D ($\beta = 1.03$, $P < .001$) (Fig 3C). Overall, 40.6% of flare episodes had a weighted score ≥ 3.5, hereafter referred to as high-risk flares. These flares could arise from either severe or mild/moderate SFI subcriteria, or from mixed combinations thereof (Supplementary Table S9), suggesting that damage-inducing flare activity is not confined to conventionally defined severe or mild/moderate SFI events. By classifying damage as definitely, possibly, or not related to GC [22], we observed that high-risk flares were associated with increased accrual risk across all 3 categories, with the strongest association for steroid-independent damage (Supplementary Table S10).

At the operational level, using a >0 threshold (any flare) preserved comparable sensitivity with the original SFI model, while marginally improving specificity (73.6% vs 72.9%) (Supplementary Table S11). Raising the threshold to ≥ 3.5 (high-risk flares), further enhanced specificity for predicting future damage (89.6%) and improved overall model fit.

Association of high-risk flares with organ damage and external validation

We assessed the cumulative, patient-level impact of flare burden, testing whether individuals who experienced repeated high-risk flares had greater overall damage accrual during the entire follow-up as compared with those meeting the original SFI definition. As shown in Figure 3D, the risk for damage accumulation rose progressively with increasing number of high-risk flares, becoming significant even after 1 or 2 events (IRR: 1.47, 95% CI: 1.03-2.08) and further increasing for patients with 3 to 4 (IRR: 1.63, 95% CI: 1.06-2.52) or ≥ 5 events (IRR: 2.49, 95% CI: 1.51-4.10). By contrast, SFI flares showed a significant effect only after ≥ 5 episodes (IRR: 2.52, 95% CI: 1.68-3.78 for severe flares).

Using an independent cohort of 110 patients with SLE (Supplementary Table S12) observed over a median (IQR) of 66.3 (34.7) months, high-risk flares were more strongly associated with increases in SDI than SFI flares (IRR: 2.29, 95% CI 1.99-2.64 vs 1.57, 95% CI 1.34-1.84) (Fig 3E). Consistently, the discriminative ability for detecting any increase in SDI (≥ 1 vs 0) was higher for high-risk flares (AUC: 0.89, 95% CI: 0.83-0.96) compared with SFI flares (AUC: 0.82, 95% CI: 0.74-0.90) (Fig 3F). Collectively, these findings reinforce the high discriminatory performance of the high-risk flare definition, capturing the prognostic impact of even a limited number of high-intensity flare episodes.

High-risk flares may be prevented only by sustained attainment of treatment targets

Achieving DORIS or LLDAS has been linked to reduction of severe SFI flares [19]. To determine whether this association extends to high-risk flares, we quantified for each patient in the derivation cohort the cumulative proportion of time spent in either target during the first 12 months after inclusion. Cox-regression models were then used to assess the risk of flares occurring beyond the 12-month landmark, stratified by predefined target duration thresholds. Spending >25% of the first year in DORIS was associated with a significantly lower hazard of subsequent severe SFI (hazard ratio [HR]: 0.40, 95% CI: 0.23-0.68) and high-risk flares (HR: 0.54, 95% CI: 0.34-0.85), whereas no significant effect was observed for smaller exposure (> 0%-25%) (Fig 4A,B). Attainment of LLDAS during the first 12 months showed a graded protective effect against severe SFI flares, with borderline benefit for > 0% to 25% and > 25% to 50% exposure, and a significant risk reduction when LLDAS was maintained for > 50% of the first year (HR: 0.55, 95% CI: 0.35-0.84). In contrast, LLDAS exposure did not meaningfully reduce the hazard of high-risk flares, irrespective of duration, including LLDAS > 50% of time (HR: 0.85, 95% CI: 0.57-1.27) (Fig 4C,D).

DISCUSSION

In this study, we show that the prognostic impact of SLE flares varies according to the specific SFI subcriteria involved. Weighting subcriteria derived from both mild/moderate and severe SFI classification identified new or worsened organ manifestations, GC escalation, and addition of immunosuppressants/biologics as key features of high-risk flares. These flares were associated with a markedly higher risk of subsequent organ damage, outperforming conventional SFI definitions, and remained prognostically informative already after only 1 or 2 events. Notably, high-risk flares occurred across both mild/moderate and severe SFI-defined categories, indicating that this conventional dichotomy does not fully capture the heterogeneity of flare severity. In this context, our findings suggest that sustained DORIS remission, rather than intermittent low disease activity, may be needed to meaningfully reduce the occurrence of high-risk flares and subsequent organ damage.

Flare occurrence in our SLE population was higher than that reported in other longitudinal studies, which estimated IRs of ~0.25 to 0.30 flares per patient-year [23−25]. Direct comparisons across studies, however, are challenging due to differences in patient populations, baseline disease activity, use of treatments, flare definitions, and duration of follow-up. The higher flare frequency observed in our cohort likely reflects the inclusion of patients with active disease at baseline.

Although severe SFI flares exerted a greater prognostic impact, mild/moderate flares also contributed to damage. Consistently, in the Asia-Pacific Lupus Collaboration, mild/moderate SFI flares increased the risk of subsequent damage by 32% and severe flares by 58% [13]. Similar results were reported in the Latin American Group for the Study of Lupus (GLADEL) cohort, in which each flare—irrespective of severity—increased the likelihood of damage accrual by almost 2-fold [14]. Altogether, these data suggest that even mild/moderate flares are clinically relevant contributors to long-term morbidity in SLE. Part of this burden might reflect less intensive treatment approaches for mild/moderate flare episodes in earlier therapeutic eras, compared with contemporary strategies favouring earlier immunosuppressive/biologic escalation [3].

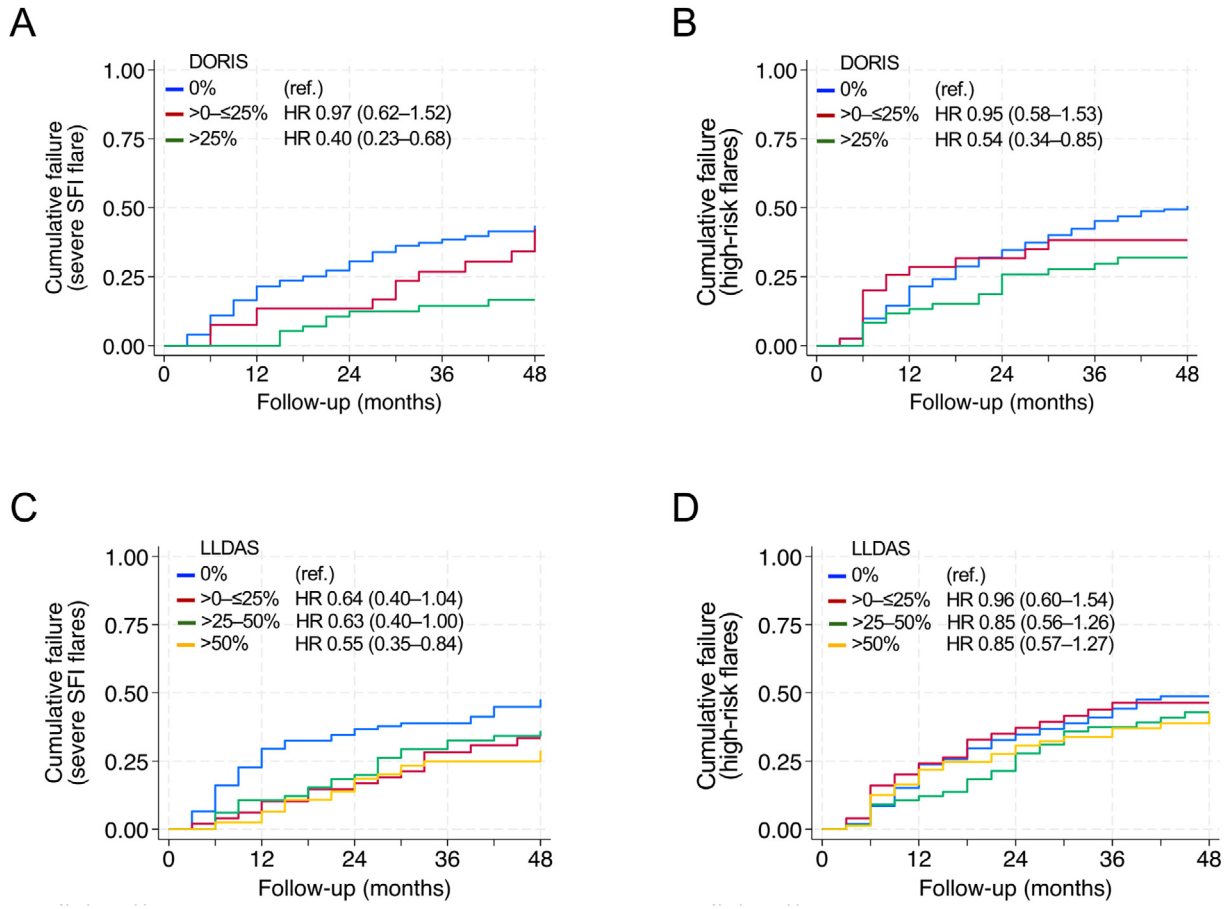


Figure 4. Treatment target attainment and risk of subsequent flares. Landmark-based analysis showing the cumulative incidence of flares occurring after the first 12 months of follow-up, stratified by the cumulative proportion of time spent in DORIS remission during the first 12-month period divided into 3 groups (0%[reference], > 0 to ≤ 5%, and > 25%). Analysis was performed in the derivation cohort (n = 354 patients with SLE). Curves show time to first severe SFI flare (A) and time to first high-risk flare (B). Corresponding analyses stratified by the cumulative proportion of time spent in LLDAS (divided into 4 groups; 0%[reference], > 0 to ≤ 25%, > 25% to 50%, and > 50%) during the first 12 months are shown for severe SFI flares (C) and high-risk flares (D). Hazard ratios (HR) were estimated using Cox models with predefined exposure groups and are shown on the graphs (A-D). DORIS, Definition of Remission in systemic lupus erythematosus; LLDAS, Lupus Low Disease Activity State; SFI, Safety of Estrogens in Lupus Erythematosus National Assessment-systemic lupus erythematosus Disease Activity Index Flare Index; SLE, systemic lupus erythematosus.

Our SLE flare score is driven primarily by treatment-related features, with GC escalation and intensification of ongoing therapy emerging as most influential. These elements reflect clinically meaningful changes in disease status that may not be explicitly captured by standard activity indices. The lower data-driven weight of standalone higher-dose GC escalation relative to mild/moderate oral escalation may reflect differences in GC tapering dynamics and disease control between the 2 dosing strategies, although this requires further investigation.

The SFI was introduced nearly 2 decades ago, when escalation to immunosuppressive agents (and later biologics) was reserved for more severe disease, such that treatment intensification itself partly encoded flare severity. Current recommendations advocate earlier introduction of these therapies both for flare prevention and as GC-sparing strategies, including in patients who might historically have fulfilled only the mild/moderate SFI definition. Consequently, an episode classified as mild/moderate SFI flare may reflect either limited disease activity or a clinically meaningful relapse managed without therapeutic escalation. Importantly, the weights in our score were data-derived, without any *a priori* assumption regarding the appropriateness of specific treatment strategies. Accordingly, within the score, intensification of immunosuppressive/biologic therapy should be interpreted primarily as a clinician-derived marker of perceived flare severity rather than as an independent driver of damage accrual; therefore, recommendations-concordant treatment escalation should not be viewed as being 'penalised' by the scoring system.

The SFI, British Isles Lupus Assessment Group (BILAG)-2004, and more recently the SLE-DAS flare tool each capture different dimensions of flare activity. However, the relative complexity of instruments such as SFI and BILAG-2004 may hinder widespread implementation in routine practice [26]. Previous deconstruction of the SFI suggested that medication-only criteria may misclassify flare severity when considered in isolation [5]. Consistently, we demonstrated that treatment intensification acquires prognostic value especially when it occurs in conjunction with clinical worsening. Accordingly, combinations of clinical manifestations and treatment escalation frequently exceeded the predefined threshold of ≥ 3.5, thereby designating flares linked to subsequent damage.

This study has several limitations. Recruitment from tertiary centres likely enriched the cohort with more severe and treatment-refractory cases. The cohort was almost exclusively of white patients with a modest representation of severe renal involvement (15.8% at inclusion), consistent with its Southern European composition; both features may limit generalizability. Our study spanned a prolonged calendar period, during which

treatment strategies evolved, particularly with respect to GC dosing and the increasing, albeit still limited, use of biologics with potential disease-modifying effects [27]. BILAG-2004 was not systematically scored, precluding direct comparison with this instrument. For SLE-DAS, head-to-head analyses were performed using a modified version constructed from the available variables. This modification may have reduced sensitivity to longitudinal changes in disease activity, and the possibility that the original SLE-DAS captures damage-relevant signal beyond that detected by the SFI cannot be entirely excluded. The SFI architecture also inherently couples clinical features with treatment decisions, which may vary and lack standardisation across clinicians and centres. As with all observational cohorts relying on scheduled visits, self-limiting mild/moderate flares may be incompletely captured; such underascertainment would, however, bias the corresponding association towards the null, whereas a significant association with damage was nonetheless observed. Laboratory parameters obtained within 30 days of clinical assessment may have introduced temporal discordance between serological and clinical activity. Finally, the retrospective design may be subject to residual confounding.

In summary, we developed a clinician-friendly flare score that provides an outcome-based refinement of SLE flares—beyond the classic mild/moderate vs severe dichotomy—and improves prognostic stratification. High-risk flares (score ≥ 3.5) may arise either from major organ involvement requiring high-dose GCs or from combinations of milder organ disease with treatment escalation, including GC dose increases or immunosuppressive/biologic initiation. These high-score episodes were robustly linked to subsequent damage accrual, and the prognostic relevance of even a single high-risk flare underscores the importance of achieving durable disease control. These findings support further evaluation of this scoring approach and suggest potential utility for clinical monitoring and trial design.

Competing interests

ES received consulting fees from AstraZeneca out of the present work. MG has received fees for sponsored lectures from GSK and AstraZeneca. AB reports consulting fees from GSK. GB reports grants from AstraZeneca, MSD, honoraria and/or consulting fees from GSK, Lilly, Novartis, AstraZeneca, and Otsuka. The remaining authors declare no conflict of interest.

Acknowledgements

We are thankful to the staff physicians and nurses of the Rheumatology Departments of the University Hospital of Heraklion and the University of Ferrara for providing care to the patients with systemic lupus erythematosus.

Contributors

DN and GB conceived and designed the study, interpreted the results, and drafted the manuscript. GB performed statistical analysis. PG and SK performed statistical analysis and drafted parts of the manuscript. SP, DNM, and MN collected data and performed data entry and data curation. ES, AR, AM, CB, AR, and AB evaluated patients and completed data collection forms. MG, PS, AF, DB, and GS contributed to results interpretation and evaluated patients. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication. GB supervised the study and acts as guarantor for the overall content of the manuscript.

Funding

The study received funding from the Research Account of the University of Crete (KA10210) and the Pancretan Health Association. DN has received grants from the Swedish Rheumatism Association (R-995557 and R-1028196), King Gustaf V's 80-Year Foundation (FAI-2023-1006 and FAI-2025-1632), and the Ulla and Roland Gustafsson Foundation (2024-49 and 2025-54).

Patient consent for publication

Patients' informed consent was obtained for data collection, analysis, and publication of results for both datasets used in this study.

Ethics approval

This study involves human participants, and data collection and analysis were approved by the ethics committees of the participating centres (protocol no. 13960/10-10-2018 and 516/2019/Oss/AOUFe). All patients gave informed consent upon inclusion in the respective registries.

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.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data availability statement

All data collected for the study are available upon request.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ard.2026.06.013.

Orcid

Dionysis Nikolopoulos: http://orcid.org/0000-0002-9894-6966

George Bertsias: http://orcid.org/0000-0001-5299-1406

REFERENCES

[1] Fanouriakis A, Tziolos N, Bertsias G, Boumpas DT. Update omicronn the diagnosis and management of systemic lupus erythematosus. Ann Rheum Dis. 2021 Jan;80(1):14–25.

[2] Adamichou C, Bertsias G. Flares in systemic lupus erythematosus: diagnosis, risk factors and preventive strategies. Mediterr J Rheumatol. 2017 Mar;28(1):4–12.

[3] Fanouriakis A, Kostopoulou M, Andersen J, Aringer M, Arnaud L, Bae SC, et al. EULAR recommendations for the management of systemic lupus erythematosus: 2023 update. Ann Rheum Dis. 2024 Jan;83(1):15–29.

[4] Petri M, Buyon J, Kim M. Classification and definition of major flares in SLE clinical trials. Lupus. 1999;8(8):685–91.

[5] Thanou A, Chakravarty E, James JA, Merrill JT. 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. Rheumatol (Oxf Engl). 2014 Dec;53(12):2175–81.

[6] Buyon JP, Petri MA, Kim MY, Kalunian KC, Grossman J, Hahn BH, 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 Jun;142(12 Pt 1):953–62.
[7] Garantziotis P, Nikolopoulos D, Katechis S, Temiz A, Nothling DM, Adamichou C, et al. Systemic lupus erythematosus damage risk index (SLE-DRI): a simple machine learning-based tool for identifying patients at risk for early organ damage. RMD Open 2025 Sep;11(3):e006009.
[8] Petri M, Orbai AM, Alarcon GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum. 2012 Aug;64(8):2677–86.
[9] Aringer M, Costenbader K, Daikh D, Brinks R, Mosca M, Ramsey-Goldman R, et al. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. Arthritis Rheumatol. 2019 Sep;71(9):1400–12.
[10] Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. J Rheumatol 2002 Feb;29(2):288–91.
[11] Piga M, Chessa E, Morand EF, Ugarte-Gil MF, Tektonidou M, van Vollenhoven R, et al. Physician Global Assessment International Standardisation COnsensus in Systemic Lupus Erythematosus: the PISCOS study. Lancet Rheumatol. 2022 Jun;4(6):e441–9.
[12] Petri M, Kim MY, Kalunian KC, Grossman J, Hahn BH, Sammaritano LR, et al. Combined oral contraceptives in women with systemic lupus erythematosus. N Engl J Med. 2005 Dec;353(24):2550–8.
[13] Kandane-Rathnayake R, Milea D, Louthrenoo W, Hoi A, Golder V, Cho J, et al. Longitudinal associations of flare and damage accrual in patients with systemic lupus erythematosus. Lupus Sci Med. 2025 Jan;12(1):e001363.
[14] Ugarte-Gil MF, Acevedo-Vasquez E, Alarcon GS, Pastor-Asurza CA, Alfaro-Lozano JL, Cucho-Venegas JM, et al. The number of flares patients experience impacts on damage accrual in systemic lupus erythematosus: data from a multiethnic Latin American cohort. Ann Rheum Dis. 2015 Jun;74(6):1019–23.
[15] Gladman D, Ginzler E, Goldsmith C, Fortin P, Liang M, Urowitz M, et al. The development and initial validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology damage index for systemic lupus erythematosus. Arthritis Rheum. 1996 Mar;39(3):363–9.
[16] Correction: 2021 DORIS definition of remission in SLE: final recommendations from an international task force. Lupus Sci Med. 2022 Feb;9(1):e000538corr1.
[17] Golder V, Kandane-Rathnayake R, Huq M, Nim HT, Louthrenoo W, Luo SF, et al. Lupus low disease activity state as a treatment endpoint for systemic lupus erythematosus: a prospective validation study. Lancet Rheumatol. 2019;1(2):e95–102.
[18] Katechis S, Pitsigavdaki S, Nikoloudaki M, Silvagni E, Repa A, Marangoni A, et al. Combination of clinical factors predicts successful glucocorticoid withdrawal in systemic lupus erythematosus (SLE): results from a multicentre, retrospective cohort study. RMD Open. 2025 Jan 6;11(1):e005118.
[19] Pitsigavdaki S, Nikoloudaki M, Garantziotis P, Silvagni E, Repa A, Marangoni A, et al. Pragmatic targets for moderate/severe SLE and their implications for clinical care and trial design: sustained DORIS or LLDAS for at least 6 months is sufficient while their attainment for at least 24 months ensures high specificity for damage-free progression. Ann Rheum Dis. 2024 Mar;83(4):464–74.
[20] Jesus D, Matos A, Henriques C, Zen M, Larosa M, Iaccarino L, 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 Mar;78(3):365–71.
[21] Saraiva L, Cunha RN, Jesus D, Gatto M, Zen M, Iaccarino L, et al. The SLE-DAS provides an accurate and feasible flare tool in the clinical setting: a validation study. Rheumatol (Oxf Engl). 2024 Apr;63(4):1123–9.
[22] Gladman DD, Urowitz MB, Rahman P, Ibañez D, Tam LS. Accrual of organ damage over time in patients with systemic lupus erythematosus. J Rheumatol. 2003 Sep;30(9):1955–9.
[23] Zen M, Iaccarino L, Gatto M, Bettio S, Saccon F, Ghirardello A, et al. The effect of different durations of remission on damage accrual: results from a prospective monocentric cohort of Caucasian patients. Ann Rheum Dis. 2017 Mar;76(3):562–5.
[24] Kwon OC, Park MC. Risk of systemic lupus erythematosus flares according to autoantibody positivity at the time of diagnosis. Sci Rep. 2023 Feb;13(1):3068.
[25] Hao Y, Ji L, Gao D, Fan Y, Geng Y, Zhang X, et al. Flare rates and factors determining flare occurrence in patients with systemic lupus erythematosus who achieved low disease activity or remission: results from a prospective cohort study. Lupus Sci Med. 2022 Mar;9(1):e000553.
[26] Isenberg D, Sturgess J, Allen E, Aranow C, Askanase A, B Sang-Cheol, et al. Study of Flare Assessment in Systemic Lupus Erythematosus Based on Paper Patients. Arthritis Care Res (Hoboken). 2018 Jan;70(1):98–103.
[27] Askanase AD, Furie RA, Dall'Era M, Bomback AS, Schwarting A, Zhao MH, et al. Disease-modifying therapies in systemic lupus erythematosus for extrarenal manifestations. Lupus Sci Med. 2024 May;11(1):e001124.