



Sensitivity analyses of four systemic lupus erythematosus disease activity indices in predicting the treatment changes in consecutive visits: a longitudinal study

Alimohammad Fatemi¹ · Ahmad Raeisi² · Zahra Sayedbonakdar¹ · Abbas Smiley³

Received: 26 April 2017 / Revised: 9 November 2017 / Accepted: 4 December 2017 / Published online: 18 December 2017
© International League of Associations for Rheumatology (ILAR) 2017

Abstract

This study was conducted to assess the ability of the British Isles Lupus Assessment Group-2004 (BILAG-2004), the SLE Disease Activity Index-2K (SLEDAI-2K), the European Consensus Lupus Activity Measurement (ECLAM), and the Revised Systemic Lupus Activity Measure (SLAM-R) to detect the need to treatment change in daily clinical practice. One hundred and two patients with SLE were enrolled and followed up for 2 to 8 months and visited at least 3 times. Physician Global Assessment, BILAG-2004, SLEDAI-2K, SLAM-R, and ECLAM, were calculated in every visit. Treatment change, dependent variable, was categorized as decrease/no change vs. increase. The aforementioned indices, independent variables, were compared to learn their ability in predicting the treatment change. The probability of treatment change was measured by generalized linear-mixed effect model (GLMM) and generalized estimating equations (GEE). Adjusted odds ratios were calculated. Predictive power of indices was compared by area under the curve (AUC) in plots of sensitivity vs. 1-specificity and application of receiver operating characteristic curves (ROC). BILAG-2004 and SLEDAI-2K had substantial correlation with treatment change. Among different GLMM models, BILAG-2004 followed by SLEDAI-2K showed the highest associations with treatment change. Among various GEE models, similar findings were observed. Also, these 2 indices had the highest sensitivity (the largest AUC) towards treatment change; BILAG-2004 (AUC = 0.779, 95% CI = 0.710–0.848, $p = 0.001$) and SLEDAI-2K (AUC = 0.771, 95% CI = 0.698–0.843, $p = 0.001$). BILAG-2004 followed by SLEDAI-2K had the highest predictability of treatment change.

Keywords BILAG-2004 · ECLAM · SLAM-R · SLEDAI · Systemic lupus erythematosus

Introduction

Systemic lupus erythematosus (SLE) is an autoimmune, multisystem disease with protean clinical and immunological manifestations. Moreover, even in a single patient, it has a very variable course with periods of flares and remissions [1, 2]. Therefore, assessment of disease activity is a great challenge for both physicians and researchers.

Measurement of disease activity has a central role in proper and prompt management of SLE in clinical daily practice. Moreover, it is important in observational and randomized clinical trials to evaluate whether the primary endpoint or efficacy of treatment has been reached [3–5]. Although serological markers and acute phase reactants have been used widely to assess lupus activity, their drawbacks and pitfalls should be considered [4]. For instance, anti-double-stranded DNA (anti-ds DNA) and complement levels may be abnormal despite a low disease activity state. Similarly, high erythrocyte sedimentation rate or C-reactive protein does not necessarily mean active SLE. Furthermore, differentiating reversible active disease from irreversible tissue/organ damage is another important issue in assessment of disease activity [1].

Although more than 60 indices have been developed to measure the disease activity since past mid-century, no assessment tool has been proved to be the gold standard yet [1, 6].

Physician Global Assessment (PGA), or physician's opinion, has also been used to capture disease activity [3, 7, 8].

✉ Alimohammad Fatemi
a_fatemi@med.mui.ac.ir

¹ Department of Rheumatology, Alzahra Hospital, Isfahan University of Medical Sciences, Isfahan, Iran

² Department of Internal Medicine, Alzahra Hospital, Isfahan University of Medical Sciences, Isfahan, Iran

³ Department of Epidemiology and Biostatistics, School of Public Health, Indiana University, Bloomington, IN, USA

However, it relies solely on physician's judgment. In addition, laboratory results could confound its determination [9].

Among several disease activity indices, some considered global disease activity such as the SLE Disease Activity Index-2K (SLEDAI-2K), the European Consensus Lupus Activity Measurement (ECLAM), and the Revised Systemic Lupus Activity Measure (SLAM-R) [1]. They are useful to compare the patients with widely different clinical presentations, but they may overlook the high activity in a single organ in the face of overall measured mild disease activity. Conversely, some other indices such as the British Isles Lupus Assessment Group-2004 (BILAG-2004) index point out individual organ activity [10]. Therefore, they may detect improvement or deterioration more precisely than the former group, and consequently, they may be more conclusive towards the change of treatment.

The validity and reliability of these indices have been studied before [11–14]. However, previous studies have demonstrated inconsistent results about their sensitivity to detect disease activity. For instance, Yee et al. showed more sensitivity of BILAG than SLEDAI-2K [15], whereas Gladman et al. found the vice versa [16]. Likewise, although some studies showed the more sensitivity of SLAM-R than SLEDAI to record the clinical changes [8, 17], some others demonstrated the opposite [16]. Moreover, most previous studies have compared only two/three indices or with low sample size. We aimed to longitudinally compare SLEDAI-2K, BILAG-2004, SLAM-R, and ECLAM on detection of changes in disease activity and to understand their predictability on treatment alteration in routine clinical practice.

Patients and methods

Study design

In a longitudinal cohort study conducted from August 2014 to April 2015, 102 patients who fulfilled at least four of the American College of Rheumatology classification criteria [18] of SLE were recruited. The patients were referred from the Lupus Clinic affiliated to the Isfahan University of Medical Sciences. The regional Ethical Committee approved the study design. Informed consent was obtained from all patients.

Data collection

Clinical and laboratory parameters as well as prescribed medications were recorded during each visit. The interval between the two consecutive assessments was not determined in advance. Rather, patients were visited in routine clinical practice "as needed" according to their health status in order to either modify the dosage of medications, prescribe stronger/lighter medications, or follow the patients more closely. Activity

indices including SLEDAI-2K, BILAG-2004, ECLAM, and SLAM-R were calculated during each visit.

We selected measurement tools used most frequently in clinical settings and research investigations with the ability of disease categorization. They were different in measurement format and definition of disease severity. Overlooking the change of medications used by patients was the common drawback of all tools. A rheumatologist (the corresponding author) conducted the interview and physical examination during each visit, and recorded his opinion about disease activity according to PGA. The PGA scales were as follow: 0 (no active disease), 1 (mildly active disease), 2 (moderately active disease), and 3 (severely active disease) [19].

BILAG-2004 included nine categories of symptoms and laboratory evaluations as constitutional, musculoskeletal, mucocutaneous, cardiorespiratory, neuropsychiatric, gastrointestinal, renal, ophthalmic, and hematology [11]. Each item was scored based on presence, absence, recurrence, or progress of symptom or laboratory parameter in the last 4 weeks in relation to the previous 4 weeks. BILAG-2004 grades would be nominal scaling of alphabetical orders A, B, C, D, and E which stand for "requiring urgent disease-modifying therapy," "demanding close attention, often symptomatic therapy," "stable, controlled on current therapy," and "no involvement," inactive disease but previously affected, respectively. Active disease was defined if any previous score changed to A score or change of C, D, or E score to B score. Improvement (decreased disease activity) was defined as disappearance of A and B scores in all organs/systems with no new A or B scores [1].

SLEDAI-2K included 16 clinical manifestations and eight laboratory parameters (four urinary and four hematologic/immunologic items) [14]. Presence of any item during the past 10 days was recorded and summed up towards weighted scoring. The final score would be from 0 (no disease activity) and 105 (the most severe disease activity). An increase in SLEDAI > 3 and a decrease in SLEDAI > 3 were considered as increased and decreased disease activity, respectively [1].

ECLAM is also a global index of SLE disease activity in the past month [20]. It scored the disease activity from 0 to 10 using 15-weighted clinical, hematological, and serological findings. The latter included ESR and complement levels.

SLAM-R involved 23 clinical active vs. non-active symptoms and signs plus 7 laboratory parameters [12]. They were weighted based on a Likert scale of 1 (mild) to 3 (severe) and categorized in 10 categories leading to maximum score of 81. The categories included laboratory, cardiovascular, pulmonary, ophthalmic, gastrointestinal, dermatologic, neuromotor, rheumatologic, and reticuloendothelial systems. Similar to BILAG-2004, the immunologic parameters are excluded from SLAM-R. ECLAM and SLAM-R were considered as continuous variables and one unit change was defined as the index change.

Treatment change was classified as decrease/no change vs. increase in treatment. The classification was based on the difference in treatment after patient evaluation and comparing with the treatment before evaluation. Any increase in any of the medications of interest, irrespective of concurrent decrease in other medications, was considered as treatment increase. Any decrease in any of the medications of interest, without concurrent increase in other medications, was considered as treatment decrease. The followings were considered as the medications of interest: glucocorticoids, immunosuppressive and anti-malarial agents, immunoglobulins, and plasmapheresis. Treatment change was used as the standard reference of disease activity. PGA and the four aforementioned indices were compared to the standard reference. Also, treatment change was used as the response variable (dependent variable) to build the models according to indices (see below).

Statistical methods

Data analyses were conducted using SPSS program (SPSS, Chicago, IL). Patients were divided into two groups on each visit as increased treatment and decreased/not-changed treatment. The Spearman correlation was applied to assess the correlation of treatment change and each measurement index. Frequency distributions of increased disease scores in relation to the last visit scores were compared between the two groups of treatment change for each index by chi-square test. The probability of increased treatment vs. decreased/not-changed

treatment was evaluated by longitudinal analysis using generalized linear-mixed effect model (GLMM) and generalized estimating equations (GEE), adjusted for fixed effects of sex, age of disease diagnosis, current age and disease duration, and random effects of subjects. Information criteria to compare the models were based on Akaike’s information criterion (AIC) in GLMM and Quasi-likelihood under Independence model Criteria (QIC) in GEE. Significant models with smaller AIC/QIC values fit better. Although some patients were visited more than three times, we only included the first three visits in the current study to have a balanced matrix which results in optimum model buildings by GEE. Adjusted odds ratios (ORs) and 95% confidence intervals (CIs) were reported. Predictive power of indices was compared through area under the curves (AUC) in plots of sensitivity vs. 1-specificity and application of receiver operating characteristic (ROC) curves. The larger the AUC, the higher the predictability of index. *p* value less than 0.05 was considered significant.

Results

This study comprised 89 women and 13 men. Patients were followed up for 2 to 8 months and they were visited at least 3 times. There was no significant difference between women and men in terms of mean age (35.5 vs. 34.8 years, respectively), average disease duration (95.8 vs. 79.4 months, respectively), average follow-up duration, and average number of visits.

Table 1 Demographics of patients recruited into the study and their baseline scores (*n* = 102)

Patient Characteristics	Mean	Median	Q1–Q3	Min–Max
Age, years	35.6	34.5	26–42	16–82
Age at diagnosis, years	27	25.5	18–32	10–78
Disease duration, months	92.5	98	38–134	1–326
Follow-up duration, months	5	5	4–6	2–8
Baseline scores				
PGA	NA	0	0–1	0–3
BILAG	NA	C	D–B	E–A
SLEDAI	4.6	4	0–8	0–20
SLAM	2.3	2	1–3	0–11
ECLAM	2	2	0–3	0–8
Baseline laboratory values				
WBC	6050	5700	4250–7300	1700–15,200
Hb	12.4	12.1	11.8–13.1	6.4–16.9
Platelets	216,000	211,000	177,000–264,000	64,000–405,000
ESR	22	16	11–29	0–76
GFR	85	85	68–99	11–165
Proteinuria (mg/day)	462	117	78–349	25–3623

Q1 quartile 25, *Q3* quartile 75, *PGA* Physician Global Assessment, *NA* not applicable, *BILAG* British Isles Lupus Assessment Group, *SLEDAI* Systemic Lupus Erythematosus Disease Activity Index, *SLAM* Systemic Lupus Activity Measure, *ECLAM* European Consensus Lupus Activity Measurement, *WBC* white blood cell, *Hb* hemoglobin, *ESR* erythrocyte sedimentation rate, *GFR* glomerular filtration rate

Table 2 Correlation among the indices and treatment change. Significant correlations at the 0.01 level (two-tailed) are characterized by an asterisk

	PGA	BILAG	SLEDAI	SLAM	ECLAM	Treatment change
PGA	1	0.62*	0.68*	0.47*	0.58*	0.46*
BILAG	0.62*	1	0.78*	0.71*	0.59*	0.35*
SLEDAI	0.68*	0.78*	1	0.58*	0.70*	0.33
SLAM	0.47*	0.71*	0.58*	1	0.62*	0.26
ECLAM	0.58*	0.59*	0.70*	0.62*	1	0.29*
Treatment change	0.46*	0.35*	0.33	0.26	0.29*	1

PGA Physician Global Assessment, BILAG British Isles Lupus Assessment Group, SLEDAI Systemic Lupus Erythematosus Disease Activity Index, SLAM Systemic Lupus Activity Measure, ECLAM European Consensus Lupus Activity Measurement

Other demographic characteristics of patients are presented in Table 1. High anti-ds DNA and low complement levels were seen in 57 and 30.5% of patients at baseline, respectively. Additional laboratory indices are accessible in Table 1. In the first visit, renal involvement was the most frequent organ treated (72%) followed by skin (15%), joints (6%), blood (6%), and central nervous system affection (1%). The same pattern with minimal changes in percentage was observed in the second and third visits. Hydroxychloroquine was the most frequent prescribed medication (84%) followed by prednisolone (83%), azathioprine (25%), mycophenolate mofetil (21%), cyclophosphamide (14%), tacrolimus (12.5%), methotrexate (6%), and cyclosporine (3%). The mean (SD) dosage of prescribed prednisolone in the first, second, and third visits were 8.8 (12), 7.2 (7.8), and 6.1 (6.5) mg/day, respectively. The correlations between treatment change and all indices are shown in Table 2. BILAG-2004 and SLEDAI-2K substantially correlated with treatment change; whereas, SLAM-R and ECLAM demonstrated moderate correlation with treatment change.

The distribution of the change of indices on each visit in relation to the previous visit according to treatment change is presented in Table 3. Concordant cases of increased treatment/increased index are shown in *italic*. Similarly, concordant cases of decreased/not-changed treatment and decreased/not-changed index are in *italic*. In the second visit, 15 patients had increased treatment. An ideal index might be able to show concurrent increased disease activity index in relation to the first visit in those 15 patients. The concurrent increased index/increased treatment (concordant cases) was from 40% (6/15) in SLEDAI-2K to 73% (11/15) by SLAM-R. In the third visit, the treatment was increased for 13 patients. This treatment change was again detected differently by the indices; from 38% (5/13) by SLAM-R to 85% (11/13) by BILAG-2004. Similarly, 87 patients had decreased/not-changed treatment in the second visit, of which, the concurrent decreased/not-changed index cases (concordant ones) were from 73% (64/87) by SLAM-R to 84% (73/87) by BILAG-2004. Also, 89 patients had decreased/not-changed treatment in the third visit,

Table 3 Distribution of treatment increased vs. treatment decreased/not changed according to each index. The significant distributions are marked by an asterisk

Visits	Index	Treatment increased, <i>n</i> (%)		Treatment decreased/not-changed, <i>n</i> (%)	
		Index increased	Index not increased	Index increased	Index decreased not changed
2nd (<i>n</i> = 102)	PGA*	<i>8 (7.8)</i>	7 (6.9)	7 (6.9)	<i>80 (78.4)</i>
	BILAG*	<i>9 (8.8)</i>	6 (5.9)	14 (13.7)	<i>73 (71.5)</i>
	SLEDAI	<i>6 (5.9)</i>	9 (8.8)	16 (15.7)	<i>71 (69.6)</i>
	SLAM*	<i>11 (10.7)</i>	4 (3.9)	23 (22.5)	<i>64 (62.7)</i>
	ECLAM	<i>7 (6.9)</i>	8 (7.8)	22 (21.6)	<i>65 (63.7)</i>
3rd (<i>n</i> = 102)	PGA*	<i>10 (9.8)</i>	3 (2.9)	5 (4.9)	<i>84 (82.3)</i>
	BILAG*	<i>11 (10.7)</i>	2 (1.9)	13 (12.7)	<i>76 (74.5)</i>
	SLEDAI*	<i>7 (6.9)</i>	6 (5.9)	10 (9.8)	<i>79 (77.4)</i>
	SLAM	<i>5 (4.9)</i>	8 (7.8)	20 (19.6)	<i>69 (67.6)</i>
	ECLAM*	<i>7 (6.9)</i>	6 (5.9)	15 (14.7)	<i>74 (72.5)</i>

PGA Physician Global Assessment, BILAG British Isles Lupus Assessment Group, SLEDAI Systemic Lupus Erythematosus Disease Activity Index, SLAM Systemic Lupus Activity Measure, ECLAM European Consensus Lupus Activity Measurement

*The 2×2 chi-square test of the index with treatment change was significant

Table 4 Longitudinal analyses using GLMM and GEE. Treatment was the dependent variable and each index was the independent variable

Model terms	GLMM					GEE				
	Coefficient	OR	95%CI	P	AIC	Coefficient	OR	95%CI	P	QIC
PGA Model										
PGA = 0	Reference group									
PGA = 1	3.811	45.18	10.7–191	0.0001						
PGA = 2	4.057	57.77	13.1–254	0.0001		1.122	3.072	2.33–4.05	0.0001	
PGA = 3	4.344	77.02	16.1–366	0.0001						
Sex = male	0.523	1.687	0.58–4.92	0.35	162.41	0.821	2.272	0.85–6.05	0.10	209.95
Diagnosis age	–0.465	0.628	0.006–70	0.85		–0.085	0.918	0.28–3.02	0.89	
Current age	0.486	1.626	0.01–182	0.85		0.101	1.106	0.33–3.66	0.87	
Disease duration	–0.039	0.961	0.65–1.42	0.85		–0.007	0.993	0.89–1.10	0.90	
BILAG Model										
BILAG = E	Reference group									
BILAG = D	–19.22	0.001	0.01–NA	0.999						
BILAG = C	0.940	2.560	0.54–12.1	0.232		1.029	2.798	1.88–4.15	0.0001	
BILAG = B	2.096	8.131	1.99–33.2	0.004						
BILAG = A	3.163	23.65	5.28–106	0.0001	188.04					219.35
Sex = male	0.622	1.863	0.65–5.37	0.25		0.591	1.807	0.78–4.18	0.17	
Diagnosis age	–0.783	0.457	0.004–57	0.75		–0.740	0.477	0.15–1.48	0.20	
Current age	0.774	2.168	0.02–271	0.75		0.730	2.076	0.67–6.44	0.20	
Disease duration	–0.066	0.936	0.63–1.40	0.75		–0.063	0.939	0.85–1.03	0.20	
SLEDAI-2K Model										
SLEDAI-2K	0.195	1.216	1.13–1.30	0.001		0.194	1.215	1.13–1.29	0.0001	
Sex = male	0.398	1.489	0.51–4.30	0.46		0.385	1.469	0.70–3.06	0.310	
Diagnosis age	0.900	2.460	0.015–41	0.73	210.91	1.076	2.932	1.08–7.96	0.035	226.05
Current age	–0.910	0.402	0.002–67	0.73		–1.085	0.338	0.12–0.92	0.034	
Disease duration	0.075	1.078	0.70–1.65	0.73		0.090	1.094	1.01–1.19	0.040	
SLAM Model										
SLAM	0.141	1.151	1.03–1.28	0.01		0.324	1.383	1.21–1.58	0.0001	
Sex = male	–0.232	0.790	0.41–1.52	0.48		0.186	1.204	0.45–3.32	0.712	
Diagnosis age	0.429	1.536	0.01–187	0.86	218.49	0.456	1.578	0.51–4.88	0.429	242.15
Current age	–0.447	0.639	0.005–78	0.86		–0.474	0.622	0.20–1.94	0.413	
Disease duration	0.034	1.035	0.69–1.54	0.86		0.036	1.037	0.94–1.14	0.463	
ECLAM Model										
ECLAM	0.365	1.440	1.22–1.69	0.0001		0.365	1.441	1.25–1.65	0.0001	
Sex = male	0.569	1.766	0.58–5.38	0.31		0.558	1.746	0.79–3.86	0.17	
Diagnosis age	0.777	2.175	0.07–697	0.79	226.29	1.031	2.804	0.89–8.80	0.077	238.74
Current age	–0.782	0.457	0.01–147	0.79		–1.036	0.355	0.11–1.12	0.079	
Disease duration	0.063	1.065	0.66–1.72	0.79		0.084	1.088	0.98–1.20	0.093	

GLMM generalized linear-mixed effect model, GEE generalized estimating equations, AIC Akaike’s information criterion, QIC quasi-likelihood under independence model criteria, PGA Physician Global Assessment, BILAG: British Isles Lupus Assessment Group, SLEDAI Systemic Lupus Erythematosus Disease Activity Index, SLAM Systemic Lupus Activity Measure, ECLAM European Consensus Lupus Activity Measurement

of which, the concurrent decreased/not-changed index cases (concordant ones) were from 77.5% (69/89) by SLAM-R to 89% (79/89) by SLEDAI-2K. No index showed the ability to capture all concordant pairs in both second and third visits.

Treatment was considered as the dependent variable and each index was considered as independent variable in GLMM and GEE models of longitudinal analyses (Table 4). The models were adjusted for sex, age of disease diagnosis,

current age, and disease duration. In addition to PGA, BILAG-2004 followed by SLEDAI-2K revealed the smallest AIC among GLMM models and the smallest QIC among the GEE models. Sensitivity analysis according to AUC and ROC are presented in Table 5 and Fig. 1. Aside from PGA, BILAG-2004 followed by SLEDAI-2K demonstrated the largest AUC indicating the highest ability to distinguish patients who need increased treatment among four main indices. The difference

Table 5 Area under curves (AUC) for all indices

Index	AUC	95% Confidence interval	<i>p</i> value
PGA	0.833	0.780–0.886	0.0001
BILAG	0.779	0.710–0.848	0.0001
SLEDAI	0.771	0.698–0.843	0.0001
ECLAM	0.733	0.658–0.809	0.0001
SLAM	0.709	0.631–0.788	0.0001

AUC area under the curves, PGA Physician Global Assessment, BILAG British Isles Lupus Assessment Group, SLEDAI Systemic Lupus Erythematosus Disease Activity Index, SLAM Systemic Lupus Activity Measure, ECLAM European Consensus Lupus Activity Measurement

between BILAG-2004 and SLEDAI-2K was less than 1 %, 77.9 vs. 77.1%, respectively (Table 5).

Discussion

In this study, we evaluated the changes of four SLE disease activity indices in daily clinical practice and their relationships to physician's assessment and treatment change. We tried to show the best index that might be able to predict treatment changes in longitudinal analyses.

All indices substantially correlated with PGA. SLEDAI-2K had the highest rank followed by BILAG-2004, ECLAM, and SLAM-R in decreasing order. This finding, more or less, was consistent with the previous studies [8, 21, 22]. In addition, PGA, BILAG-2004, and SLEDAI-2K showed substantial correlation with treatment change. BILAG-2004 seems to be the only index (A, B, C, D, E) which covers the treatment decision

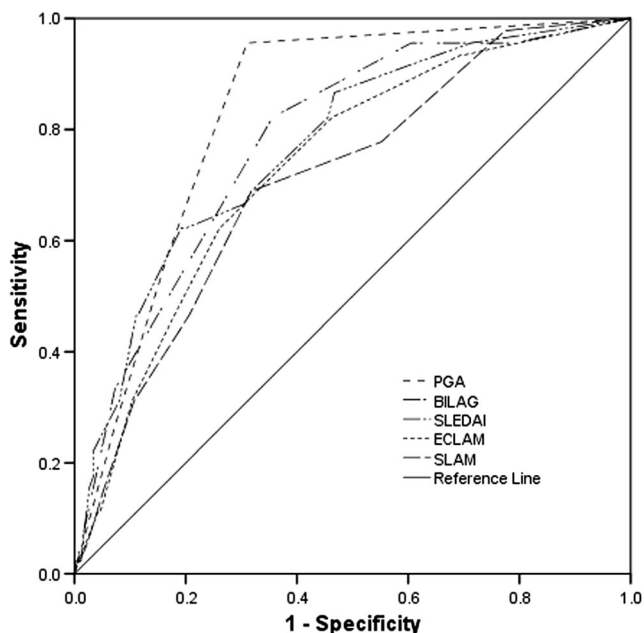


Fig. 1 Sensitivity analyses of all indices according to AUC and ROC

among the indices. It has more items than the other three indices. This may explain its stronger correlation with treatment change compared with the other ones. Also, SLEDAI-2K includes objective items such as immunologic markers, which might justify its substantial correlation with treatment change [1]. On the other hand, some indices such as SLAM-R includes more subjective items like fatigue or arthralgia that might be unrelated to SLE activity, and, consequently, to treatment modification [1].

Table 3 showed an inconclusive pattern of detecting disease activity by five indices requiring treatment change. That's why we needed model building in order to take into account other factors such as age and sex of patients. GLMM models in Table 4 lead us to the probability of increased treatment relative to decreased/not-changed treatment in the current study. For example, the odds of requiring increased treatment instead of decreased/not-changed treatment for patients with BILAG-2004 = A category are estimated to be 23.65 (OR) times the corresponding odds for patients with BILAG-2004 = E category, all other variables being equal. Patients with BILAG-2004 = B are 8 times more likely to need increased treatment relative to those with BILAG-2004 = E. Similarly, patients with BILAG-2004 = C are roughly 2.5 times more likely to need increased treatment relative to patients with BILAG-2004 = E. GEE models in Table 4 lead us to the similar information provided by GLMM but in population level. In other words, GLMM models provide subject-specific results; whereas, GEE models provide average population information.

The AUC measured discrimination. It reflected the ability of index to correctly distinguish between the patients who needed treatment increase vs. those who did not, based on the sensitivity and specificity of each index. An index with perfect discrimination, 100% sensitivity and 100% specificity, has a ROC curve that passes through the upper left corner. Then, the closer the ROC curve to the upper left corner, the higher the overall accuracy of the index. For instance, AUC of 0.779 for BILAG-2004 in Table 5 means, on average, a patient with increased treatment had higher BILAG-2004 score than 77.9% of those with no increased treatment. In other words, BILAG-2004 had the predictive ability to discriminate patients who needed increased treatment than those who did not. This was true for all indices since all AUCs were significant. But, the detectability power was different, from the minimum AUC of 71% for SLAM-R to the maximum AUC of 78% for BILAG-2004. The difference between BILAG and SLEDAI-2K in terms of AUC was minimal. The ability of indices to capture disease activity and the need to increase treatment was different in previous studies. Although BILAG-2004 was more able to detect this issue in the study of Yee et al. [15], there are some reports for better response of SLEDAI [16], SLAM [8, 23], or ECLAM [24]. This inconsistency among studies may result from differences in

statistical analysis, sample size, or study design (for instance, cross-sectional vs. longitudinal study).

This study had some limitations. Our results could be improved if patients' assessment of disease status were considered. In addition, one rheumatologist decided on all treatments, and there was no second check. If the patients were evaluated by a second rheumatologist, less errors were probably expected.

In summary, the current study showed that BILAG-2004 followed by SLEDAI-2K had the highest predictability of the need to increase the treatment in three consecutive visits among the four disease activity indices in our study. It may be expected that the outcome of BILAG-2004 in terms of treatment will be more meaningful.

Compliance with ethical standards The regional Ethical Committee approved the study design. Informed consent was obtained from all patients.

Disclosures None.

References

- Griffiths B, Mosca M, Gordon C (2005) Assessment of patients with systemic lupus erythematosus and the use of lupus disease activity indices. *Best Pract Res Clin Rheumatol* 19(5):685–708. <https://doi.org/10.1016/j.berh.2005.03.010>
- Romero-Diaz J, Isenberg D, Ramsey-Goldman R (2011) Measures of adult systemic lupus erythematosus: updated version of British Isles Lupus Assessment Group (BILAG 2004), European Consensus Lupus Activity Measurements (ECLAM), Systemic Lupus Activity Measure, Revised (SLAM-R), Systemic Lupus Activity Questionnaire for Population Studies (SLAQ), Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K), and Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI). *Arthritis Care Res (Hoboken)* 63(Suppl 11):S37–S46
- Ceccarelli F, Perricone C, Massaro L, Cipriano E, Alessandri C, Spinelli FR, Valesini G, Conti F (2015) Assessment of disease activity in systemic lupus erythematosus: lights and shadows. *Autoimmun Rev* 14(7):601–608. <https://doi.org/10.1016/j.autrev.2015.02.008>
- Rao V, Gordon C (2014) Advances in the assessment of lupus disease activity and damage. *Curr Opin Rheumatol* 26(5):510–519. <https://doi.org/10.1097/BOR.0000000000000085>
- Mikdashi J, Nived O (2015) Measuring disease activity in adults with systemic lupus erythematosus: the challenges of administrative burden and responsiveness to patient concerns in clinical research. *Arthritis Res Ther* 17(1):183. <https://doi.org/10.1186/s13075-015-0702-6>
- Haq I, Isenberg DA (2002) How does one assess and monitor patients with systemic lupus erythematosus in daily clinical practice? *Best Pract Res Clin Rheumatol* 16:181–194
- Petri M, Hellmann D, Hochberg M (1992) Validity and reliability of lupus activity measures in the routine clinic setting. *J Rheumatol* 19(1):53–59
- Ward MM, Marx AS, Barry NN (2000) Comparison of the validity and sensitivity to change of 5 activity indices in systemic lupus erythematosus. *J Rheumatol* 27(3):664–670
- Aranow C (2014) Optimal determination of physician global assessment of lupus disease activity: a pilot study (abstract). *Arthritis Res Ther* 16(Suppl 1):A42
- Yee CS, Farewell V, Isenberg DA, Rahman A, Teh LS, Griffiths B et al (2014) British Isles Lupus Assessment Group 2004 index is valid for assessment of disease activity in systemic lupus erythematosus. *Arthritis Rheum* 56:4113–4119
- Isenberg DA, Rahman A, Allen E, Farewell V, Akil M, Bruce IN, D'Cruz D, Griffiths B, Khamashta M, Maddison P, McHugh N, Snaith M, Teh LS, Yee CS, Zoma A, Gordon C (2005) 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)* 44(7):902–906. <https://doi.org/10.1093/rheumatology/keh624>
- Bae SC, Koh HK, Chang DK, Kim MH, Park JK, Kim SY (2001) Reliability and validity of systemic lupus activity measure-revised (SLAM-R) for measuring clinical disease activity in systemic lupus erythematosus. *Lupus* 10(6):405–409. <https://doi.org/10.1191/096120301678646146>
- Bencivelli W, Vitali C, Isenberg DA, Smolen JS, Snaith ML, Sciuto M, Bombardieri S (1992) Disease activity in systemic lupus erythematosus: report of the Consensus Study Group of the European Workshop for Rheumatology Research. III. Development of a computerised clinical chart and its application to the comparison of different indices of disease activity. The European Consensus Study Group for Disease Activity in SLE. *Clin Exp Rheumatol* 10(5):549–554
- Gladman DD, Ibanez D, Urowitz MB (2002) Systemic lupus erythematosus disease activity index 2000. *J Rheumatol* 29(2):288–291
- Yee CS, Isenberg DA, Prabu A, Sokoll K, Teh LS, Rahman A, Bruce IN, Griffiths B, Akil M, McHugh N, D'Cruz D, Khamashta MA, Maddison P, Zoma A, Gordon C (2008) BILAG-2004 index captures systemic lupus erythematosus disease activity better than SLEDAI-2000. *Ann Rheum Dis* 67(6):873–876. <https://doi.org/10.1136/ard.2007.070847>
- Gladman DD, Goldsmith CH, Urowitz MB, Bacon P, Bombardieri C, Isenberg D, Kalunian K, Liang MH, Maddison P, Nived O (1994) Sensitivity to change of 3 Systemic Lupus Erythematosus Disease Activity Indices: international validation. *J Rheumatol* 21(8):1468–1471
- Chang E, Abrahamowicz M, Ferland D, Fortin PR (2002) Comparison of the responsiveness of lupus disease activity measures to changes in systemic lupus erythematosus activity relevant to patients and physicians. *J Clin Epidemiol* 55(5):488–497. [https://doi.org/10.1016/S0895-4356\(01\)00509-1](https://doi.org/10.1016/S0895-4356(01)00509-1)
- Hochberg MC (1997) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 40:1725
- Petri M, Genovese M, Engle E, Hochberg M (1991) Definition, incidence, and clinical description of flare in systemic lupus erythematosus. A prospective cohort study. *Arthritis Rheum* 34(8):937–944. <https://doi.org/10.1002/art.1780340802>
- Vitali C, Bencivelli W, Isenberg DA, Smolen JS, Snaith ML, Sciuto M, Neri R, Bombardieri S (1992) Disease activity in systemic lupus erythematosus: report of the Consensus Study Group of the European Workshop for Rheumatology Research. II. Identification of the variables indicative of disease activity and their use in the development of an activity score. The European Consensus Study Group for Disease Activity in SLE. *Clin Exp Rheumatol* 10(5):541–547
- Liang MH, Socher SA, Roberts WN, Esdaile JM (1988) Measurement of systemic lupus erythematosus activity in clinical research. *Arthritis Rheum* 31(7):817–825. <https://doi.org/10.1002/art.1780310701>

22. Uribe AG, Vila LM, McGwin G Jr, Sanchez ML, Reveille JD, Alarcon GS (2004) The Systemic Lupus Activity Measure-revised, the Mexican Systemic Lupus Erythematosus Disease Activity Index (SLEDAI), and a modified SLEDAI-2K are adequate instruments to measure disease activity in systemic lupus erythematosus. *J Rheumatol* 31(10):1934–1940
23. Fortin PR, Abrahamowicz M, Clarke AE, Neville C, Du Berger R, Fraenkel L, Liang MH (2000) Do lupus disease activity measures detect clinically important change? *J Rheumatol* 27(6):1421–1428
24. Bombardieri S, Vitali C, Caponi L, Manca L, Bencivelli W (1994) Activity criteria in systemic lupus erythematosus. *Clin Exp Rheumatol* 12(Suppl 11):S45–S48