Comparison of responsiveness of BILAG-2004, SLEDAI-2000 and BILAG Systems Tally (BST)

Yee, Chee-seng; Gordon, Caroline; Isenberg, David A; Griffiths, Bridget; Teh, Lee-suan; Bruce, Ian N; Ahmad, Yasmeen; Rahman, Anisur; Prabu, Athiveeraramapandian; Akil, Mohammed; Mchugh, Neil; Edwards, Christopher J.; D'cruz, David; Khamashta, Munther A; Farewell, Vernon T

DOI:
10.1002/acr.24606

Title: Comparison of responsiveness of BILAG-2004, SLEDAI-2000 and BILAG Systems Tally (BST).

Authors: Dr Chee-Seng Yee PhD FRCP(UK)

Doncaster and Bassetlaw Teaching Hospitals NHS Foundation Trust, Doncaster, UK

Prof Caroline Gordon MD FRCP(UK)

University of Birmingham, Birmingham, UK

Prof David A Isenberg MD FRCP(UK)

University College London, UK

Dr Bridget Griffiths MD FRCP(UK)

Newcastle-upon-Tyne Hospitals NHS Foundation Trust, Newcastle-upon-Tyne, UK

Prof Lee-Suan Teh MD FRCP(UK)

Royal Blackburn Hospital, Blackburn, UK

Prof Ian N Bruce FRCP(UK)

University of Manchester, Manchester, UK

Dr Yasmeen Ahmad PhD MRCP(UK)

Betsi Cadwaladr University Health Board, Wales, UK

Prof Anisur Rahman PhD FRCP(UK)

University College London, London, UK

Dr Athiveeraramapandian Prabu MD FRCP(UK)

University of Birmingham, Birmingham, UK

Dr Mohammed Akil FRCP(UK)

Sheffield Teaching Hospitals NHS Trust, Sheffield, UK

Prof Neil McHugh FRCP(UK)

Royal National Hospital for Rheumatic Diseases NHS Trust, Bath, UK

Prof Christopher J. Edwards MD FRCP(UK)

University Hospital Southampton NHS Foundation Trust, Southampton, UK

Prof David D'Cruz FRCP(UK)

Guy's and St Thomas' NHS Foundation Trust, London, UK

Prof Munther A Khamashta MD FRCP(UK)

Guy's and St Thomas' NHS Foundation Trust, London, UK

Prof Vernon T Farewell PhD

MRC Biostatistics Unit, Cambridge, UK

Corresponding Author: Dr Chee-Seng Yee

Consultant Rheumatologist

Department of Rheumatology

Doncaster Royal Infirmary

Armthorpe Road

Doncaster

DN2 5LT

United Kingdom

E-mail: csyee@ymail.com

Funding

This study was supported by a grant from Versus Arthritis (Grant No. 16081), Medical Research Council (UK) (Grant U105261167) and Vifor Pharma/Aspreva Pharmaceuticals. IB is funded by Versus Arthritis, NIHR Manchester Biomedical Research Unit and NIHR

Manchester Wellcome Trust Clinical Research Facility. The Birmingham SLE clinics are supported by Lupus (UK).

Acknowledgements

We would like to thank all the nurse specialists and doctors who contributed to this study at the participating centres. Some of the work was carried out at the National Institute for Health Research (NIHR)/Wellcome Trust Birmingham Clinical Research Facility. The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR or the Department of Health.

Key words: SLEDAI-2000; BILAG-2004; BST; sBST; SLE; Responsiveness; Disease activity; Longitudinal study

Word count: 3751

Abstract

Objective:

To compare the responsiveness of BILAG-2004 and SLEDAI-2000 disease activity indices and determine if there was any added value in combining BILAG-2004, BILAG System Tally (BST) or simplified BST (sBST) with SLEDAI-2000.

Methods:

This was a multi-centre longitudinal study of SLE patients. Data were collected on BILAG-2004, SLEDAI-2000 and therapy on consecutive assessments in routine practice. The external responsiveness of the indices was assessed by determining the relationship between change in disease activity and change in therapy between two consecutive visits. Comparison of indices and their derivatives was performed by assessing the main effects of the indices using logistic regression. ROC curves analysis was used to describe the performance of these indices individually and in various combinations and comparisons of AUC were performed.

Results:

There were 1414 observations from 347 patients. Both BILAG-2004 and SLEDAI-2000 maintained an independent relationship with change in therapy when compared. There was some improvement in responsiveness when continuous SLEDAI-2000 variables (change in score and score of previous visit) were combined with BILAG-2004 system scores. Dichotomisation of BILAG-2004 or SLEDAI-2000 resulted in poorer performance. BST and sBST had similar responsiveness as the combination of SLEDAI-2000 variables and BILAG-2004 system scores. There was little benefit in combining SLEDAI-2000 with BST or sBST.

Conclusions:

The BILAG-2004 index had comparable responsiveness to SLEDAI-2000. There was some benefit in combining both indices. Dichotomisation of BILAG-2004 and SLEDAI-2000 leads

to suboptimal performance. BST and sBST performed well on their own; sBST is recommended for its simplicity and clinical meaningfulness.

Significance and Innovations

Various ways of analysing the BILAG-2004 index and SLEDAI (and their derivatives) have been employed in longitudinal studies of SLE especially in clinical trials. However, there has not been a direct comparison of these two indices and their various combinations to determine the best way of using them without the addition of a physician's global assessment. The results of this analysis provide guidance on the use of these indices as disease activity outcome measures in longitudinal studies of SLE. The key findings from this analysis are:

1. Both the BILAG-2004 index and SLEDAI-2000 have similar responsiveness and there is some improvement when they are combined.
2. Dichotomisation of BILAG-2004 index and SLEDAI-2000 may reduce performance as an outcome measure.
3. sBST may have an advantage due to its simplicity and clinical meaningfulness.

Systemic lupus erythematosus (SLE) is a complex multi-system disease and assessment of this disease is challenging given the multiple outcome domains to be considered. The two commonly used disease activity indices that allow the results from different cohorts of SLE patients to be compared in clinical trials or observational studies are BILAG-2004 index (BILAG-2004)(1–5) and SLEDAI-2000(6–8).

A strong correlation between Classic BILAG index and original SLEDAI was demonstrated using patient vignettes, but there has been no direct comparison of the performance of BILAG-2004 and SLEDAI-2000 using real-world clinical data(9,10). Various attempts have been made to combine SLEDAI (or its derivatives) with BILAG-2004 or Classic BILAG indices in clinical trials, in the belief that a combination might be superior to either index on its own(11–14). However, there are little data to support this presumption and there are concerns about the impact of variable recording of the physician's global assessment (PhGA) by different physicians(9) in composite responder indices such as SLE Responder Index (SRI) and its derivatives(11,13–16) and BILAG Composite Lupus Assessment (BICLA)(12,17,18). These composite clinical trial end-points focus on changes, specifically patients showing specific levels of improvement in one index at the final trial visit as compared to baseline visit and require no worsening in the alternative index and PhGA. Both SRI and BICLA are currently used as endpoints in clinical trials of SLE but trial results have been inconsistent including some with promising results in Phase 2 studies but negative results in Phase 3 or disappointing results generally(12,15,17,19–21). One of the concerns with trials that failed was with the outcome measure used as the primary endpoint being not optimal(22). This study reports on the analysis comparing BILAG-2004 and SLEDAI-2000, and to determine the best way of using them without PhGA in longitudinal studies.

We have previously demonstrated the external responsiveness of BILAG-2004 and SLEDAI-2000(4,23). Employing similar robust methodology(24), the analyses presented here examine whether the use of both indices improves the responsiveness of each alone using data from a large longitudinal study of SLE patients seen in routine practice. We also compare the performance of BILAG-2004 systems tally (BST). BST is an alternative way of representing BILAG-2004 scores in a longitudinal assessment, that combines the flexibility and simplification of overall numerical scoring of BILAG-2004 with the clinical intuitiveness of BILAG-2004 structure(25).

Patients and Methods

Data from a multi-centre prospective longitudinal study in the United Kingdom, which was primarily designed to validate BILAG-2004, were used in this analysis(4). This same dataset was used to demonstrate the external responsiveness of SLEDAI-2000 and to develop BST and simplified BST (sBST)(23,25). All patients satisfied the revised ACR criteria for classification of SLE(26,27). This study received multi-centre research ethical approval and was carried out in accordance with the Helsinki Declaration. Written consent was obtained from all patients.

This study had been described in detail previously(4). In summary, patients were followed up prospectively in routine clinical practice and data (BILAG-2004, SLEDAI-2000 and treatment) were collected for all consecutive visits and physician encounters. Previously we demonstrated, based on receiver operating characteristic (ROC) curve analyses, that BST, sBST and BILAG-2004 global numerical variables (combination of change in BILAG-2004 global numerical score(5) and the score from the previous visit), were comparably related to change in therapy and provided better discrimination than a model including variables for changes in all nine BILAG-2004 system scores(25). In the analyses presented here, we had included disease activity as assessed by SLEDAI-2000, BILAG-2004 individual system scores and global BILAG-2004 numerical score.

Changes in disease activity and treatment between two consecutive visits were analysed. Each observation for the analysis was derived from two consecutive visits. A robust definition for change in therapy between consecutive visits was used as the external reference for change in disease activity as described previously (see Supplementary Material)(3–5,23,25). Three categories of changes in therapy were defined: ‘no change’, ‘increase in therapy’ and ‘decrease in therapy’.

Statistical Analysis

All statistical analyses were performed using Stata for Windows version 8 (Stata Corporation, Texas) and R(28). Robust variance estimation was used to allow for correlation between multiple assessments from the same patients(29).

External responsiveness was used to compare the performance of the indices in this longitudinal study(24). It assessed the extent to which changes in the index over time relate to corresponding changes in therapy between two consecutive visits. As such, clinically meaningful change was assessed. Change in therapy was chosen as the external reference as there was no better objective alternative and this was used in multiple validation studies on BILAG-2004, SLEDAI-2000 and BST(3–5,23,25). The pros and cons of using change in therapy as the external reference was discussed previously(3).

Maximum-likelihood multinomial and binary logistic regression were used to assess external responsiveness with change in therapy as the outcome variable and changes in disease activity (as determined by the indices) as the explanatory variables. For comparison purposes, the main effects of the indices were assessed within a common regression model. The baseline comparator for change in disease activity used in the analysis was ‘minimal or no change in activity' while the baseline comparator for change in treatment was 'no change in therapy'. The results were reported as odds ratios (OR) with 95% confidence intervals (CI) and Wald tests were used for model comparison where needed.

In the multinomial regression analyses, the baseline category of 'no change in therapy' was compared with both 'increase in therapy' and 'decrease in therapy' respectively. There was no direct comparison between 'increase in therapy' and 'decrease in therapy'. An OR value greater than 1 for one unit increase in the variable defined by the index of interest, within the comparison between 'increase in therapy' and 'no change in therapy', indicated that the increase in index score was associated with 'increase in therapy'. On the other hand, an OR of less than 1 for the same comparison implied that the increase in index score was associated with 'no change in therapy' (and not with 'decrease in therapy') or equivalently an inverse association with 'increase in therapy'. Similar interpretation was applicable to the reported OR for the comparison between 'decrease in therapy' and 'no change in therapy'.

Various combinations of SLEDAI-2000 and BILAG-2004 (including BST and sBST) as dichotomised or regarded as continuous variables were examined and compared to determine if there was added value in combining both of these indices. For some analyses, BILAG-2004 global numerical score was calculated based on the system scores using A=12, B=8, C=1, D/E=0 coding scheme(5). ROC curves analysis was used to describe the performance of these indices and the various combinations(30). Logistic regression was used to estimate the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and area under the curve (AUC). The analyses were performed from two perspectives: deterioration in scores as predictor of 'increase in therapy' and improvement in scores as predictor of 'decrease in therapy'. Calculation of an asymptotic confidence interval for AUC and comparison of AUCs were performed using a non-parametric approach(31).

AUC, with a value from 0 to 1, quantified the performance of the index with the value of 1 corresponding to the index providing perfect discrimination.

Definition of BILAG-2004 deterioration and improvement

Deterioration in BILAG-2004 was defined to have occurred if there was worsening in the score to Grades A or B in any of the systems. The deteriorations were classified as (in order of ranking):

1) *Major deterioration*: when there was worsening from Grade C/D/E to A or from Grade D/E to B
2) *Minor A deterioration*: when there was worsening from Grade B to A
3) *Minor B deterioration*: when there was worsening from Grade C to B

A change from Grade D/E to C was considered minor and not clinically significant. As such, it is excluded from the definition of deteriorations.

Improvement in BILAG-2004 was deemed to have occurred if there was reduction in the score in any system in the absence of any deterioration in the other systems. The improvements were classified as (in order of ranking):

1) *Major improvement*: when there was improvement from Grade A to C/D or Grade B to D
2) *Minor A improvement*: when there was improvement from Grade A to B
3) *Minor B/C improvement*: when there was improvement from Grade B to C or Grade C to D

These classifications were used to define BILAG-2004 based explanatory variables in regression analyses. The definition and gradation above were based on the principle of intention to treat, that underlay BILAG-2004 scoring, whereby active disease requiring therapy was graded A or B depending on the item, while grade C usually required symptomatic therapy(1). It was accepted that at individual patient and organ level, there may be variation in the severity of the disease items and the need for change in therapy within each grade.

BST and sBST

BST and its simplified version sBST were counts of systems with specified changes in scores between two assessments(25).

BST comprised 6 components:

1. Number of systems with major deterioration (change of Grade B/C/D/E to A or Grade D/E to B)
2. Number of systems with minor deterioration (change of Grade C to B)
3. Number of systems with persistent significant activity (no change from Grade A or B)
4. Number of systems with major improvement (change of Grade A to C/D or Grade B to D)
5. Number of systems with minor improvement (change of Grade A to B or B to C)
6. Number of systems with persistent minimal or no activity (change of Grade C/D/E to C/D/E)

sBST had 3 components:

1. Number of systems with active/worsening disease (systems with major deterioration, minor deterioration and persistent significant activity)
2. Number of systems with improving disease (systems with major improvement and minor improvement).
3. Number of systems with persistent minimal or no activity.

Results

There were 347 SLE patients with 1761 assessments that contributed 1414 observations for this analysis. There was an increase in treatment in 22.7% of observations while 37.3% had therapy decreased, and in 40.0%, there was no change in treatment as previously reported(4). The demographics and distribution of change in disease activity for each system were summarised in Supplementary Tables A and B.

Comparison of BILAG-2004 with SLEDAI-2000

To examine the combined performance of SLEDAI-2000 and BILAG-2004, we undertook multinomial logistic regression analysis of change in therapy using the changes in both BILAG-2004 and SLEDAI-2000 with and without their respective values at the previous visit. We had demonstrated previously that although change in SLEDAI-2000 score was significantly associated with changes in treatment, the strongest relationship was observed in a model that included both the change in SLEDAI-2000 score and the score at the previous visit as continuous variables (hereby referred as SLEDAI-2000 variables)(23).

In the analysis of external responsiveness reported here, changes in the individual system scores of BILAG-2004 and SLEDAI-2000 variables (as a continuous variable) were included as explanatory variables for the outcome variable of change in therapy. Table 1 demonstrated that SLEDAI-2000 variables and individual BILAG-2004 system scores retained independent relationships with change in therapy. Consistent with our earlier work(23), if only the change in SLEDAI-2000 score was included (i.e. SLEDAI-2000 score of previous visit was omitted), change in SLEDAI-2000 score was no longer significantly associated with change in therapy (increase or decrease) while changes in BILAG-2004 system scores maintained their significant association with change in therapy (data not shown).

When we undertook a multinomial logistic regression analysis of change in therapy using change in the numerical score of BILAG-2004 and in SLEDAI-2000 along with their respective values at the previous visit (Supplementary Table C), we observed the expected relationships between the changes in the numerical scores and changes in therapy. Both pairs of variables, the two based on BILAG-2004 and the two based on SLEDAI-2000, added predictive power for increase in therapy (p= 0.02 for addition of SLEDAI-2000 variables to BILAG-2004 numerical score variables and p<0.01 for addition of BILAG-2004 numerical score variables to SLEDAI-2000 variables, Wald test). For decrease in therapy, SLEDAI-2000 variables did not provide additional predictive power (p=0.50, Wald test).

From Table 2, we observed that BST variables were related to changes in therapy in the expected manner and SLEDAI-2000 variables provided additional predictive power for increase in therapy (p=0.007 based on Wald test from separate logistic regression) but not for decrease in therapy (p=0.30, Wald test). Similar results were obtained with sBST (Supplementary Table D).

Comparison of performance of combinations of BILAG-2004 and SLEDAI-2000

Table 3 summarised the results of further analyses using various combinations of information from BILAG-2004 and SLEDAI-2000. It presented the AUC measures based on ROC curves derived from binary regression analyses of both ‘increase in therapy’ and ‘decrease in therapy’ versus ‘no change in therapy’. For completeness, we performed similar analyses of ‘increase in therapy’ versus ‘no increase in therapy’ and ‘decrease in therapy’ versus ‘no decrease in therapy’.

The comparison of AUC measures from this exploratory analysis for 'increase in treatment' versus 'no increase in treatment' was summarised in Supplementary Table E that provided the significance levels for the comparison of the various models. The p values should be regarded as illustrative as no adjustment for multiplicity had been performed. Similar results were obtained for analysis for 'increase in treatment' versus 'no change in treatment' (Supplementary Table F). The analysis showed that there was no evidence that either BILAG-2004 system scores or SLEDAI-2000 variables were more predictive of changes in therapy individually than the other (p=0.89, Wald test). There was some improvement in the performance from the combination of both BILAG-2004 system scores and SLEDAI-2000 variables (p<0.001 for the addition of each to the other, Wald test). BST and sBST had comparable performance (p=0.107, Wald test) and were respectively similar to (p=0.128, Wald test) or slightly worse (p<0.001, Wald test) than BILAG-2004 numerical score variables (change in numerical score and previous visit numerical score). BST, sBST and BILAG-2004 numerical score variables appeared to be more predictive of increase in therapy compared to BILAG-2004 system scores (p<0.001, p=0.002 and p<0.001 respectively, Wald test) and SLEDAI-2000 variables (p<0.001, p=0.013 and p<0.001 respectively, Wald test). Furthermore, BST, sBST and BILAG-2004 numerical score variables were comparable to or slightly better than the combination of BILAG-2004 system scores and SLEDAI-2000 variables (p=0.63, p=0.26 and p=0.03 respectively, Wald test). Finally, the addition of SLEDAI-2000 variables provided little improvement to the performance of BST, sBST or BILAG-2004 numerical score variables (p=0.60, p=0.16 and p=0.22 respectively, Wald test).

Dichotomisation of indices

Dichotomised versions of BILAG-2004 and SLEDAI-2000 have been used for a variety of purposes. In Supplementary Material sections of analysis of deterioration of activity using dichotomised variables, analysis of improvement in activity using dichotomised variables, Table G and Table H, the results for clinically relevant dichotomisations were given for the two indices, separately and in combination. These were based on multinomial regressions with a single binary explanatory variable.

Two particular categorisations of changes in the combination of these measures that were of similar magnitude to those used in the definition of SRI (SLEDAI-2000 decrease ≥ 4 and no BILAG-2004 deterioration)(11) and BICLA (all improvements in BILAG-2004 with no SLEDAI-2000 increase ≥ 1)(12) were included in the table examining improvement in disease activity but without PhGA and the change was between two consecutive visits (not between the start and end of study). The estimated sensitivities and specificities were 1.5% and 98.9% respectively for the SRI-like variable, and, 48.2% and 70.0% respectively for the BICLA-like variable when used to predict decrease in therapy (versus no decrease). The AUC values for these two variables were 0.50 and 0.59 respectively compared with AUCs > 0.65 for BST, sBST and BILAG-2004 numerical variables (Table 3). Other dichotomised variables also did not perform as well as these numerical variables in relation to both decrease in therapy and increase in therapy.

Discussion

This multi-centre observational study directly compared the responsiveness of BILAG-2004 with SLEDAI-2000 in longitudinal fashion and assessed the potential value of combining the two indices using a comprehensive range of approaches. Our analyses showed that there was some non-overlapping relationship with change in therapy when both BILAG-2004 and SLEDAI-2000 were included in the model, confirming that both indices had similar responsiveness. Responsiveness was optimal if both the change in SLEDAI-2000 score and SLEDAI-2000 score of the previous visit were included in the model as continuous variables. The use of only change in SLEDAI-2000 score was associated with inferior performance.

Outcome in clinical trials is determined by three factors: efficacy of intervention, study design and effectiveness of the outcome measure used. Our discussion is focused on properties of outcome measure that would affect its ability to differentiate the efficacy of the different treatment arms. It is beyond the scope of this paper to discuss the other factors.

Many clinical trials in SLE had reported their results using various combinations of SLEDAI-2000 or SELENA-SLEDAI (and its variants) with BILAG-2004 or Classic BILAG index as the primary endpoints(22). In the belimumab phase 3 trials, the SRI was used in which a response was defined as an improvement in SELENA-SLEDAI score of at least 4 points with no new Grade A and no more than 1 new Grade B Classic BILAG system score, and, no deterioration of PhGA(13,14). This combination was selected using the dataset from the Phase 2 trial of belimumab to derive the best separation in efficacy between belimumab and placebo, with the presumption that belimumab was effective(11). Using a similar combination of improvement in SLEDAI-2000 score of at least 4 points with no worsening of the BILAG-2004 system score to Grade A or B, we found that this combination performed poorly when assessed using the reference of change in therapy. This was surprising as we would have expected these 2 indices to exert a greater role than PhGA which is subject to variable reporting due to individual physicians scoring lupus manifestations differently to each other in the absence of a glossary, particularly in patients with more than one system involved(9). The indices used in this study were different to the original SRI (BILAG-2004 instead of Classic BILAG index, SLEDAI-2000 instead of SELENA-SLEDAI and no PhGA). This was very similar to the indices used successfully in the phase 2 trial of ustekinumab(16), but which failed as the primary end-point in phase 3 trials of anifrolumab(15) and Lupuzor(19). These trials used a modification of the SRI in which response was driven by a 4-point reduction in SLEDAI-2000 with no more than one new B grade in BILAG-2004 and no more than 10% worsening of PhGA.

A different combination (BICLA) was used in other clinical trials, in which a response was defined as an improvement in BILAG-2004 system score (in the absence of new Grade A or B score) with no worsening of SLEDAI-2000 score (≥1) and no worsening of PhGA(12,15,17,18). The results of our study, presented in Supplementary Table H, supported the use of this combination of BILAG-2004 and SLEDAI-2000 indices which although not successful in the epratuzumab phase 3 trial(17), was successful in phase 3 trials of anifrolumab as primary (TULIP-2) and secondary end-points (TULIP-1)(15,18).

Currently, the combination of the two indices (BILAG-2004 with SLEDAI-2000) used in clinical trials involves dichotomisation(s) of the outcome variables. Our data suggested that the benefit was minimal when combining these two indices in this specific way and the value of PhGA was debatable(9). Dichotomisation involves using a cut-off to determine if a response is achieved (Yes/No response). However, dichotomisation of variables may result in loss of efficiency as it does not allow for a graded response and a partial response might be considered lack of response if the cut-off is not achieved(32). We demonstrated that dichotomisation of both BILAG-2004 and SLEDAI-2000 resulted in poorer responsiveness in our longitudinal study. With better efficiency and performance of the outcome measure used, fewer patients would be required in a study to demonstrate differences between groups which then facilitates target recruitment and reduce the cost of running the study. It was calculated that in comparison to the use of a continuous outcome, the size of a trial may need to be increased by a factor of 30% if a binary outcome with a uniform distribution was used with a median cut-off, with greater gains for a normal distribution(32). By using BST or sBST, which were based on counts of systems with specified transitions in BILAG-2004 scores, the problem of dichotomisation could be avoided.

Although BILAG-2004 numerical score variables and combination of the SLEDAI-2000 variables with BST or sBST had slightly better performance than BST or sBST alone, BST and sBST performed better than BILAG-2004 system scores and SLEDAI-2000. In addition, there was difficulty with interpretation of the clinical meaningfulness of BILAG-2004 numerical score variables and combination of SLEDAI-2000 variables with BST or sBST. Our analyses supported the use of BST or sBST alone and that there was minimal advantage of combining SLEDAI-2000 with BST or sBST. Consequently, there could be simplification in study methodology by using only one disease activity index (BILAG-2004) which would avoid confusion and reduce errors due to differences in BILAG-2004 and SLEDAI-2000 glossaries.

One limitation of this study which might affect the applicability of the results to clinical trials was the time reference used to define change in disease activity. This study looked at the changes between consecutive visits. In contrast, clinical trials generally compare the disease activity between the beginning and the end of the study (and not between consecutive visits) which might be one year apart. With a longer time interval, it is far more likely for a larger effect to occur. However, comparing the outcome measures at only two time points (the beginning and the end of study) ignores the level of disease activity between these two time points. The use of counts or a continuous variable over the study period (such as flare rate) could overcome this disadvantage. Another limitation was that BST and sBST were developed using this same dataset which might have provided an advantage. Validation of our result with an independent dataset is needed.

In conclusion, both BILAG-2004 and SLEDAI-2000 have similar responsiveness longitudinally. There is some benefit in combining the two indices, but dichotomisation of the indices leads to suboptimal performance. BST and sBST performed well on their own and the addition of SLEDAI-2000 variables only resulted in minimal improvement. There is no significant difference with the responsiveness of BST or sBST. Given that sBST has only 3 components, we would recommend the use of sBST in longitudinal analysis of disease activity for its simplicity and clinical meaningfulness.

References

1. Isenberg DA, Rahman A, Allen E, Farewell V, Akil M, Bruce IN, 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. 2005;44(7):902–6.
2. Yee CS, Farewell V, Isenberg DA, Prabu A, Sokoll K, Teh LS, et al. Revised British isles lupus assessment group 2004 index: A reliable tool for assessment of systemic lupus erythematosus activity. Arthritis Rheum. 2006;54(10):3300–5.
3. Yee CS, Farewell V, Isenberg DA, Rahman A, Teh LS, Griffiths B, et al. British Isles Lupus Assessment Group 2004 index is valid for assessment of disease activity in systemic lupus erythematosus. Arthritis Rheum. 2007;56(12):4113–9.
4. Yee CS, Farewell V, Isenberg DA, Griffiths B, Teh LS, Bruce IN, et al. The BILAG-2004 index is sensitive to change for assessment of SLE disease activity. Rheumatology. 2009;48(6):691–5.
5. Yee CS, Cresswell L, Farewell V, Rahman A, Teh LS, Griffiths B, et al. Numerical scoring for the BILAG-2004 index. Rheumatology. 2010;49(9):1665–9.
6. Gladman DD, Ibañez D, Urowltz MB. Systemic lupus erythematosus disease activity index 2000. J Rheumatol. 2002;29(2):288–91.
7. Touma Z, Urowitz MB, Gladman DD. SLEDAI-2K for a 30-day window. Lupus. 2010;19(1):49–51.
8. Bombardier C, Gladman DD, Urowitz MB, Caron D, Chang CH, Austin A, et al. Derivation of the sledai. A disease activity index for lupus patients. Arthritis Rheum. 1992;35(6):630–40.
9. Wollaston SJ, Farewell VT, Isenberg DA, Gordon C, Merrill JT, Petri MA, et al. Defining response in systemic lupus erythematosus: A study by the Systemic Lupus International Collaborating Clinics group. J Rheumatol. 2004;31(12):2390–4.
10. Liang MH, Fortin P, Schneider M, Abrahamowicz M, Alarcón GS, Bombardieri S, et al. The American College of Rheumatology response criteria for systemic lupus erythematosus clinical trials: Measures of overall disease activity. Arthritis Rheum. 2004;50(11):3418–26.
11. Furie RA, Petri MA, Wallace DJ, Ginzler EM, Merrill JT, Stohl W, et al. Novel evidence-based systemic lupus erythematosus responder index. Arthritis Care Res. 2009;61(9):1143–51.
12. Wallace DJ, Kalunian K, Petri MA, Strand V, Houssiau FA, Pike M, et al. Efficacy and safety of epratuzumab in patients with moderate/severe active systemic lupus erythematosus: Results from EMBLEM, a phase IIb, randomised, double-blind, placebo-controlled, multicentre study. Ann Rheum Dis. 2014;73(1):183–90.
13. Furie R, Petri M, Zamani O, Cervera R, Wallace DJ, Tegzová D, et al. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. Arthritis Rheum. 2011;63(12):3918–30.
14. Navarra S V., Guzmán RM, Gallacher AE, Hall S, Levy RA, Jimenez RE, et al. Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: A randomised, placebo-controlled, phase 3 trial. Lancet. 2011;377(9767):721–31.
15. Furie RA, Morand EF, Bruce IN, Manzi S, Kalunian KC, Vital EM, et al. Type I interferon inhibitor anifrolumab in active systemic lupus erythematosus (TULIP-1): a randomised, controlled, phase 3 trial. Lancet Rheumatol. 2019;1(4):e208–19.
16. van Vollenhoven RF, Hahn BH, Tsokos GC, Wagner CL, Lipsky P, Touma Z, et al.

Efficacy and safety of ustekinumab, an IL-12 and IL-23 inhibitor, in patients with active systemic lupus erythematosus: results of a multicentre, double-blind, phase 2, randomised, controlled study. Lancet. 2018;392(10155):1330–9.

17. Clowse MEB, Wallace DJ, Furie RA, Petri MA, Pike MC, Leszczyński P, et al. Efficacy and Safety of Epratuzumab in Moderately to Severely Active Systemic Lupus Erythematosus: Results From Two Phase III Randomized, Double-Blind, Placebo-Controlled Trials. Arthritis Rheumatol. 2017;69(2):362–75.
18. Morand EF, Furie R, Tanaka Y, Bruce IN, Askanase AD, Richez C, et al. Trial of Anifrolumab in Active Systemic Lupus Erythematosus. N Engl J Med. 2020;382(3):211–21.
19. A 52-Week, Randomized, Double-Blind, Parallel-Group, Placebo-Controlled Study to Evaluate the Efficacy and Safety of a 200-mcg Dose of IPP-201101 Plus Standard of Care in Patients With Systemic Lupus Erythematosus - Study Results - ClinicalTrials.gov [Internet]. NIH U S National Library of Medicine. 2019. Available from: https://clinicaltrials.gov/ct2/show/results/NCT02504645?view=results
20. Furie R, Khamashta M, Merrill JT, Werth VP, Kalunian K, Brohawn P, et al. Anifrolumab, an Anti–Interferon-α Receptor Monoclonal Antibody, in Moderate-to-Severe Systemic Lupus Erythematosus. Arthritis Rheumatol. 2017;69(2):376–86.
21. Zimmer R, Scherbarth HR, Rillo OL, Gomez-Reino JJ, Muller S. Lupuzor/P140 peptide in patients with systemic lupus erythematosus: A randomised, double-blind, placebo-controlled phase IIb clinical trial. Ann Rheum Dis. 2013;72(11):1830–5.
22. Merrill JT. For lupus trials, the answer might depend on the question. Vol. 1, The Lancet Rheumatology. 2019. p. e196–7.
23. Yee CS, Farewell VT, Isenberg DA, Griffiths B, Teh LS, Bruce IN, 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. 2011;50(5):982–8.
24. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: A critical review and recommendations. J Clin Epidemiol. 2000;53(5):459–68.
25. Yee CS, Gordon C, Isenberg DA, Griffiths B, Teh LS, Bruce IN, et al. The BILAG-2004 systems tally-a novel way of representing the BILAG-2004 index scores longitudinally. Rheumatol (United Kingdom). 2012;51(11):2099–105.
26. Tan EM, Cohen AS, Fries JF, Masi AT, Mcshane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1982;25(11):1271–7.
27. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Vol. 40, Arthritis and rheumatism. 1997. p. 1725.
28. Kurt Hornik. The R FAQ [Internet]. Comprehensive R Archive Network. 2020. p. 50. Available from: https://cran.r-project.org/doc/FAQ/R-FAQ.html
29. Williams RL. A note on robust variance estimation for cluster-correlated data. Vol. 56, Biometrics. 2000. p. 645–6.
30. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. Vol. 39, Clinical Chemistry. 1993. p. 561–77.
31. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1988;44(3):837.
32. Farewell VT, Tom BDM, Royston P. The impact of dichotomization on the efficiency

of testing for an interaction effect in exponential family models. J Am Stat Assoc. 2004;99(467):822–31.