RHEUMATOLOGY

Rheumatology 2012;51:2099–2105
doi:10.1093/rheumatology/kes207
Advance Access publication 20 August 2012

Original article

The BILAG-2004 systems tally—a novel way of representing the BILAG-2004 index scores longitudinally

Chee-Seng Yee[1], Caroline Gordon[2], David A. Isenberg[3], Bridget Griffiths[4], Lee-Suan Teh[5], Ian N. Bruce[6], Yasmeen Ahmad[7], Anisur Rahman[3], Athiveeraramapandian Prabu[2], Mohammed Akil[8], Neil McHugh[9], Christopher Edwards[10], David D'Cruz[11], Munther A. Khamashta[11] and Vernon T. Farewell[12]

Abstract

Objective. This was an exploratory analysis to develop a new way of representing BILAG-2004 system scores longitudinally that would be clinically meaningful and easier to analyse in comparison with multiple categorical variables.

Methods. Data from a multicentre longitudinal study of SLE patients (the BILAG-2004 index and therapy collected at every visit) were used. External responsiveness analysis of the index suggested the possibility of using counts of systems with specified transitions in scores as a basis to analyse the system scores. Exploratory analyses with multinomial logistic regression were used to examine the appropriateness of this new method of analysing BILAG-2004 system scores. Receiver operating characteristic (ROC) curve analysis was used to assess the performance of this approach.

Results. There were 1414 observations from 347 patients. A novel method was devised based on counts of systems with defined transitions in score (BILAG-2004 systems tally, BST). It has six components (systems with major deterioration, systems with minor deterioration, systems with persistent significant activity, systems with major improvement, systems with minor improvement and systems with persistent minimal or no activity). This was further simplified (simplified BST, sBST) into three components (systems with active/worsening disease, systems with improving disease and systems with persistent minimal or no activity). Both versions had expected associations with change in therapy. ROC curve analyses demonstrated that both versions had similar good performance characteristics (areas under the curve >0.80) in predicting increase in therapy.

Conclusion. The BST and sBST provide alternative approaches to representing BILAG-2004 disease activity longitudinally. Further validation of their use is required.

Key words: BILAG-2004, SLE, disease activity, longitudinal study, BILAG-2004 systems tally, BST, sBST.

CLINICAL SCIENCE

[1]Department of Rheumatology, Doncaster and Bassetlaw Hospitals NHS Foundation Trust, Doncaster, [2]Rheumatology Research Group, University of Birmingham, Birmingham, [3]Centre for Rheumatology Research, University College London, [4]Department of Rheumatology, Freeman Hospital, Newcastle upon Tyne, [5]Department of Rheumatology, Royal Blackburn Hospital, Blackburn, [6]Arthritis Research UK Epidemiology Unit, Manchester Academic Health Sciences Centre, The University of Manchester, Manchester, [7]Department of Rheumatology, Betsi Cadwaladr University Health Board, Wales, [8]Department of Rheumatology, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, [9]Department of Rheumatology, Royal National Hospital for Rheumatic Diseases NHS Foundation Trust, Bath, [10]Department of Rheumatology, University of Southampton, Southampton, [11]Lupus Unit, St Thomas' Hospital, London and [12]MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.

Submitted 8 February 2012; revised version accepted 26 June 2012.

Correspondence to: Chee-Seng Yee, Department of Rheumatology, Doncaster Royal Infirmary, Armthorpe Road, Doncaster DN2 5LT, UK.
E-mail: csyee@ymail.com

Introduction

The BILAG-2004 index is a system-based disease activity index that has been validated for SLE and is one of the recommended disease activity outcome measures used in clinical trials of SLE [1–4]. This index provides disease activity scores across nine systems using an ordinal scale score (from A to E). Although this is clinically intuitive, the analysis of nine separate system scores poses a great challenge in clinical studies. The development of a global numerical scoring for the BILAG-2004 index offers one approach to overcome the difficulty with analysis, but it is hampered by the difficulty in interpretation of its clinically meaningful change [5].

Currently, most analyses of the classic BILAG or BILAG-2004 index scores are based on categorization of the outcome variable, especially dichotomization into a binary (yes/no) response. In the rituximab EXPLORER trial, the outcome end points were major clinical response, partial clinical response or no response [6]. Furthermore, the criteria for the clinical responses were based on dichotomization of whether certain classic BILAG scores were achieved and maintained. Similarly the combinations of indices used in the belimumab and epratuzumab trials are also based on dichotomization of responses [7, 8]. However, dichotomization of variables does not allow the gradation of response according to severity to be taken into account in the analysis. It is potentially worthwhile capturing the amount of improvement or worsening in different systems, particularly over time. A drug that induced improvement in more systems than another would then be more easily recognized as doing so. This is essentially why, from a technical point of view, dichotomization of variables results in loss of efficiency [9]. In addition, there is difficulty in accounting for all the changes in disease activity occurring between the beginning and end of a study. It has to be borne in mind that prolonged remission is uncommon in the context of clinical trials of SLE, and changes in disease activity in both directions (improvement and worsening) may occur concomitantly across different systems. Although a numerical score is able to capture this globally, it is not possible to identify how changes in disease activity occurred across different systems.

The ideal solution would be a representation of the BILAG-2004 system scores that is clinically relevant and meaningful that allows for the gradation of response according to severity, avoids dichotomization and takes into consideration the categorical nature of the data. Furthermore, it has to be relatively easy to analyse as compared with categorical data. Using these principles, we performed exploratory analyses on the data from a large longitudinal study of SLE patients seen in routine clinical practice. Based on our analysis, we propose a novel way of representing the BILAG-2004 index scores longitudinally in clinical studies, especially for clinical trials. This new method captures the total number of systems with defined transitions in disease activity (reflecting improvement, worsening or persistence) that can be analysed as a single observation (between two assessments) or cumulatively over time.

Patients and methods

Data from a multicentre longitudinal study in the UK that was primarily designed to validate the BILAG-2004 index were used in the analysis [4]. All patients satisfied the revised ACR criteria for classification of SLE [10, 11]. Patients were excluded from the study if they were pregnant, aged <18 years or unable to give valid consent. The original study to which this study is an additional analysis received ethics approval (Hull and East Riding Research Ethics Committee) and was carried out in accordance with the Declaration of Helsinki. Written consent was obtained from all patients.

This study has been described in detail previously [4]. In summary, patients were followed up prospectively and data (BILAG-2004 index and treatment) were collected for all consecutive visits or encounters that the patients had with the physician. For this analysis, changes in disease activity and treatment between two consecutive visits were studied. A robust definition for change in therapy between consecutive visits was used as the external reference for change in disease activity. This definition has been described previously [4] and is available as supplementary data (Section A), available at *Rheumatology* Online. Three categories of changes in therapy were defined as follows: no change, increase in therapy and decrease in therapy.

Statistical analysis

Exploratory analyses were performed using external responsiveness methodology [12]. It assesses the extent to which changes in the index over time relate to corresponding changes in therapy (external reference) between two consecutive visits. Therefore each observation for the analysis was derived from two consecutive visits.

Initially, maximum-likelihood multinomial logistic regression was used to assess external responsiveness, with change in therapy as the outcome variable and changes in disease activity (according to the index) as the explanatory variables. This methodology essentially fits two binary logistic regression models. The first discriminates between increases in therapy and no change in therapy between visits, with no change in therapy being regarded as the baseline category. The second discriminates between decreases in therapy and no change in therapy. Explanatory variables are defined based on changes in disease activity, with minimal change in activity taken as the reference category. Thus the association between change in disease activity and change in therapy was assessed in both directions (increase and decrease). The comparison between increase in therapy and decrease in therapy is implicit in this model and is not directly estimated. The results were reported as coefficients, with 95% CI. A coefficient value >0 for a particular change in score within the comparison between increase in therapy and no change in therapy indicates that the change in score is associated with increase in therapy. Conversely,

a negative coefficient value (<0) for a particular change in score within the comparison between increase in therapy and no change in therapy indicates that the change in score is associated with no change in therapy (and not with decrease in therapy) or equivalently an inverse association with increase in therapy. This interpretation applies similarly to the comparison between decrease in therapy and no change in therapy.

In particular, we devised a new method of classifying changes in the BILAG-2004 system scores, and the appropriateness of this scheme was assessed using the external responsiveness analysis. In addition, receiver operating characteristic (ROC) curve analysis was used to describe the performance of the various models for representing the BILAG-2004 index scores [13]. For this purpose, areas under the curve (AUC) were estimated from relevant logistic regression models. One model focused on the analysis of increase in therapy (indicating deterioration in disease activity) versus no change in therapy, and another model focused on increase in therapy versus no increase in therapy (combination of no change in therapy and decrease in therapy). Similarly two models were fitted to focus on decrease in therapy (indicating improvement in disease activity) versus no change in therapy and decrease in therapy versus no decrease in therapy (combination of no change in therapy and increase in therapy).

All statistical analyses were performed using Stata for Windows version 8 (StataCorp, College Station, TX, USA) and R [14]. Robust variance estimation was used in the analyses to accommodate multiple assessments from the same patients [15].

Results

There were 347 SLE patients (92.9% female patients, 57.9% Caucasian, 20.5% Afro–Caribbean and 19% South Asian) with 1761 assessments that contributed 1414 observations for analysis (demographics summarized in supplementary Table B, available at *Rheumatology* Online). The median duration of follow-up was 11 months (range 1–26 months). Increase in treatment between consecutive visits occurred in 22.7% of observations, while 37.3% of observations had therapy decreased, and in 40% of observations there was no change in treatment. The distribution of changes in disease activity according to BILAG-2004 and change in therapy is available as supplementary Table C, available at *Rheumatology* Online.

Development of BILAG-2004 systems tally

Based on the external responsiveness of the changes in the BILAG-2004 index system scores [4], we devised a new method of classifying changes in the BILAG-2004 index system scores, using counts of systems with specified transitions in scores. It records the number of systems in which activity increased, decreased or remained the same between two consecutive visits and expresses this as a tally (BILAG-2004 systems tally, BST). It has the following six components:

(i) Number of systems with major deterioration (change from grade B/C/D/E to A or grade D/E to B).
(ii) Number of systems with minor deterioration (change from grade C to B).
(iii) Number of systems with persistent significant activity (no change from grade A or B).
(iv) Number of systems with major improvement (change from grade A to C/D or grade B to D).
(v) Number of systems with minor improvement (change from grade A to B or grade B to C).
(vi) Number of systems with persistent minimal or no activity (change from grade C/D/E to C/D/E).

This was further simplified into three components (simplified BILAG-2004 systems tally, sBST) by grouping major deterioration, minor deterioration and persistent activity into a single group, and major improvement with minor improvement into another group:

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

Further details on the development of the BST and sBST are available in the supplementary data (Section B), available at *Rheumatology* Online. We examined both these measures, as we did not wish to be limiting or prescriptive in our investigation.

Assessment of BST and sBST

The external responsiveness analysis of BST demonstrated that it was appropriate for use, as it had the expected significant associations with change in therapy (Table 1). The number of systems with major deterioration, minor deterioration and persistent significant activity were independently associated with increase in therapy. On the other hand, the number of systems with major improvement and minor improvement were independently associated with decrease in therapy. Similarly the sBST was shown to be appropriate for use, as the number of systems with active/worsening disease were independently associated with increase in therapy, and the number of systems with improving disease were independently associated with decrease in therapy (Table 2). Table 3 summarizes the distribution of changes in therapy according to the sBST components, and it has to be noted that these components are not mutually exclusive.

A formal test of the linearity assumption for the sBST components in Table 2 compared a model with five level factor versions of each variable (corresponding to values of 0, 1, 2, 3 and 4) with the model of Table 2. This generated a significance level of 0.42 (χ^2 value of 12.3 on 12 degrees of freedom). As there are few values of 4 in either variable, and this creates some numerical instability, these were also grouped with values of 3, and this led to a *P*-value of 0.23

Table 1 External responsiveness of the BST with multinomial logistic regression ($n = 1414$)

BST[a]	Number of observations	Increase in therapy[b] coefficient (95% CI)	Decrease in therapy[b] coefficient (95% CI)
Number of systems with major deterioration[c]		2.82 (2.34, 3.30)[d]	−0.22 (−0.79, 0.35)
One system with major deterioration	144	–	–
Two systems with major deterioration	23	–	–
Three systems with major deterioration	3	–	–
Number of systems with minor deterioration[c]		1.88 (1.22, 2.55)[d]	−0.02 (−0.63, 0.59)
One system with deterioration	94	–	–
Two systems with deterioration	9	–	–
Number of systems with persistent significant activity[c]		1.64 (1.21, 2.06)[d]	−0.38 (−0.79, 0.03)
One system with persistent significant activity	166	–	–
Two systems with persistent significant activity	11	–	–
Three systems with persistent significant activity	2	–	–
Number of systems with minor improvement[c]		−0.28 (−0.73, 0.17)	0.33 (0.04, 0.62)[d]
One system with improvement	158	–	–
Two systems with improvement	23	–	–
Three systems with improvement	7	–	–
Number of systems with major improvement[c]		0.18 (−0.26, 0.63)	0.56 (0.23, 0.89)[d]
One system with major improvement	150	–	–
Two systems with major improvement	18	–	–
Three systems with major improvement	5	–	–
Number of systems with persistent minimal or no activity[c]		0	0
Five or fewer systems with persistent minimal or no activity	51	–	–
Six systems with persistent minimal or no activity	144	–	–
Seven systems with persistent minimal or no activity	441	–	–
Eight or nine systems with persistent minimal or no activity	778	–	–

[a]Major deterioration: change of grade B/C/D/E to A or grade D/E to B. Minor deterioration: change of grade C to B. Persistent significant activity: no change from grade A or B. Major improvement: change of grade A to C/D or grade B to D. Minor improvement: change of grade A to B or grade B to C. Persistent minimal or no activity: change of grade C/D/E to C/D/E. [b]As compared with no change in therapy. [c]Breakdown of number of observations with the respective number of systems in each of the BST components. [d]Statistically significant association.

Table 2 External responsiveness of the sBST with multinomial logistic regression ($n = 1414$)

sBST[a]	Number of non-zero observations	Increase in therapy[b] coefficient (95% CI)	Decrease in therapy[b] coefficient (95% CI)
Number of systems with active/worsening disease	391	2.08 (1.72, 2.43)[c]	−0.24 (−0.58, 0.09)
Number of systems with improving disease	330	−0.05 (−0.39, 0.29)	0.43 (0.21, 0.65)[c]
Number of systems with persistent minimal or no activity	1414	0	0

[a]Active/worsening disease: systems with major deterioration, minor deterioration and persistent significant activity. Improving disease: systems with major improvement and minor improvement. Persistent minimal or no activity: change of grade C/D/E to C/D/E. [b]As compared with no change in therapy. [c]Statistically significant association.

for a test of non-linearity (χ^2 value of 10.5 on 8 degrees of freedom). From this analysis, the estimated coefficients for the factors corresponding to one system, two systems and three or four systems with active/worsening disease were 2.4, 3.6 and 4.4 for the increase in therapy component of the model. Comparable coefficients for the number of systems improving for the decrease in therapy component were 0.62, 0.66, and 1.01.

Performance of BST and sBST

ROC curve analyses for BST and sBST are summarized in Table 4, with other methods for analysing the BILAG-2004 index scores included for comparative purposes. For the BILAG-2004 numerical scoring [5], the combination of change in numerical score and the numerical score of the previous visit has superior performance to the single variable of just change in score, in a

Table 3 Cross tabulation of the sBST with change in therapy ($n = 1414$)

sBST[a]	Number of observations	Increase in therapy (%)	No change in therapy (%)	Decrease in therapy (%)
Number of systems with active/worsening disease[b]				
One system with active/worsening disease	295	155 (52.5)	77 (26.1)	63 (21.4)
Two systems with active/worsening disease	80	65 (81.3)	10 (12.5)	5 (6.3)
Three systems with active/worsening disease	14	13 (92.9)	1 (7.1)	0 (0)
Four systems with active/worsening disease	2	2 (100)	0 (0)	0 (0)
Number of systems with improving disease[b]				
One system with improving disease	256	55 (21.5)	81 (31.6)	120 (46.9)
Two systems with improving disease	56	3 (5.4)	21 (37.5)	32 (57.1)
Three systems with improving disease	14	1 (7.1)	5 (35.7)	8 (57.1)
Four systems with improving disease	4	1 (25.0)	0 (0)	3 (75.0)

[a]Active/worsening disease: systems with major deterioration, minor deterioration and persistent significant activity. Improving disease: systems with major improvement and minor improvement. [b]Breakdown of the number of observations with the respective number of systems in each of the sBST components.

Table 4 AUC values from ROC curve analysis of the different models of the BILAG-2004 index scores

Model	Increase in therapy		Decrease in therapy	
	Versus no change in therapy	Versus no increase in therapy	Versus no change in therapy	Versus no decrease in therapy
Original BILAG-2004 index system scores (nine separate changes in system scores)	0.75	0.75	0.59	0.65
Counts of BILAG-2004 systems with transition in scores (as in supplementary Table C, available at *Rheumatology* Online)	0.85	0.86	0.59	0.68
BST (as in Table 2)	0.82	0.83	0.57	0.66
sBST (as in Table 3)	0.81	0.81	0.57	0.65
Change in BILAG-2004 numerical score	0.73	0.75	0.58	0.65
BILAG-2004 numerical score variables (change in numerical score and previous visit numerical score included in model)	0.84	0.85	0.58	0.67

similar fashion to SLEDAI-2000, which we have previously demonstrated [16]. BST and sBST performed well, in particular with increase in therapy (indicating deterioration in disease activity), with an AUC >0.80. Apart from the analysis using the original BILAG-2004 index system scores (with nine categorical system scores) and the single variable of change in BILAG-2004 numerical score, all the other models for analysing the BILAG-2004 index scores have similar performance characteristics.

Table 4 is included to give a simple illustration of the relative value of the different forms of the BILAG-2004 data, and formal statistical analyses are not presented. Any particular weighted sums of the different components of BST or sBST, and in particular the weightings derived from the logistic regressions that underlie Table 4, are not being advocated. In a clinical trial, treatment comparisons would be based on relevant aspects of these tallies (see Discussion). However, for indicative purposes, we note that a 95% bootstrap CI for the AUC value of 0.83, estimated for use of the BST when related to increase in therapy versus no increase in therapy, is (0.81, 0.86). Comparable intervals for the sBST and original BILAG-2004 index system score values of 0.81 and 0.75 are (0.78, 0.85) and (0.70, 0.78). This illustrates that the precision of the estimated AUCs in Table 4 supports the general conclusions drawn. In addition, because the regression coefficients underlying the estimates in

Table 4 are generally distinct, the ranking of the model scores, which essentially defines the estimated AUC values, would not differ greatly if the coefficients varied somewhat. Thus the problem of optimism of the estimates, because they are evaluated on the same data as are used to define the models, is also not a major concern. For example, a calculation of the estimated optimistic bias in the AUC for the BST value of 0.83, using the bootstrap like resampling procedure of Harrell *et al.* [17], gives the small value of 0.005. Nevertheless, this does not negate the value of validating the observed performance in other observational and clinical trial data sets that may have a different patient mix (see Discussion).

Discussion

We have devised a data-driven method of representing the BILAG-2004 index system scores longitudinally from our large data set of patients. This BST is based on counts of the number of systems with active/worsening disease and improving disease. It has a comprehensive form with six components (BST) and a simplified form (sBST). As there are several components to the BST and sBST, it is not as simple as a global score.

Although these tallies seem complex with two different forms and made up of several components, they are essentially a simplification of the changes in the nine categorical system scores of the BILAG-2004 index that is clinically meaningful and maintains gradation according to severity. It provides a summary of the number of systems that have active/worsening disease and those that have improved between two time points. It is an alternative to the BILAG-2004 numerical scoring or a global score. A potential advantage is that drug regimens that induce more or less improvement (or worsening) in multiple systems in a group of patients can be distinguished without requiring disease activity in all systems to change in the same direction.

The components of sBST can be regarded as the composite outcome of the respective components of the BST, in similar fashion to the DAS28 score, which is the composite of inflammatory joint counts, ESR and patient's visual analogue score. Even though data will be collected on all components of the tally in clinical studies, not every component needs to be used in the analysis and different emphasis can be placed on different components, depending on the question or hypothesis of the study. Each of the components of the tally can be analysed separately, and it is anticipated that, in many cases, one or more of them would be used as the outcome of interest, rather than all the different components. In a clinical trial it might be the number of systems with active/worsening disease over time that would be of interest in the analysis. The primary outcome of interest could be the difference in the number of systems with major deterioration over time between the treatment arms, whereas the secondary outcomes could be the number of systems with minor deterioration, the number of systems with persistent significant activity and the composite of the number of systems with active/worsening disease. In other studies the primary or initial focus might be on improvements. The choice between BST and sBST may be application specific or one or the other may prove to be generally preferred. The nature of the study and the hypothesis to be tested will most likely determine whether BST or sBST should be used in a clinical study. Based on our analysis, we could not make a recommendation of one over the other.

As BST and sBST represent measurement of change in disease activity between two time points, the summation of the number of systems with active/worsening disease for patients over time would reflect the burden of disease activity during the period of study. If the time intervals between visits are equal or have little variation across patients in the study, this summation of the number of systems with active/worsening disease over the study period may be considered similar to an AUC-type measure. Although calculation of the AUC is an attractive form of analysis, caution should be exercised, as there may be difficulty in its interpretation, especially when the interval between assessments is prolonged (particularly >3 months apart), and if the intervals are variable (as in observational studies). An alternative, and more flexible, analysis of the number of active/worsening systems for two groups of patients could be based on ordinal regression models. Ordinal regression of this outcome could be performed at a specified time point or longitudinally. One approach to the latter is to use a generalized linear mixed model with random patient effects and a complementary log-log link. Further advantages of a regression approach is that the analysis can be stratified on relevant entry characteristics, such as disease activity levels, and patients with partially missing data can contribute to the analysis. In addition, patterns of disease activity over time can be investigated.

In essence, BST and sBST combine the flexibility and simplicity of numerical scoring with the clinical intuitiveness and meaningfulness of the original BILAG-2004 index categorical score. Specifically, this new scheme does not involve dichotomization of the outcome, hence it does not suffer from the loss of efficiency associated with reliance on a simple yes/no response [9]. Greater efficiency in the detection of group differences means that fewer patients will be required in comparative studies, and this will reduce the cost of running such studies. Thus, in our opinion, this new scheme is a better and more efficient way of reporting differences in efficacy between treatment arms in clinical trials as compared with the dichotomous variables that are currently being used most frequently.

The main limitation of this exploratory analysis is that the BST and sBST are derived and assessed using the same data set. Further work will require assessment using an independent data set and validation of its use in clinical trials.

In conclusion, the BST represents a new method of representing the BILAG-2004 system scores longitudinally.

Rheumatology key messages

- The BST and sBST provide alternative approaches to representing BILAG-2004 disease activity longitudinally.
- The BST and sBST combine the flexibility and simplicity of numerical scoring with the clinical intuitiveness of the BILAG-2004 score.

Acknowledgements

The authors would like to thank the nurse specialists of all participating centres, Wellcome Trust Clinical Research Facility (Birmingham), Lupus (UK), Arthritis Research (UK) and Vifor Pharma/Aspreva Pharmaceuticals for their support. V.T.F. was supported by MRC (UK) funding U105261167.

Funding: This study was supported by a grant from Arthritis Research (UK) (grant no. 16081).

Disclosure statement: I.N.B. has received speaker's and/or consultancy fees from Human Genome Sciences, Roche Pharmaceuticals, Glaxo Smith Kline and UCB. He has also received grant income from Roche Pharmaceuticals, Glaxo Smith Kline and UCB. C.G. has received consultancy payments and honoraria from Roche Pharmaceuticals, Biogen, Merck Serono, Genentech, Glaxo Smith Kline, Bristol Myers Squibb, UCB and Vifor Pharma/Aspreva Pharmaceuticals. All other authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* Online.

References

1 Isenberg DA, Rahman A, Allen E *et al*. BILAG 2004. Development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. Rheumatology 2005;44:902–6.

2 Yee CS, Farewell V, Isenberg DA *et al*. Revised British Isles Lupus Assessment Group 2004 index: a reliable tool for assessment of systemic lupus erythematosus activity. Arthritis Rheum 2006;54:3300–5.

3 Yee CS, Farewell V, Isenberg DA *et al*. British Isles Lupus Assessment Group 2004 index is valid for assessment of disease activity in systemic lupus erythematosus. Arthritis Rheum 2007;56:4113–19.

4 Yee CS, Farewell V, Isenberg DA *et al*. The BILAG-2004 index is sensitive to change for assessment of SLE disease activity. Rheumatology 2009;48:691–5.

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

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

7 Furie RA, Petri MA, Wallace DJ *et al*. Novel evidence-based systemic lupus erythematosus responder index. Arthritis Rheum 2009;61:1143–51.

8 Wallace DJ, Kalunian K, Petri M *et al*. Epratuzumab demonstrates clinically meaningful improvements in patients with moderate to severe systemic lupus erythematosus (SLE): results from EMBLEM a phase IIb study. Lupus 2010;19(1 Suppl):14.

9 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;99:822–31.

10 Tan EM, Cohen AS, Fries JF *et al*. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.

11 Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1997;40:1725.

12 Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000;53:459–68.

13 Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993;39:561–77.

14 Hornik K. The R FAQ. Comprehensive R Archive Network 2010. http://CRAN.R-project.org/doc/FAQ/R-FAQ.html (26 July 2012, date last accessed).

15 Williams RL. A note on robust variance estimation for cluster-correlated data. Biometrics 2000;56:645–6.

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

17 Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361–87.