

# The BILAG-2004 index is sensitive to change for assessment of SLE disease activity

Chee-Seng Yee<sup>1</sup>, Vernon Farewell<sup>2</sup>, David A. Isenberg<sup>3</sup>, Bridget Griffiths<sup>4</sup>, Lee-Suan Teh<sup>5</sup>, Ian N. Bruce<sup>6</sup>, Yasmeen Ahmad<sup>7</sup>, Anisur Rahman<sup>3</sup>, Athiveeraramapandian Prabu<sup>1</sup>, Mohammed Akil<sup>8</sup>, Neil McHugh<sup>9</sup>, Christopher Edwards<sup>10</sup>, David D'Cruz<sup>11</sup>, Munther A. Khamashta<sup>11</sup>, Peter Maddison<sup>12</sup> and Caroline Gordon<sup>1</sup>

**Objective.** To determine if the BILAG-2004 index is sensitive to change for assessment of SLE disease activity.

**Methods.** This was a prospective multi-centre longitudinal study of SLE patients. At every assessment, data were collected on disease activity (BILAG-2004 index) and treatment. Analyses were performed using overall BILAG-2004 index score (as determined by the highest score achieved by any of the individual systems) and all the systems scores. Sensitivity to change was assessed by determining the relationship between change in disease activity and change in therapy between two consecutive visits. Statistical analyses were performed using multinomial logistic regression.

**Results.** There were 1761 assessments from 347 SLE patients that contributed 1414 observations for analysis. An increase in therapy between visits occurred in 22.7% observations, while 37.3% had a decrease in therapy and in 40.0% therapy was unchanged. Increase in overall BILAG-2004 index score was associated with increase in therapy and inversely associated with decrease in therapy. Decrease in overall BILAG-2004 index score was associated with decrease in therapy and was inversely associated with increase in therapy. Changes in overall BILAG-2004 index score were differentially related to change in therapy, with greater change in score having greater predictive power. Increase in the scores of most systems was independently associated with an increase in treatment and there was no significant association between decreases in the score of any system with an increase in therapy.

**Conclusions.** The BILAG-2004 index is sensitive to change and is suitable for use in longitudinal studies of SLE.

**KEY WORDS:** BILAG-2004, SLE, Outcome measures, Epidemiology, Statistics, Sensitivity to change, Responsiveness, Disease activity.

## Introduction

SLE is a multi-system autoimmune disease with diverse immunological and clinical manifestations. The BILAG-2004 index, a comprehensive composite clinical disease activity index, has been developed for the assessment of SLE disease activity [1–3]. It has been demonstrated to be reliable and has construct and criterion validity [1–3]. However, it needs to be shown to be sensitive to change before it can be used in longitudinal studies of SLE.

Sensitivity to change or responsiveness of an index implies its ability to change with time. There are two forms of responsiveness: internal and external responsiveness [4]. Internal responsiveness is the ability of the index to change over a particular period of time. Traditional statistical methods that have been used to assess sensitivity to change of several SLE disease activity indices fall into this category. The main disadvantage of these methods is that the changes in the index do not relate to changes in an external measure, hence statistically significant change in the index may occur without corresponding change in clinical status. The comparison is made at the population level and may not reflect clinical change at individual patient level. Furthermore,

comparison across studies is difficult as the statistical methods used are not independent of study design.

External responsiveness refers not just to the ability of the index to change over time, but also includes how the changes in the index relate to the corresponding changes in an external reference. It characterizes the relationship between change in the index and change in the external reference at the individual patient level. The result is generalizable across studies allowing for comparison. Therefore, external responsiveness is a more robust method of assessing sensitivity to change of an index.

We report on the sensitivity to change of the BILAG-2004 index, using the external responsiveness method, for the assessment of SLE disease activity.

## Patients and methods

This longitudinal study involved eight centres in the UK. All patients met the revised ACR criteria for classification of SLE [5, 6]. Patients were excluded from the study if they were pregnant, <18 years of age or unable to give valid consent. This study received multi-centre research ethical approval from Hull and East Riding Research Ethics Committee and approval from local research ethics committees. Written consent was obtained from all patients. This study was carried out in accordance with the Declaration of Helsinki.

Patients were followed up prospectively and data (disease activity using BILAG-2004 index and treatment) were collected for all consecutive visits/encounters (inpatient or outpatient) that the patients had with their physicians. The majority of the patients were also involved in the cross-sectional validation study that has been reported [3]; however, this longitudinal study was of longer duration and the analysis was completely different.

## BILAG-2004 index (BILAG-2004)

This is an ordinal scale index with items distributed across nine systems. It was developed based on the principle of the

<sup>1</sup>Rheumatology Research Group, University of Birmingham, Birmingham, <sup>2</sup>MRC Biostatistics Unit, University of Cambridge, Cambridge, <sup>3</sup>Centre for Rheumatology, University College London, London, <sup>4</sup>Department of Rheumatology, Freeman Hospital, Newcastle-upon-Tyne, <sup>5</sup>Department of Rheumatology, Royal Blackburn Hospital, Blackburn, <sup>6</sup>ARC Epidemiology Unit, University of Manchester, Manchester, <sup>7</sup>Department of Rheumatology, North West Wales NHS Trust, Wales, <sup>8</sup>Department of Rheumatology, Sheffield Teaching Hospitals NHS Trust, Sheffield, <sup>9</sup>Department of Rheumatology, Royal National Hospital for Rheumatic Diseases NHS Trust, Bath, <sup>10</sup>Department of Rheumatology, University of Southampton, Southampton, <sup>11</sup>Lupus Research Unit, St Thomas Hospital, London and <sup>12</sup>Department of Rheumatology, University of Wales, Bangor, UK.

Submitted 19 September 2008; revised version accepted 26 February 2009.

Correspondence to: Chee-Seng Yee, Rheumatology Research Group, Division of Infection and Immunity, The Medical School (East Wing), University of Birmingham, Birmingham, B15 2TT, UK. E-mail: cseyee@yemail.com

'physician's intention to treat'. Disease activity is categorized into five levels (Grade A—very active, Grade B—moderate activity, Grade C—mild activity, Grade D—no activity but previously affected and Grade E—no current or previous disease activity).

Changes were made to the scoring of the renal system for Grades A and B during the progress of this study to improve the scoring system with regards to proteinuria. The changes made were:

- (i) urine protein dipstick result is superseded by other methods of urine protein estimation (urine albumin-creatinine ratio, urine protein-creatinine ratio or 24 h urine protein) where available;
- (ii) the threshold for definition of improvement in proteinuria was reduced from 50 to 25%;
- (iii) an additional Category B criterion was added for urine protein excretion which was of at least 0.5 g/day (or equivalent) that has not improved by at least 25%.

These changes were laboratory based and had no impact on data collection as it only affected the way renal system score was calculated. The renal system score incorporating these changes was used in the analysis.

During this study, a few other issues related to the glossary were noted and minor changes were made to the index after the completion of this study. Although these changes were not accounted for in the analysis, they would not have had a major effect on the results of this study. The revised index (BILAG-2004 index form, glossary and scoring scheme) incorporating all the above changes is available as supplementary data at *Rheumatology* Online.

### Change in therapy

As there is no gold standard for disease activity, change in therapy between consecutive visits was used as the external reference for change in disease activity in the analysis. Change in therapy was the change in treatment between two consecutive visits (or the difference in treatment after the patient was assessed at the index visit, as compared with the therapy prescribed following the previous visit).

The medications of interest were immunosuppressives, anti-malarials, glucocorticoids, biological therapy, topical glucocorticoids, topical immunosuppressives, intravenous immunoglobulins, plasmapheresis, anti-coagulation, prasterone, thalidomide and retinoids.

A robust definition for change in therapy was used with three categories of changes in therapy defined: 'no change', 'increase in therapy' and 'decrease in therapy'. This definition is similar to the one used in the criterion validity analysis of the cross-sectional study [3] and is available as supplementary data (Definition of Change in Therapy and Table A) at *Rheumatology* Online.

### Statistical analysis

The sensitivity to change of the index was assessed using the external responsiveness method as outlined by Husted *et al.* [4]. The extent to which changes in BILAG-2004 score between two consecutive visits relate to the corresponding changes (actual) in therapy (external reference) was studied. Therefore, two consecutive visits give rise to one observation. This is different from the criterion validity analysis of the cross-sectional study which looked at the correlation between the score of the index and change in therapy at a single visit [3].

The overall BILAG-2004 score (overall score), as determined by the highest score achieved by any of the individual systems, was used in the analysis. Further analysis using all the nine system scores was also performed. Scores of Grades D and E were combined as both indicate inactivity.

Maximum likelihood multinomial logistic regression was used in the analysis, with change in therapy as the outcome variable and change in BILAG-2004 score as the explanatory variable. The baseline comparator for change in BILAG-2004 score was 'no change in score' or 'minimal change in score'. In addition to 'no change in score', 'minimal change in score' also includes the change from Grade D/E to C, as this change is considered minor and therapy rarely changes. The baseline comparator for change in treatment was 'no change in therapy'. As the association between change in BILAG-2004 score and change in therapy was assessed in both directions (increase and decrease), the baseline comparator for these variables were chosen as such.

The results were reported in coefficients with 95% CIs. As this was a multinomial regression analysis with 'no change in therapy' as the baseline comparator, two separate analyses were involved:

- (i) comparison between 'increase in therapy' and 'no change in therapy' and
- (ii) comparison between 'decrease in therapy' and 'no change in therapy'.

There was no direct comparison between 'increase in therapy' and 'decrease in therapy'. A coefficient value  $>0$  for a particular category of change in BILAG-2004 score within the comparison between 'increase in therapy' and 'no change in therapy' indicates that the change in score category is associated with 'increase in therapy'. Inversely, a negative coefficient value ( $<0$ ) for a particular category of change in BILAG-2004 score within the comparison between 'increase in therapy' and 'no change in therapy' indicates that the change in score category 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'.

As the majority of the patients contributed more than one observation, independence of observations from the same patient could not be assumed and robust variance estimation was used [7]. All statistical analyses were performed using Stata for Windows version 8 (Stata Corporation, TX, USA).

## Results

There were 1761 assessments from 347 SLE patients (92.9% females, 57.9% Caucasian, 20.5% Afro-Caribbean and 19% South Asian) that contributed 1414 observations for analysis. The mean age was 40.9 years (s.d. 12.9) and mean disease duration was 8.2 years (s.d. 7.8). The median duration of follow-up was 11 months (range 1–26) and median number of assessments per patient was four (range 2–18). Increase in treatment between consecutive visits occurred in 22.7% of the observations while 37.3% had therapy decreased, and in 40.0%, there was no change in treatment. The distribution of the changes in disease activity according to BILAG-2004 and change in therapy is available as supplementary data (Table B) available at *Rheumatology* Online.

### Analysis using overall score

Increase in the overall score was associated with increase in therapy (coefficient 1.35; 95% CI 1.01, 1.70) and inversely associated with decrease in therapy (coefficient  $-0.44$ ; 95% CI  $-0.81$ ,  $-0.06$ ). Decrease in the overall score was associated with decrease in therapy (coefficient 0.44; 95% CI 0.16, 0.71) and inversely associated with increase in therapy (coefficient  $-0.79$ ; 95% CI  $-1.27$ ,  $-0.32$ ).

When the minor change of score from Grade D/E to C was excluded from the definition of increase in activity (defined as minimal change in activity), increase in the overall scores had a much greater predictive power of increase in therapy (coefficient 2.25; 95% CI 1.81, 2.70).

TABLE 1. Analysis of the subgroups of changes in overall BILAG-2004 index score and its association with change in therapy using multinomial logistic regression ( $n = 1414$ )

Change in overall BILAG-2004 score	Number of observations	Increase in therapy <sup>a</sup> coefficient (95% CI)	Decrease in therapy <sup>a</sup> coefficient (95% CI)
Minimal change in activity	848	0	0
Increase in activity to Grade A	48	2.86 (1.85, 3.86)	-0.88 (-2.55, 0.78)
Increase in activity to Grade B	147	2.07 (1.59, 2.56)	-0.14 (-0.69, 0.41)
Decrease in activity to Grade B	41	-0.37 (-1.33, 0.59)	0.37 (-0.28, 1.01)
Decrease in activity to Grade C or D	330	-0.77 (-1.27, -0.27)	0.51 (0.24, 0.79)

<sup>a</sup>As compared with 'no change in therapy'.

Further analysis revealed that increase in the score to Grade A was more likely than increase in the score to Grade B to be associated with increase in therapy. Decrease in the score to Grade C/D was much less likely to have increase in treatment than decrease in the score to Grade B (Table 1). Therefore, changes in overall score were differentially related to change in therapy, with greater change in score having greater predictive power.

#### Analysis using system scores

Similar analysis using all the system scores (instead of overall score) as explanatory variables for the outcome variable of change in therapy is summarized in Table 2. Apart from the ophthalmic, haematological and renal systems, increase in the scores of the other systems was independently associated with treatment increase. The association between decrease in the system score and treatment reduction could not be demonstrated as consistently. There was no significant association between decreases in score of any system with increase in therapy.

There were too few observations with an increase in the ophthalmic system score to demonstrate an association between increase in score and increase in therapy (Table B of supplementary data, available at *Rheumatology* Online). In the haematological system, the increases in score were due to minor worsening of leucopenia, neutropenia or anaemia of chronic disease (resulting in an increase to Grade B) which were not treated. There was no occurrence of major haematological manifestation of active disease in this cohort of patients, such as severe thrombocytopenia or haemolytic anaemia.

The non-significant association between increase in the renal system score with treatment increase was due to seven observations with an increase in the score that had therapy unchanged, while there were 13 observations with an increase in the score in which the treatment was increased. In addition, there were 15 observations in which therapy was decreased. Thus, there were in total 22 observations with an increase in the score to Grades A/B that were not accompanied by an increase in therapy (Table B of supplementary data, available at *Rheumatology* Online). There were three main reasons for this discrepancy.

First, many of these observations were due to persistent activity in the renal system. As these patients were assessed frequently (about every 1–2 months), minor changes in the levels of proteinuria between consecutive visits led to fluctuations in the system score between adjacent categories (from Grade B to A or Grade C to B). For example, a change of urine albumin-creatinine ratio from 122 to 57 mg/mmol and subsequently to 73 mg/mmol on three consecutive visits would result in the score changing from Grade B to C and finally back to B.

TABLE 2. Sensitivity to change analysis of the BILAG-2004 index using system scores with multinomial logistic regression after excluding change of Grade D/E to C from definition of increase in activity ( $n = 1414$ )

Change in BILAG-2004 system score	Number of observations	Increase in therapy <sup>a</sup> coefficient (95% CI)	Decrease in therapy <sup>a</sup> coefficient (95% CI)
<b>Constitutional</b>			
Minimal change	1339	0	0
Increase	11	$\infty^b$	0.02 (-0.35, 0.38)
Decrease	64	0.36 (-0.68, 1.40)	0.81 (0.22, 1.39)
<b>Mucocutaneous</b>			
Minimal change	991	0	0
Increase	119	2.32 (1.80, 2.84)	-0.60 (-1.32, 0.12)
Decrease	304	-0.13 (-0.54, 0.29)	0.46 (0.17, 0.75)
<b>Neuropsychiatric</b>			
Minimal change	1373	0	0
Increase	13	$\infty^b$	0.57 (-1.79, 2.93)
Decrease	28	0.08 (-1.39, 1.55)	0.75 (-0.05, 1.55)
<b>Musculoskeletal</b>			
Minimal change	1034	0	0
Increase	88	2.94 (2.19, 3.70)	-0.04 (-1.00, 0.91)
Decrease	292	-0.28 (-0.70, 0.14)	-0.01 (-0.33, 0.31)
<b>Cardiorespiratory</b>			
Minimal change	1302	0	0
Increase	35	1.25 (0.23, 2.26)	-0.41 (-1.46, 0.65)
Decrease	77	0.22 (-0.65, 1.08)	0.35 (-0.20, 0.90)
<b>Gastrointestinal</b>			
Minimal change	1397	0	0
Increase	6	2.29 (-0.22, 4.81)	$-\infty^c$
Decrease	11	-0.11 (-1.40, 1.17)	0.10 (-1.23, 1.44)
<b>Ophthalmic</b>			
Minimal change	1400	0	0
Increase	4	0.03 (-5.76, 5.83)	$-\infty^d$
Decrease	10	1.39 (-0.58, 3.36)	0.48 (-1.40, 2.37)
<b>Renal</b>			
Minimal change	1313	0	0
Increase	35	1.11 (-0.01, 2.22)	0.87 (-0.23, 1.97)
Decrease	66	-0.07 (-0.87, 0.73)	0.57 (0.05, 1.10)
<b>Haematological</b>			
Minimal change	1219	0	0
Increase	17	-0.01 (-1.36, 1.37)	-0.03 (-1.04, 0.98)
Decrease	178	0.02 (-0.51, 0.54)	-0.13 (-0.48, 0.22)

<sup>a</sup>As compared with 'no change in therapy'. <sup>b</sup>All observations had increase in therapy and coefficient significantly  $>0$  ( $P < 0.05$ ). <sup>c</sup>All observations did not have decrease in therapy and coefficient significantly  $<0$  ( $P < 0.05$ ). <sup>d</sup>All observations did not have decrease in therapy but coefficient not significantly different from 0 ( $P \geq 0.05$ ).

Secondly, active urinary sediments may be intermittently present during the treatment of lupus nephritis. The absence of it at one visit followed by the reappearance in the following visit would result in an increase in the system score by one category. Finally, some patients had pre-existing damage but developed superimposed active disease which caused difficulty in distinguishing the contribution of disease activity from damage.

#### Discussion

This multi-centre longitudinal study is the largest prospective validation study to date assessing the sensitivity to change of a SLE disease activity index as compared with previous studies of other disease activity indices [8–13]. As external responsiveness method was used, changes in both directions (increase and decrease) with regards to the index score and change in therapy (external reference) were assessed. It is imperative that the selected external reference measure represents an accepted indicator of change in the patient's status. As change in therapy was chosen as the external reference, clinically meaningful change in the index was studied. Physician's global assessment could have been used, but this has performed unsatisfactorily with poor agreement between physicians in several studies [10, 14, 15].

This study has demonstrated that the BILAG-2004 index is sensitive to change. There was a hierarchical effect of the different subgroups of changes in the index score in its association with



change in therapy, with greater change in scores having greater predictive power of change in therapy. Furthermore, increases in the score in most of the systems were independently associated with increase in therapy and decreases in score in all the systems were not associated with treatment increase. All of this indicated that the changes in the index score had performed as expected, with the possible exception of the renal system. The results may suggest that the index appears to perform better at detecting increase in disease activity (stronger association with treatment increase) as compared with improvement in disease activity in its association with decrease in therapy. However, this may be because decrease in therapy is not a good marker of improvement in disease activity. In practice, increase in therapy is very likely to occur with increase in disease activity and is unlikely to occur with decreasing disease activity. The reverse does not hold true as reduction in disease activity does not necessarily result in treatment reduction, particularly if the patient is already on low-dose therapy, as there are other important factors to be considered such as risk of flares if therapy is reduced further. Hence, the relationship between disease activity and increase in therapy is possibly more informative than that of disease activity with decrease in therapy.

With change in therapy as the reference standard, we could not establish definitively that this index is adequately responsive to improvement in disease activity from very active to moderately active (Grade A to B) as this degree of change in activity is not always reflected by a reduction in therapy. This is because an improvement of the overall score from Grade A to B may occur in three situations with rather different treatment decisions and there are insufficient observations to look at each of these separately. First, this occurs with improvement in the manifestation that had resulted in Grade A previously, whereby there would usually be a reduction or no change in therapy. Another possibility is when the manifestation that resulted in Grade A previously has resolved but there is occurrence of another manifestation within the same system resulting in Grade B. In this situation, the treatment may be increased or left unchanged. Finally, the situation arises when the manifestation that resulted in Grade A has resolved, but there is development of a new Grade B in another system. With this scenario, it is likely that there will be an increase in therapy but it may also be left unchanged depending on the nature of the previous treatment decision. Hence, different circumstances leading to the change in overall score of Grade A to B have different likelihood and direction (increase, decrease or no change) of change in therapy. Nevertheless, we have demonstrated that major improvement from Grade A/B to C/D is significantly associated with decrease in therapy. Furthermore, clinical trials should use the efficacy criteria of improvement to low level activity (Grade C/D) as the main outcome, instead of improvement from Grade A to B.

Some concerns were noted regarding the renal scoring scheme following the completion of the cross-sectional validation study [3], which were also highlighted in another study on renal response rating [16]. This resulted in modifications of the renal scoring system [17]. However, despite these modifications, the results revealed that the changes in the renal system score between consecutive visits do not necessarily accord with the expected change in treatment.

This index is well designed to detect new onset of lupus nephritis or significant resolution of disease activity in the renal system. However, there is an issue with assessments that are close together with regards to fluctuations in the level of proteinuria and intermittent active urine sediments during treatment (resulting in fluctuation of the renal system score between adjacent categories of Grades A and B or Grades B and C). These fluctuations are in fact persistent disease activity that does not necessarily trigger a change in treatment. Treatment is not commonly increased on the basis of a single rise in proteinuria or a single failure of the proteinuria to improve. Trends over a longer period are more

important in assessing the response of disease activity to a change in therapy. Another concern is that isolated sterile pyuria (an element of active urinary sediment) is usually not treated. Differentiating the contribution of disease activity from damage in a patient having both processes occurring concomitantly remains challenging. Indeed, there is a need for an effective way to determine whether (and by how much) a renal manifestation (especially proteinuria) is due to disease activity or damage without resorting to multiple renal biopsies. All these issues with the renal system are not specific to the BILAG-2004 index but apply to all disease activity indices.

Although this index was developed on the principle of the 'physician's intention to treat', using change in therapy as the external reference should not bias the analysis in favour of the index as change in treatment does not determine the scoring. Only the presence of manifestations of active disease will determine the scoring. Furthermore, the scoring of the index was not available to the physician when the treatment decision was made and it is difficult to determine the scoring of the index in routine practice without the appropriate documentation. It should be noted that actual change in therapy involves consideration of many factors other than change in disease activity (such as patients' opinions, previous therapy and presence of co-morbidities) and this is different from intention to treat whereby the main consideration is change in disease activity. Therefore, it is not surprising that change in therapy does not mirror change in disease activity perfectly.

In conclusion, the BILAG-2004 index is sensitive to change. It is suitable for use in longitudinal studies of SLE if the outcome of interest is worsening of disease activity or major improvement in disease activity to low level of activity. However, further work is required to confirm that this index is responsive to improvement in disease activity from very active to moderately active. Furthermore, the issues regarding the renal system (especially the problem of concomitant disease activity and damage) need to be considered in longitudinal studies looking at renal outcome, particularly in clinical trials. We would recommend using more specific criteria to define response in longitudinal studies on lupus nephritis. Nevertheless, this index remains very useful in identifying new activity or significant deterioration due to disease activity in the renal system.

#### Rheumatology key messages

- The BILAG-2004 index is sensitive to change.
- Differentiation of disease activity from damage in SLE is a challenging problem.

#### Acknowledgements

We would like to thank the nurse specialists of all participating centres, Dr Madelynn Chan, Wellcome Trust Clinical Research Facility (Birmingham), Lupus (UK) and Arthritis Research Campaign for their support. Lupus (UK) funded a research associate who helped with recruitment at Royal Blackburn Hospital.

*Funding:* This study was supported by a grant from Arthritis Research Campaign (Grant No. 16081). Funding to pay the Open Access publication charges for this article was provided by the Arthritis Research Campaign.

*Disclosure statement:* D.D. has participated in Advisory Boards for Bristol-Myers Squibb and Novartis and received research grant support from Aspreva. A.R. has received an honorarium from Genexion Ltd for speaking on the subject of BILAG 2004. C.-S.Y. has received honoraria from Roche and Merck Serono and at the time of publication of this article was funded

by an unrestricted educational grant from Aspreva. C.G. has received consultancy payments and honoraria from Roche Pharmaceuticals and UCB for advice on clinical trials and training using the BILAG 2004 index. 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 Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459–68.
- 5 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.
- 6 Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;40:1725.
- 7 Williams RL. A note on robust variance estimation for cluster-correlated data. *Biometrics* 2000;56:645–6.
- 8 Ward MM, Marx AS, Barry NN. Comparison of the validity and sensitivity to change of 5 activity indices in systemic lupus erythematosus. *J Rheumatol* 2000;27:664–70.
- 9 Liang MH, Socher SA, Larson MG, Schur PH. Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus. *Arthritis Rheum* 1989;32:1107–18.
- 10 Guzman J, Cardiel MH, Arce-Salinas A, Sanchez-Guerrero J, Alarcon-Segovia D. Measurement of disease activity in systemic lupus erythematosus. Prospective validation of 3 clinical indices. *J Rheumatol* 1992;19:1551–8.
- 11 Brunner HI, Feldman BM, Bombardier C, Silverman ED. Sensitivity of the Systemic Lupus Erythematosus Disease Activity Index, British Isles Lupus Assessment Group Index, and Systemic Lupus Activity Measure in the evaluation of clinical change in childhood-onset systemic lupus erythematosus. *Arthritis Rheum* 1999;42:1354–60.
- 12 Chang E, Abrahamowicz M, Ferland D, Fortin PR. Comparison of the responsiveness of lupus disease activity measures to changes in systemic lupus erythematosus activity relevant to patients and physicians. *J Clin Epidemiol* 2002;55:488–97.
- 13 Fortin PR, Abrahamowicz M, Clarke AE *et al*. Do lupus disease activity measures detect clinically important change? *J Rheumatol* 2000;27:1421–8.
- 14 Gladman DD, Goldsmith CH, Urowitz MB *et al*. Crosscultural validation and reliability of 3 disease activity indices in systemic lupus erythematosus. *J Rheumatol* 1992;19:608–11.
- 15 Gladman DD, Goldsmith CH, Urowitz MB *et al*. Sensitivity to change of 3 Systemic Lupus Erythematosus Disease Activity Indices: international validation. *J Rheumatol* 1994;21:1468–71.
- 16 Petri M, Kasitanon N, Singh S *et al*. Systemic lupus international collaborating clinics renal activity/response exercise: comparison of agreement in rating renal response. *Arthritis Rheum* 2008;58:1789–95.
- 17 Yee CS. Development and Validation of the BILAG-2004 Index for the assessment of disease activity in Systemic Lupus Erythematosus. PhD Thesis, University of Birmingham, Birmingham, UK.