

Original article

The BILAG2004-Pregnancy index is reliable for assessment of disease activity in pregnant SLE patients

Chee-Seng Yee¹, Mohammed Akil², Munther Khamashta³, Rupa Bessant³, Rachael Kilding², Ian Giles⁴, Vernon T. Farewell⁵ and Caroline Gordon⁶

Abstract

Objective. To assess the inter-rater reliability of the BILAG2004-Pregnancy index for assessment of SLE disease activity in pregnancy.

Methods. Pregnant SLE patients were recruited from four centres and assessed separately by two raters/physicians in routine clinical practice. Disease activity was determined using the BILAG2004-Pregnancy index. Reliability was assessed using level of agreement, κ -statistics and analysis of disagreement. Major disagreement was defined as a score difference of A and C/D/E or B and D/E between the two raters, and minor disagreement was a score difference of A and B or B and C between raters.

Results. A total of 30 patients (63.3% Caucasian, 13.3% Afro-Caribbean, 16.7% South Asian) were recruited. The majority of patients had low-level disease activity according to the local rater's assessment, and there was no grade A activity, with grade B activity present in the following systems: mucocutaneous (nine patients), musculoskeletal (two patients), cardiorespiratory (one patient) and renal (one patient). The distribution of disease activity was similar to the external rater's assessment. Good levels of agreement (>70%) were achieved in all systems. κ -statistics were not appropriate for use in the gastrointestinal, ophthalmic, constitutional and neuropsychiatric systems, as there was minimal variation between patients but good levels of agreement otherwise. There were three major disagreements (0.1 per patient, all differences between B and D/E) and five minor disagreements (0.17 per patient).

Conclusion. The BILAG2004-Pregnancy index is reliable for assessment of disease activity in pregnant SLE patients.

Key words: BILAG2004-Pregnancy, SLE, disease activity, pregnancy, reliability.

Introduction

SLE is a complex multi-system autoimmune disease that predominantly affects women of child-bearing age. With improved management and survival of SLE patients, more patients with SLE are getting pregnant. The assessment of

SLE disease activity is made more challenging during pregnancy, as many pathophysiological changes in pregnancy may be confused with manifestations of SLE disease activity.

The BILAG-2004 index is a system-based disease activity measure that has been validated for use in SLE outside of pregnancy [1-3]. It is one of the preferred disease activity outcome measures used in clinical studies of SLE [4]. Therefore, there is a need to develop a system-based disease activity measure for use in pregnancy to ensure continuity of assessments when patients become pregnant in long-term longitudinal studies of SLE, which use the BILAG-2004 index as the disease activity outcome measure. Furthermore, consistent use of a standardized and validated disease activity outcome measure in the assessment of pregnant SLE patients may help

¹Department of Rheumatology, Doncaster and Bassetlaw Hospitals NHS Foundation Trust, Doncaster, ²Department of Rheumatology, Sheffield Teaching Hospitals NHS Trust, Sheffield, ³Lupus Unit, St Thomas' Hospital, London, ⁴Centre for Rheumatology, University College London, London ⁵MRC Biostatistics Unit, Institute of Public Health, Cambridge and ⁶Rheumatology Research Group, University of Birmingham, Birmingham, UK.

Submitted 22 December 2011; revised version accepted 30 April 2012.

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

address the conflicting reports of the effects of pregnancy on exacerbation of SLE disease activity.

In developing an index for use in pregnancy that is based on the BILAG-2004 index, we have made modifications to account for pathophysiological changes in pregnancy. Generally it retains similarity to the BILAG-2004 index, with the biggest change in the scoring of the renal system, whereby changes in levels of anti-dsDNA and complement (C3 and C4) influence the scoring owing to proteinuria from LN. Otherwise the other changes are mainly in the glossary and definition of items to emphasize to physicians using the index to be aware of pathophysiological changes in pregnancy that might be confused with SLE activity. We have retained the nine systems and the scoring scheme of the BILAG-2004 index, which allows for seamless transition of assessment from the pregnant to the non-pregnant state. This modified index is known as the BILAG2004-Pregnancy index (BILAG2004-P), which comprises the BILAG2004-Pregnancy index form, glossary and scoring (see supplementary data available at *Rheumatology* Online). This cross-sectional study was designed to assess the inter-rater reliability of the BILAG2004-Pregnancy index in assessment of SLE disease activity in pregnancy.

Methods

This was a multicentre cross-sectional study involving four centres across the UK. Pregnant SLE patients, who satisfied the revised ACR criteria for classification of SLE, were recruited [5, 6]. Patients were excluded from the study if they were younger than 18 years or unable to provide valid consent. This study received ethical approval from Trent Research Ethics Committee and local research and development approval from all participating centres (Birmingham Womens' Hospital, University College London Hospitals, St Thomas' Hospital and Sheffield Teaching Hospitals). This study was carried out in accordance with the Declaration of Helsinki, and written consent was obtained from patients.

Patients were assessed independently by two physicians separately, a local rheumatologist/rater and an external rheumatologist/rater (C.-S.Y.). This study was performed in the setting of routine practice, and medical records were available to both raters. The local and external raters were not aware of the outcome of each other's assessment until after completion of the study. Disease activity was determined using the BILAG2004-Pregnancy index. Following the study at each centre there was discussion between the raters on the differences in scoring between them and on issues related to the face and content validity of the index.

Statistical analysis

All statistical analyses were performed with Stata for Windows version 8 (Stata Corp., College Station, TX, USA). For this analysis, BILAG2004-Pregnancy index scores of D and E were combined, as they both indicate inactivity. Inter-rater reliability of the index was assessed

using level of agreement between raters, κ -statistics and analysis of disagreement in scores between raters.

For each system in the index, the percentage of assessments on which both raters agreed was calculated. As κ -statistics perform poorly when there is little variation within the population, they are only used where appropriate (when there is sufficient variation in scores between patients). κ -statistics are based on a simple two-way tabulation of external rater vs local rater scores [7, 8]. Both unweighted and weighted level of agreement and κ -statistics were calculated. The weighting used to calculate the weighted level of agreement and κ -statistics reflects our clinical judgement of the severity of the possible disagreements (Supplementary Table A, available as supplementary data at *Rheumatology* Online). CIs were calculated for κ -statistics, and as the scoring for each system has more than two categories, a bootstrap technique with 1000 replications was used.

Analysis of disagreement calculates the number of disagreements in scores between raters. Disagreement is classified as major disagreement or minor disagreement. The scores can be broadly divided into high-level activity, as represented by A (severe activity) and B (moderate activity) scores that are generally treated, and low-level activity, as represented by C (mild activity) and D/E (inactivity) scores that are usually not treated (treatment may actually be reduced). Therefore major disagreement is defined as the difference between high-level and low-level activity between the two raters (score difference of A and C/D/E or B and D/E), and minor disagreement is defined as a single-level difference in the level of activity between raters (score difference of A and B or B and C). Score difference of C and D/E is not included, as it is considered to be of little clinical significance.

Results

A total of 30 patients were recruited with a mean age of 30.3 years (s.d. 5.0 years), mean disease duration of 5.7 years (s.d. 4.0 years) and mean gestation of 21.8 weeks (s.d. 7.1 weeks). The ethnic distribution was 63.3% Caucasian, 13.3% Afro-Caribbean, 16.7% South Asian and 7.7% others. Based on the local rater's assessment, the majority of patients had low-level disease activity (grades C, D or E), and there was no grade A activity in any patient, with grade B activity present in the following systems: mucocutaneous (nine patients), musculoskeletal (two patients), cardiorespiratory (one patient) and renal (one patient). The distribution was similar using the external rater's assessment: 2 grade A activity (mucocutaneous 1, renal 1) and 10 grade B activity (mucocutaneous 8, musculoskeletal 2).

The two-way tabulation of the local rater's scores against the external rater's scores for five systems, where there was sufficient variation in scores among patients, is shown in Table 1. There was absence of disease activity in gastrointestinal and ophthalmic systems in all patients as assessed by both raters (perfect agreement). Similarly, there was clustering of inactivity in constitutional and neuropsychiatric systems with difference in scores

between the raters of C and D in one patient for each system.

The level of agreement (unweighted and weighted) and κ -statistics (unweighted and weighted) for each system are shown in Table 2. There was good level of agreement (weighted agreement >70%) in all the systems. There were good weighted κ -statistics (>0.40) in mucocutaneous, musculoskeletal, cardiorespiratory, renal and haematological systems. κ -statistics were not appropriate for use in the constitutional, neuropsychiatric, gastrointestinal and ophthalmic systems, as there was very little variation in scores between patients.

There were few disagreements in scores between raters in this study, with three major disagreements (0.10 per patient, all of which were differences in score between B and D/E) and five minor disagreements (0.17 per patient). The major disagreements were in the mucocutaneous (one), cardiorespiratory (one) and renal (one) systems, whereas the minor disagreements were in the mucocutaneous (one difference in score of A and B, one difference in score of B and C), musculoskeletal (two differences in score of B and C) and renal (one difference in score of A and B) systems.

Discussion

This study has demonstrated that the BILAG2004-Pregnancy index is reliable for assessment of disease activity in a representative sample of pregnant SLE women. This index is based on the BILAG-2004 index with modifications to take into account pathophysiological changes of pregnancy.

The majority of the changes were in the glossary, which reminds the physician to differentiate disease activity from pregnancy-related pathophysiological changes such as transient facial blush, melasma/chloasma, bland effusion of knees, mechanical hip/knee pain, pre-eclampsia/eclampsia, haemodilution of pregnancy and HELLP syndrome. The most significant change is in the scoring of the renal system, whereby changes in the anti-dsDNA or complement levels are taken into consideration in the scoring of proteinuria due to SLE disease activity. Furthermore, rising blood pressure is no longer considered to be due to SLE activity, and it has no bearing in the scoring for the renal system. The main reason for inclusion of serological markers is the difficulty in differentiating proteinuria of pre-eclampsia from LN in the

TABLE 1 Two-way tabulation of local rater scores against external rater scores for mucocutaneous, musculoskeletal, cardiorespiratory, renal and haematological systems

	External rater's system score																			
	Mucocutaneous				Musculoskeletal				Cardiorespiratory				Renal				Haematological			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Local rater's system score																				
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	1	7	0	1	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0
C	0	1	1	1	0	1	5	4	0	0	1	0	0	0	0	0	0	0	9	0
D	0	0	3	15	0	0	1	17	0	0	0	28	0	1	0	28	0	0	2	19

Bold values represent perfect agreement between raters.

TABLE 2 Level of agreement and κ -statistics for the BILAG2004-Pregnancy index

Systems	Agreement, % (weighted)	κ (95% CI)	Weighted κ (95% CI)
Constitutional	96.7 (98.3)	NA	NA
Mucocutaneous	76.7 (87.5)	0.59 (0.35, 0.85)	0.73 (0.48, 0.88)
Neuropsychiatric	96.7 (98.3)	NA	NA
Musculoskeletal	76.7 (88.3)	0.53 (0.27, 0.81)	0.60 (0.31, 0.80)
Cardiorespiratory	96.7 (96.7)	0.66 (0, 1)	0.48 (0.31, 0.77)
Gastrointestinal	100 (100)	NA	NA
Ophthalmic	100 (100)	NA	NA
Renal	93.3 (95.8)	0.31 (0, 0.49)	0.57 (0.32, 0.87)
Haematological	93.3 (96.7)	0.85 (0.60, 1)	0.85 (0.60, 1)

NA: κ -statistics not appropriate for use in these systems.

absence of active urinary sediments. In the context of rising anti-dsDNA levels and/or decreasing complement C3/C4 levels, proteinuria is more likely due to LN, and hence the additional weighting in the scoring of proteinuria that is provided by serological markers. However, the changes in anti-dsDNA and/or complement levels on their own have no bearing on the scoring of the index. In addition, the scoring thresholds for plasma creatinine and haemoglobin level (in grades B and C) have been adjusted to reflect the effect of haemodilution from pregnancy.

The reliability of this index was assessed using several measures in similar fashion to the reliability study of the BILAG-2004 index [1]. However, few patients recruited to this study had active disease, and those who did have active disease typically had only grade B score and none had grade A activity as assessed by both raters. This is in keeping with other recent studies of SLE disease activity in pregnancy and reflects current management of pregnant SLE patients, which starts preconception. SLE patients are advised to get pregnant after the disease is quiescent for at least 6 months to minimize the risk of flare of disease during pregnancy. Furthermore, pregnant SLE patients are monitored closely and any evidence of active disease will be treated early to reduce the extent of disease activity flare. Therefore current management strategies of SLE patients are designed to reduce the likelihood of active disease occurring in pregnancy. As such, we found clustering of patients with no activity in constitutional, neuropsychiatric, gastrointestinal and ophthalmic systems. As a result, κ -statistics were not useful in the assessment of reliability in this study, and there were a small number of disagreements in scores between raters.

A potential limitation of this study is that the index has not been tested through its entire range of possible disease activity. No major issue was identified with regards to the face and content validity of this index in this study. Therefore, as the BILAG2004-Pregnancy index is very similar to the BILAG-2004 index, which has been comprehensively validated, we are confident that this index is suitable for use to assess disease activity in pregnant SLE patients. Further validation of this index with regards to construct/criterion validity and sensitivity to change is required.

Rheumatology key message

- The BILAG2004-Pregnancy index is reliable for assessment of SLE disease activity in pregnancy.

Acknowledgements

The authors would like to thank Professor Mark Kilby, Dr Tracey Johnston, Dr Fiona Fairlie, Dr Maggie Blott, Dr David Williams. V.T.F. was supported by MRC funding (MRC MC_US_A030_0022).

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

Disclosure statement: C.-S.Y. has received consultancy payments and honoraria from Parexel, Genentech and Teva Pharmaceuticals. All other authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* Online.

References

- 1 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.
- 2 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;56:4113–9.
- 3 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.
- 4 US Food and Drug Administration. Guidance for industry: systemic lupus erythematosus—developing medical products for treatment. Silver Spring, MD: US Food and Drug Administration, 2010. Available online at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072063.pdf>.
- 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 Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
- 8 Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 1968;70:213–20.