

Further validation of the BILAG disease activity index in patients with systemic lupus erythematosus

Thomas Stoll, Gerold Stucki, Javid Malik, Stephen Pyke, David A Isenberg

Abstract

Objective—To examine the association among the BILAG disease activity index components and their relations with global assessments, health status, and laboratory tests with regard to the validity of the BILAG index.

Methods—A cross sectional study of consecutive patients with systemic lupus erythematosus (SLE) attending a specialist lupus outpatient clinic between July 1994 and February 1995. The internal consistency of the British Isles Lupus Assessment Group (BILAG) index—a disease activity assessment system for SLE patients, based on the principle of the physician's intention to treat—was examined using Cronbach's coefficient α . The association of the components of the BILAG index with health status as measured with the MOS Short Form 20 (SF-20), with patients' and doctors' global assessments of patient wellbeing and with laboratory tests was analysed with Spearman rank correlations.

Results—133 female and eight male patients, age 20.1 to 88.7 years (mean 41.1, SD 12.5), were included. With few exceptions, the components of the BILAG index which reflect disease activity in different organ systems were not associated with each other. With the exception of the mucocutaneous component, we found a significant relation between all components of BILAG and global assessment of patient wellbeing, health status, erythrocyte sedimentation rate, or serum C3 level.

Conclusions—The study confirms the validity of all but the mucocutaneous component of the BILAG index. However, disease activity in different organ systems in SLE does not follow a common pattern. Thus the individual BILAG components should be used rather than the total BILAG score as a primary endpoint in clinical and epidemiological studies. To capture the total effect of SLE on an individual measures of disease activity, damage, and health status are all needed.

both validated and reliable.¹ Examples include the SLAM (systemic lupus activity measurement) and SLEDAI (systemic lupus erythematosus disease activity index).¹ However, these are global scores and only one system—BILAG (British Isles lupus assessment group)—has been developed with the notion of reporting disease activity in the various organ systems.² It is based upon the principle of the "physician's intention to treat" and was determined by a group of rheumatologists agreeing the type of drug treatment they were likely to use across the spectrum of lupus clinical features.² BILAG, first reported in 1988,³ has been shown to be valid and reliable in two studies.² Liang *et al* found good inter-rater reliability as well as a convergent validity of the total score with other disease activity indices.¹ Hay *et al* showed that BILAG had good between-rater reliability for each organ based component.² Criterion validity was demonstrated by the finding that the start or increase of a disease modifying drug treatment was associated with most active disease in each organ based component except for the neurological component. There was also construct validity to laboratory tests, such as anti-dsDNA antibody levels and erythrocyte sedimentation rate (ESR) and the patients' treatment.² The objective of this study was to examine the validity of the BILAG system, the associations among its components, and their relations with patients' and doctors' global assessment, and with health status and laboratory tests. Moreover, we aimed to validate a previously suggested weighted total BILAG score, with A = 9 (most active disease), B = 3 (intermediate activity), C = 1 (mild, stable disease activity), D = 0 (inactive disease), and E = 0 (no activity ever), resulting in a potential range from 0 to 72 points (72 = most active disease).^{1,2,4}

Global assessment by both the doctor and the patient is an accepted external standard as there is no gold standard to measure disease activity in SLE. A significant relation between BILAG components and a global assessment would be a sign of validity. Based on the known impact of disease activity on the patient—that is, on physical, psychological, and social health—we expected a significant association between BILAG components and a general health status measure such as SF-20. However, since disease activity and health status measures are distinct, we expected this association to be moderate. A significant association of BILAG

Bloomsbury
Rheumatology
Unit/Division of
Rheumatology,
Department of
Medicine, University
College of London,
London W1P 9PG,
United Kingdom
D A Isenberg
J Malik
T Stoll

Rheumaklinik und
Institut für
Physikalische Medizin,
Universitätsspital,
CH-8091 Zürich,
Switzerland
G Stucki

Medical Statistics
Unit, London School of
Hygiene and Tropical
Medicine, London
University, London
WC1E 7HT, United
Kingdom
S Pyke

Correspondence to:
Th Stoll MD, Rheumatology
Unit, Department of Internal
Medicine, Stadtspital Waid,
Tièchestr 99, CH-8037
Zürich, Switzerland.

Accepted for publication
20 June 1996

(Ann Rheum Dis 1996;55:756-760)

Disease activity in patients with systemic lupus erythematosus (SLE) can now be assessed by several indices which have been shown to be

components with ESR or renal activity with C3 or dsDNA antibody levels would show validity, as these laboratory tests have been found to be associated with general disease activity⁵ or renal activity.⁶

Methods

STUDY DESIGN

A prospective cross sectional study of SLE patients attending a specialist lupus outpatient between July 1994 and February 1995 was undertaken.

DATA COLLECTION

Each SLE patient was assessed at the first outpatient consultation after July 1994. The date at which a patient fulfilled the fourth ARA SLE classification criterion⁷ was recorded as the date of diagnosis. Antinuclear antibodies (ANA) were determined using Hep2 cells, regarded as positive with a titre > 1:80, and antibodies against dsDNA by ELISA (Cambridge Life Sciences; normal range < 100 IU ml⁻¹). Disease activity was recorded using the BILAG system.² At the end of the consultation, patients and doctors assessed the patient's global wellbeing during the last four weeks. Shortly before or after the consultation, patients completed the MOS short form 20 with an additional question for fatigue.^{8,9}

MEASURES

BILAG

BILAG measures SLE disease activity in eight organ based components (general, mucocutaneous, neurological, musculoskeletal, cardiovascular/respiratory, vasculitis, renal, and haematological), and is based upon the principle of the physician's intention to treat. To obtain a global score, BILAG component scores can be assigned numerical values: A = 9 (most active disease), B = 3 (intermediate activity), C = 1 (mild, stable disease activity), D=0 (inactive disease), and E = 0 (no activity ever), resulting in a potential summed range of from 0 to 72 points (72 = most active disease).^{1,2,4} A simpler way of obtaining an overall score is to add up the number of organ based components, with A = most active disease (number of As, range 0-8), or the number of components with either an A or a B (=intermediate SLE activity) (number of As+B's, range 0-8). BILAG includes the following laboratory tests: haemoglobin, white blood cell, neutrophil, lymphocyte and platelet count, creatinine, and Coombs test. Proteinuria was measured by a dipstick method (negative = 0, + = 1, ++ = 2, +++ = 3, ++++ = 4), and where appropriate 24 h urine protein and/or EDTA clearance were recorded.

Global assessments

Patients' global assessment of own wellbeing during the last four weeks was recorded on a visual analogue scale (range 0-100 mm, 0 mm = completely well, 100 mm = very ill). Doctors' global assessment of patient wellbeing during the last four weeks was registered by numeric rating (range 0-10, 0 = completely well and 10 = very ill).

MOS Short Form 20 with an additional question (SF-20+)

To the 20 questions of SF-20 relating to six domains of patient general health status ("physical functioning", "role functioning", "social functioning", "mental health", "health perception", and "pain") we added a further question enquiring about recent levels of fatigue ("Have you been easily fatigued during the past month?"; numeric rating scale 0-10). The scale scores for "physical functioning", "role functioning", "social functioning", "mental health", and "health perception" of SF-20 were scored so that higher values indicated better health (range 0-100).⁸ Scores for "pain"^{8,9} and "fatigue" were scaled to have values in the range from 0 to 100, with 0 assigned to "no pain" or "not fatigued at all" and 100 assigned to "severe pain" or "yes completely fatigued".

Blood sampling

Blood was taken to determine ESR, C3 (measured by laser nephelometry, normal range 0.75 to 1.75 g litre⁻¹), and antibodies against dsDNA. These laboratory tests are not part of the BILAG system and considered external standards for disease activity.^{5,6}

STATISTICAL ANALYSES

Associations within the BILAG components were examined using Spearman rank correlations and Cronbach's coefficient α .¹⁰ Associations between the BILAG components and global assessments, health status as measured with the SF-20, and laboratory tests were analysed with the Spearman rank correlation coefficient (ρ). To account for the effect of multiple comparisons in these exploratory analyses only differences at $P < 0.01$ were considered significant.

Results

PATIENTS AND DISEASE CHARACTERISTICS

One hundred and forty one SLE patients were included in the study; 133 (94%) were female, 97 Caucasians, 16 Afro-Caribbeans, 22 Asians, and six of mixed ethnic origin. Patients ranged in age from 20.1 to 88.7 years (mean 41.1 years, SD 12.5). The disease duration ranged from 0.1 to 32 years (mean 10.2 years, SD 6.3). Twenty one (15%) of the 141 patients also had Sjögren syndrome, six (4%) had myositis, and in eight (6%) SLE overlapped with rheumatoid arthritis. According to the definition of Stewart *et al*,⁷ 44% or more of our SLE patients were in poor health in each of the following domains of SF20+: "physical functioning", "role functioning", "mental health", "health perception", and "pain". Positive ANA were found in 99% of the patients, raised anti-dsDNA antibodies in 57%, and 27% of the patients were rheumatoid factor positive. Table 1, A and B, shows the disease characteristics of the patients.

ASSOCIATION AMONG BILAG COMPONENTS

The associations among the BILAG components are shown in table 2. There were very few significant associations: notably between the

Table 1a Disease activity of the 141 SLE patients, as measured with BILAG

	Score (number in category)							
	General	Mucocutaneous	Neurological	Musculoskeletal	Cardiovascular/ respiratory	Vasculitis	Renal	Haematological
A ¹	1	1	1	7	0	0	0	1
B ²	17	19	1	18	4	6	12	21
C	85	28	30	58	25	40	10	72
D and E	38	93	109	58	112	95	119	47

¹ Only three patients suffering most active disease (A) in one organ based component had no simultaneous intermediate disease activity (B) in another organ based component; six had simultaneously 1 A and 1 B; two had 1 A and 2 B scores.

² 20 patients had simultaneously B scores in two organ based components of BILAG, four had B scores in three, and one patient had B scores in four. A = most active disease; B = intermediate disease activity; C = mild, stable disease activity; D = inactive disease; E = no activity ever.

Table 1b Disease characteristics of the 141 SLE patients

Variable (observed range)	First quartile; median; third quartile; for laboratory tests: mean (SD)
BILAG	
Total (0-17)	3; 4; 7 (96% ¹)
Global assessment	
Patients' (0-95mm)	10; 29; 52
Doctors' (0-7)	1; 2; 3
SF-20+	
Physical functioning (0-100)	33.3; 66.7; 83.3
Role functioning (0-100)	0; 50; 100
Social functioning (0-100)	60; 80; 100
Mental health (16-100)	48; 68; 84
Health perception (0-100)	25; 42.2; 63.2
Pain (0-100)	25; 50; 75
Fatigue (0-100)	40; 70; 80
Laboratory tests	
Haemoglobin (g d ⁻¹)	12.43 (1.49)
ESR (mm h ⁻¹)	29 (24)
C3 (g l ⁻¹)	0.91 (0.25)
Creatinine (μmol l ⁻¹)	109 (124)
Antibodies against dsDNA (IU ml ⁻¹)	63 (157)

¹ Percentage of patients with a score > 0.

general and the musculoskeletal score ($\rho = 0.25$), the general and the vasculitis component ($\rho = 0.23$), and between the vasculitis and the partially related cardiovascular/respiratory component ($\rho = 0.27$). None of these could be considered very strong relations. The Cronbach coefficient α (0.35) was well below that normally considered acceptable ($\alpha = 0.65$) for reliable group comparisons.¹¹ The association of BILAG components to BILAG total was stronger than to the BILAG: number of As, or the BILAG: number of As+Bs, except for the renal item.

ASSOCIATION BETWEEN BILAG COMPONENTS AND GLOBAL ASSESSMENTS (table 3)

BILAG general and total scores were most strongly associated ($\rho = 0.4$ to 0.5) with both patients' and doctors' global assessment of patient wellbeing. Somewhat weaker associations were found between these global assessments and the BILAG neurological, musculoskeletal, and vasculitis components

($\rho = 0.2$ to 0.27). Both mucocutaneous and cardiovascular/respiratory scores were more strongly related to doctors' than to patients' global assessments, though in each case the association failed to reach formal significance (at the level of $P < 0.01$). BILAG total was linked more strongly to both global assessments than BILAG: number of As or BILAG: number of As+Bs.

ASSOCIATION BETWEEN BILAG COMPONENTS AND HEALTH STATUS (table 3)

The musculoskeletal component of BILAG was associated significantly with "physical functioning", "pain", and "health perception" of SF-20+. The general component of BILAG showed a significant relation to general health status domains except for "physical" and "role functioning". The associations to "mental health" and "fatigue" ($\rho > 0.4$) were most strong. The neurological BILAG score correlated significantly with "social functioning" of SF20+. However, most of these significant associations (at the level $P < 0.01$) were not very strong ($\rho < 0.3$). BILAG total was linked more strongly to the different health status domains than BILAG: number of As or BILAG: number of As+Bs. Table 4 shows the link between the individual domains of SF-20+ providing some help in interpreting the data of table 3. For instance, "fatigue" and "mental health" are both related strongly to the general component of BILAG (table 3). Table 4 reveals that these associations were to be expected, as "fatigue" is most closely to the "mental health" domain of SF-20+.

ASSOCIATION BETWEEN BILAG COMPONENTS AND LABORATORY TESTS

Significant associations were found between BILAG and certain laboratory tests which are not part of the BILAG score. BILAG haematological, BILAG total, and BILAG:

Table 2 Association within BILAG activity index, shown as Spearman rank correlation coefficients (n=141)

Item	Mucocutaneous	Neurological	Musculoskeletal	Cardiovascular/ respiratory	Vasculitis	Renal	Haematological	Number ^o of A scores	Number ^o of A+B scores	Total ^F
General	0.11	0.21	0.25*	0.21	0.23*	-0.15	0.00	0.12	0.33	0.47
Mucocutaneous		-0.05	0.12	0.00	0.05	0.10	0.15	0.21	0.41	0.48
Neurological			0.15	0.09	0.14	0.05	0.00	0.23	0.21	0.34
Musculoskeletal				0.13	0.10	-0.04	0.02	0.32	0.37	0.51
Cardiovascular/ respiratory					0.27*	0.02	-0.07	-0.02	0.16	0.31
Vasculitis						0.03	0.00	0.09	0.23	0.40
Renal							0.17	-0.05	0.31	0.30
Haematological								0.18	0.37	0.39

* $P < 0.01$.

^o = No significance level for correlations between organ based components and total scores are given, because the former are part of the total scores and therefore a correlation between the sum and the component was to be expected.

Table 3 Association between BILAG and global assessments and health status as measured by SF-20+, shown as Spearman rank correlation coefficients

BILAG	General	Mucocutaneous	Neurological	Musculoskeletal	Cardiovascular + respiratory	Vasculitis	Renal	Haematological	Number of A scores	Number of A+B scores	Total
Patient's global assessment	0.50*	0.15	0.23*	0.27*	0.13	0.23*	-0.08	-0.02	0.25*	0.26*	0.40*
Doctor's global assessment	0.43*	0.21	0.24*	0.27*	0.21	0.20	-0.02	0.01	0.25*	0.37*	0.47*
SF-20+: Physical functioning ^o	-0.13	-0.01	-0.11	-0.23*	-0.11	0.03	0.00	-0.06	-0.15	-0.21	-0.25*
SF-20+: Role functioning ^o	-0.15	0.11	-0.20	-0.22	-0.21	0.03	0.02	-0.05	-0.11	-0.15	-0.21
SF-20+: Social functioning ^o	-0.23*	0.01	-0.27*	-0.22	-0.10	-0.06	0.09	-0.04	-0.20	-0.13	-0.24*
SF-20+: Mental health ^o	-0.41*	-0.13	-0.20	-0.13	-0.09	-0.14	0.17	-0.01	-0.18	-0.17	-0.24*
SF-20+: Health perception ^o	-0.38*	-0.09	-0.06	-0.29*	-0.09	-0.03	0.10	-0.07	-0.17	-0.15	-0.27*
SF-20+: Pain	0.27*	0.18	0.12	0.47*	0.25*	0.17	-0.15	0.00	0.17	0.20	0.34*
SF-20+: Fatigue	0.51*	0.09	0.11	0.16	0.13	0.10	-0.11	-0.10	0.04	0.12	0.23*

* P < 0.01.

^o Scale of item runs into the opposite direction from the item with which it is compared.

Table 4 Association between the individual domains of patients' self assessment using SF-20+, shown as Spearman rank correlation coefficients (n=126 to 129)

Item	Physical functioning	Role functioning	Social functioning	Mental health	Health perception	Pain
Role functioning	0.75*					
Social functioning	0.46*	0.50*				
Mental health	0.12	0.14	0.31*			
Health perception	0.41*	0.47*	0.49*	0.37*		
Pain ^o	-0.37*	-0.38*	-0.47*	-0.22	-0.60*	
Fatigue ^o	-0.22	-0.25*	-0.32*	-0.42*	-0.61*	0.44*

* P < 0.01.

^o Scale of items runs into the opposite direction from the item with which it is compared.

number of As+Bs were linked to the ESR ($\rho = 0.34$, $P < 0.01$; $\rho = 0.26$, $P < 0.01$; and $\rho = 0.24$, $P < 0.01$), and BILAG renal and serum C3 levels were also associated ($\rho = -0.26$, $P < 0.01$). The highest correlation coefficient between dsDNA levels and a BILAG component, namely haematological BILAG, was 0.14 ($P > 0.05$). The correlations between the mucocutaneous BILAG and ESR or C3 were not statistically significant ($\rho = 0.10$ or $\rho = 0.02$).

Discussion

Our study population has been shown to be comparable to other SLE populations with respect to disease activity when assessed by the BILAG index. The percentage of activity grade A (most active disease in one organ based item) per number of assessments (7.8%) is nearly identical to the rate of 8.2% previously reported.²

Hitherto, only one report has shown validity for the organ based components of BILAG.² Our study is the first to assess the degree of association between BILAG and global assessments of patient wellbeing and health status as measured with SF-20+ and to investigate validity in this way. Thus we have found validity for the general, neurological, musculoskeletal, cardiovascular/respiratory, and vasculitis BILAG component as well as for the total BILAG score, as global assessments are accepted external standards and disease activity is well known to have an impact on general health status. The haematological and renal components, respectively, showed a significant relation to ESR, and an inverse rela-

tion to serum C3 concentrations, neither of which is part of the related BILAG item. These findings suggest validity for these two BILAG components, as C3 levels are known to correlate with renal disease activity⁶ and ESR with active SLE in a group comparison.^{5, 12} We could not confirm the validity of the mucocutaneous BILAG score found by Hay *et al.*² However, the significance level for the correlation between the mucocutaneous BILAG and doctors' global assessment of patient wellbeing ($\rho = 0.21$) was between $P < 0.05$ and $P < 0.01$.

Tables 2 and 3 suggest that the total BILAG score may be superior to a BILAG assessment adding up the number of all organ based components with an A (= most active disease) or with an A and B (most active disease and intermediate disease activity), as the associations of BILAG total score to BILAG components, global assessments, health status domains, and laboratory tests were stronger than overall disease activity assessment obtained by adding up the number of A or A+B ratings. However, the present study suggests that the BILAG total score may not add much to the information revealed by the BILAG components scores. This suggestion is based upon the observation that in most cases at least one of the BILAG components correlated more strongly with a global assessment or a health status measure than did the BILAG total score. Secondly, this assertion is supported by the conclusion that disease activity in the different organ systems in SLE varies and is discordant. For example SLE may be very active in one or two organ based components but quiescent in the other

organ systems. Table 1B shows that only seven patients had an intermediate or higher disease activity in three or more organ based BILAG components. Thirdly, with very few exceptions the components of the BILAG were not associated with each other, which was reflected in a low Cronbach coefficient α . Thus the total BILAG would not fulfil the criteria for an internally consistent scale. Fourthly, in another study we found an association between disease activity and damage only for individual organ systems but not across organ systems.¹³ This indicates that disease activity and consecutive damage in one organ system is not necessarily related to disease activity and cumulative damage in another organ system. When assessing disease activity in patients with SLE it therefore seems important to grade individual components separately, and using a total score may give only limited information. Also, when defining an endpoint for clinical or epidemiological studies it appears important to define as primary or secondary endpoints the relevant BILAG components. This concept is in line with those who developed the BILAG index, who did not intend originally to produce an overall score.¹⁴ For example, in a study with outcome as an important endpoint, BILAG general and the musculoskeletal disease activity would be important measures, as these BILAG components are related to all outcome domains except for "role functioning".

Disease activity was more closely and broadly related to physical, social, or psychological outcome than SLICC/ACR damage index was to outcome¹³ and therefore seems to have more impact on general health status than damage. However, most relations between BILAG components and health status domains were moderate ($\rho < 0.3$), reflecting the fact that disease activity measures are, as expected, distinct from health status measures. The apparent association between neurological disease activity and social function may shed light on the way SLE influences socioeconomic status. As SLE is known to be associated with diminished socioeconomic status,¹⁵ our findings suggest that CNS lupus may contribute to the social decline. However, this finding warrants reconfirmation, as our study is cross sectional.

In conclusion our study confirms the validity of all the components of the BILAG except for the mucocutaneous component, and is the first

to show the neurological component to be valid, as well as the previously suggested total BILAG score.⁴ However, disease activity in different organ systems in SLE does not follow one common pattern. It is suggested that the BILAG components rather than the BILAG total score be used as a primary endpoint in clinical and epidemiological studies. To capture the total effect of SLE on a patient, measures of disease activity, damage, and patient perception of general health are all required.

We wish to thank for their support: Ciba-Geigy-Jubiläum-Stiftung, CH-4002 Basel, Theodor und Ida Herzog-Egli-Stiftung, CH-8008 Zürich, and Foundation for Specialists Training of the Swiss Society of Rheumatology and Physical Medicine, CH-8038 Zürich, Switzerland.

- 1 Liang MH, Socher SA, Larson MG, Schur PH. Reliability and clinical validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus. *Arthritis Rheum* 1989;32:1107-18.
- 2 Hay EM, Bacon PA, Gordon C, Isenberg DA, Maddison P, Snaith ML, et al. The BILAG index: a reliable and valid instrument for measuring clinical disease activity in systemic lupus erythematosus. *Q J Med* 1993;86:447-58.
- 3 Symmons DPM, Coppock JS, Bacon PA, Bresnihan B, Isenberg DA, Maddison P, et al. Development of a computerised index of clinical disease activity in systemic lupus erythematosus. *Q J Med* 1988;69:927-37.
- 4 Ehrenstein MR, Longhurst CM, Latchman DS, Isenberg DA. Serological and genetic characterization of a human monoclonal immunoglobulin anti-DNA idiotype. *J Clin Invest* 1994;93:1787-97.
- 5 Fries J, Holman H. *Systemic lupus erythematosus: a clinical analysis*. Philadelphia: WB Saunders, 1975.
- 6 Lloyd W, Schur PH. Immune complexes, complement, and anti-DNA in exacerbations of systemic lupus erythematosus (SLE). *Medicine* 1981;60:208-17.
- 7 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;25:1271-7.
- 8 Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey: reliability and validity in a patient population. *Med Care* 1988;26:724-35.
- 9 Wachtel T, Piette J, Mor V, Stein M, Fleishman J, Carpenter C. Quality of life in persons with human immunodeficiency virus infection: measurement by the medical outcomes study instrument. *Ann Intern Med* 1992;116:129-37.
- 10 Cronbach L. Coefficient α and the internal structure of tests. *Psychometrika* 1951;16:297.
- 11 Nunnally JC. Assessment of reliability. In: *Psychometric theory*. New York: McGrawHill, 1978.
- 12 Morrow WJW, Isenberg DA, Todd-Pokropek A, Parry HF, Snaith ML. Useful laboratory measurements in the management of systemic lupus erythematosus. *Q J Med* 1982;51:125-38.
- 13 Stoll T, Stucki G, Malik J, Pyke S, Isenberg DA. Association of the ACR/SLICC Damage Index with measures of disease activity and health status in patients with systemic lupus erythematosus. *J Rheumatol* (in press).
- 14 Hay E. Systemic lupus erythematosus. *Baillieres Clin Rheumatol* 1995;9:437-70.
- 15 Karlson EW, Dalroy LH, Lew RA, Wright EA, Partridge AJ, Roberts WN, et al. The independence and stability of socioeconomic predictors of morbidity in systemic lupus erythematosus. *Arthritis Rheum* 1995;38:267-73.