Clinical Composite Measures of Disease Activity and Damage Used to Evaluate Patients With Systemic Lupus Erythematosus: A Systematic Literature Review

Isabel Castrejón, a,∗ Inigo Rúa-Figueroa, b María Piedad Rosario, c Loreto Carmona d

a Departamento de Medicina, División de Reumatología, New York University School of Medicine, NYU Hospital for Joint Diseases, Nueva York, United States
b Departamento de Reumatología, Hospital Dr. Negrín, Las Palmas de Gran Canaria, Spain
c Unidad de Investigación, Sociedad Española de Reumatología, Madrid, Spain
d Universidad Camilo Jose Cela, Instituto de Salud Musculoesquelética, Madrid, Spain

A R T I C L E   I N F O

Article history:
Received 9 August 2013
Accepted 31 January 2014
Available online 10 August 2014

Keywords:
Validation
Systemic lupus erythematosus
Disease activity
Damage
Systematic review

A B S T R A C T

Objectives: To determine the most appropriate indices to evaluate the disease activity and damage in patients with systemic lupus erythematosus (SLE).

Methods: A systematic literature search was performed to identify validation studies of indices used to evaluate disease activity and damage. We collected information for each instrument on every aspect of validation including feasibility, reliability, validity and sensitivity to change using ad hoc forms.

Results: A total of 38 articles were included addressing the validation of 6 composite indices to evaluate disease activity (BILAG, ECLAM, SLAM, SLEDAI, LAI and SLAQ). and 3 indices to evaluate damage (SLICC/ACE-DI, LD1Q and BILD). Only the SLAQ, LD1Q and the BILD were self-administered. Feasibility and internal consistency was only studied in 3 indices (BILAG, SLAQ and SDI) with a Cronbach α ranging from 0.35 to 0.87. The intra-observer reliability was examined by the intraclass correlation coefficient for BILAG with a result of 0.48 (95% CI: 0.23–0.81) and using analysis of variance for SLAM-R (0.78), SLEDAI (0.33) and the LAI (0.81). The inter-observer feasibility was evaluated using the correlation coefficient for ECLAM (0.90–0.93), the SLAM (0.86) and MEX-SLEDAI (0.97–0.89). The construct validity was examined by means of convergence with other instruments, specifically with global assessment by the physician, with similar results between indices (0.48–0.75). Lastly, responsiveness was tested in all indices except LAI, SLEDAI and LD1Q, with a standardized response mean ranging from 0.12 to 0.75.

Conclusions: Although multiple instruments have been validated for use in SLE it was not possible to find direct evidence of which is the most appropriate. BILAG and SLEDAI, with moderate feasibility and low responsiveness, are the 2 indices with a most complete validation and more extensively used.

© 2013 Elsevier España, S.L.U. All rights reserved.

Índices compuestos para evaluar la actividad de la enfermedad y el daño estructural en pacientes con lupus eritematoso: revisión sistemática de la literatura

R E S U M E N

Objetivos: Determinar que índices compuestos serían los más apropiados para la evaluación de la actividad y el daño estructural en pacientes con lupus eritematoso sistémico (LES).

Métodos: Se realizó una revisión sistemática para identificar estudios de validación de índices de actividad y daño en pacientes con LES. De cada instrumento se recogió información sobre cada aspecto de la validación, como la viabilidad, la fiabilidad, la validez y la sensibilidad al cambio utilizando un formulario ad hoc.
Introduction

Measuring disease activity and irreversible damage in patients with systemic lupus erythematosus (SLE) is of vital importance to evaluate patient outcomes and prognosis, differences between patient groups and responses to new treatments. There are several validated indices, available since the early 1980s, although none have shown a clear superiority and therefore do not have universal acceptance.

Assessing patients with SLE is challenging for several reasons. On the one hand, the complex multisystem involvement caused by the disease and, on the other, its fluctuating course, which makes patients have very different evolutions and affections. Furthermore, the absence of a gold standard or a single variable to assess the disease activity makes the use of composite indices or scales necessary. The physicians global assessment (PGA) can be strongly influenced by the clinical experience of the physician and therefore produce a wide interobserver variability complicating comparison between patients.  

The objective of this review is to evaluate aspects of the validation of the indices used to measure both disease activity and accumulated damage in patients with SLE.

Methods

As part of the consensus of the Spanish Society of Rheumatology for the use of biological therapies in patients with SLE, a systematic review was conducted to examine the validity of the composite indexes used to assess disease activity and cumulative damage.

Search Strategy

A search strategy was designed using the following bibliographic databases: MEDLINE, EMBASE and Cochrane Central Library up until March 2012. The search included Mesh terms and free text. The search strategies for MEDLINE, EMBASE and the Cochrane library were included in the Appendix (available on the web). The search was limited to human studies and studies published in English, French and Spanish. In addition, a manual search for the included articles was conducted in the bibliography.

The search was limited to human studies and studies published in English, French.

Selection of Studies

Validation studies, cohort studies, meta-analyses and systematic reviews were included. Regarding the type of participants, studies were selected with patient age ≥ 18 and diagnosed with SLE according to the criteria of the American College of Rheumatology (ACR). An intervention was considered as any index, instrument or scale used to assess disease activity and structural damage. The comparator was any index compared with itself or with other indices. Finally, the outcome measures were evaluated aspects of validation and sensitivity, specificity, feasibility, reliability, validity and sensitivity to change. More detailed information on aspects of validation, for an easy interpretation of the results, is given in Table 1.

One reviewer (CI) selected the articles by title and abstract, following the inclusion criteria. Table 2 shows the characteristics of the included studies. Information concerning the validation of each composite index was collected using standardized forms. Evidence tables were generated to summarize every aspect of validation as well as the method used for evaluation.

Results

Fig. 1 details the search results. In the search strategy, a total of 704 articles, 50 were selected for a detailed review after exclusion of 519 references by title and abstract. A total of 38 articles were included and 11 were excluded because they did not meet the inclusion criteria. Table 3 details the reasons for the exclusion of studies.

The included articles address the validation of 6 composite indices of disease activity: British Isles Lupus Assessment Group (BILAG), European Consensus Lupus Activity Measurement (ECLAM), Systemic Lupus Activity Index (SLAM), Systemic Lupus Erythematosus Disease Activity Measure (SLEDAI), UCSF/JHU Lupus Activity Index (LAI) and Systemic Lupus Activity Questionnaire (SLAQ), and 3 indices to assess cumulative damage: Systemic Lupus International Collaborating Clinics/American College of Rheumatology-Damage Index (SLICC/ACR-DI), Lupus Damage Index Questionnaire (LDIQ) and Brief Index of Lupus Damage (BILD).

The characteristics and validation of each index were then reviewed in detail. Table 4 shows aspects of validation of these indices to allow for easy comparison.

Indices to Assess Disease Activity

British Isles Lupus Assessment Group

The British Isles for the Assessment of Lupus Group (BILAG) began to meet regularly in 1984 and in 1988 developed for the first time an index to measure disease activity in patients with SLE. This index was developed based on the physicians intention to...
treat and evaluates specific manifestations requiring treatment in a total of 8 organs or systems: general, mucocutaneous, neurological, musculoskeletal, cardiorespiratory, vasculitis, renal and hematologic. Unlike other indices which provide a global assessment of the disease, BILAG provides an assessment by organ system.

Subsequently, certain minor modifications of this index were conducted and assessed for both their reliability and validity. A study of Hay et al., evaluated the BILAG version 3, which includes a glossary with explanations and recommendations for ease of use. BILAG showed good intraobserver reliability, with kappa ranging from 0.79 to 0.97, depending on the target system. To assess the criterion validity of the BILAG, it was compared to a gold standard defined as starting or increasing treatment with either prednisone or other immunosuppressive therapy. BILAG sensitivity for grade A in any organ or system was 87% and specificity of 99%. BILAG's Cronbach coefficient \( \alpha \) calculated to assess the internal consistency and, therefore, the association between its components, is 0.35, lower than what is recommended for reliable comparisons. The BILAG showed moderate correlation with the physician global assessment (\( \rho = 0.43 \)) and patient assessment (\( \rho = 0.50 \)).

The BILAG was also used to evaluate the occurrence of relapses in patients with lupus. The presence of a severe flare is defined as a score of A, new appearance, and a moderate flare is defined with a rating of B, if it was previously in D or E. A study published in 2003 by Gordon et al. assessed the degree of agreement between the definition of flare by BILAG and the real attitude about treatment by rheumatologists in routine clinical practice. 792% of patients with a severe flare (A) received intensive treatment by their doctor and only 41% received it in case of a moderate flare (B). Not only BILAG is useful for evaluating the presence of a lupus flare. In a study by Isenberg et al., 3 methods to assess flares were compared; BILAG 2004 index flair index generated for the Safety of Estrogens in Lupus Erythematosus National Assessment (SELENA) called the SELENA. SLEDAI, the Flare Index (SFI) and the VGM: The BILAG 2004 had the highest ICC (0.54 [95% CI: 0.32–0.78]) compared to the SFI (0.21 [95% CI: 0.08–0.48]) and VGM (0.18 [95% CI: 0.06–0.45]).

The validity of the BILAG has been evaluated in numerous studies. In a 2010 study construct validity was assessed with the hypothesis that patients with higher scores on the BILAG presented higher levels of erythrocyte sedimentation rate (ESR), anti-dsDNA titers and SLEDAI scores, and conversely, presented lower levels of complement. The association between these variables was expected, with an OR=2.6 (1.2–4.3) for ESR>60 mm/h, OR=2.5 (1.4–3.6) for anti-dsDNA greater than 5 times the normal value, OR=4.8 (1.4–15.1) for C3 below half the normal value, OR=4.1 (2.3–5.8) for C4 below half the normal value and OR=215.6 (99.8–387.6) for a value of SLEDAI above 6.

Only one study has been specifically designed to evaluate the sensitivity to change of BILAG. It was a longitudinal multicenter study where the relationship between the change in BILAG 2004 and the change in treatment between 2 consecutive visits was evaluated. BILAG increase was associated with an increase in treatment (coefficient multinomial logistic regression: 1.35, 95% CI: 1.01–1.70). In the opposite direction, this association was also found.

The BILAG can be manually calculated, but there is a computer software known as the British Lupus Integrated Prospectve System (Blips) incorporating demographic variables and clinical information necessary to calculate not only the BILAG, but other composite indices such as SLAM, SLEDAI, and SDI the SF-36.

### Table 1
Guidelines to Interpret the Results on Aspects of Validation.

<table>
<thead>
<tr>
<th>Term</th>
<th>Aspects</th>
<th>Analysis technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viability (feasibility)</td>
<td>Time spent. Clarity of elements (simple). Accepted by patients and users</td>
<td>Pilot study (30 patients)</td>
</tr>
<tr>
<td>Degree (reliability) to which the instrument accurately measures: without error. Reliable, accurate and free of error (systematic error/bias and random error)</td>
<td>Intraobserver or test-retest reliability; measures the stability of the scores given by the same evaluator in the same subjects. Interobserver reliability or measurement error: the extent of agreement between 2 or more reviewers. Ability (validity) to measure that for which it is designed.</td>
<td>ICC (intraclass correlation coeff.): quantitative Cohen’s kappa: qualitative. Standard error of measurement, minimum detectable change agreement limits. Expert opinion on the relevance and understandability: redaction of questions. Structural: factor analysis. Test of Ho: correlations. Cross-cultural validation. Gold standard. Continuous variable: correlations and ROC curves. Dichotomous variables: sensitivity and specificity. Multiple methods.</td>
</tr>
<tr>
<td>Ability to change (responsiveness)</td>
<td>Ability to measure changes. Intrinsic and extrinsic</td>
<td></td>
</tr>
</tbody>
</table>

---

**Fig. 1.** Articles retrieved by the search strategy and results of the selection.
### Table 2
Characteristics of Studies Included.

<table>
<thead>
<tr>
<th>First author and year</th>
<th>Type of study and participants</th>
<th>Index</th>
<th>Evaluated aspects of validation</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symmons 1988</td>
<td>Multicenter study (5 centers in the UK and Ireland) on symptoms that require treatment</td>
<td>BILAG</td>
<td>Face validity; the score was compared with actual treatment decisions</td>
<td>Description and preliminary validation</td>
</tr>
<tr>
<td>Hay 1993</td>
<td>Multicenter study with 82 patients</td>
<td>BILAG version 3 (explanatory glossary)</td>
<td>Reliability; criterion validity (GS = treatment decision); construct validity (ESR, dsDNA)</td>
<td>Most patients inactive or less active</td>
</tr>
<tr>
<td>Stoll 1996</td>
<td>Cross-sectional study of 141 patients</td>
<td>BILAG</td>
<td>Internal consistency and construct validity</td>
<td>BILAG vs PGA</td>
</tr>
<tr>
<td>Isenberg 2000</td>
<td>Narrative description 250 SLE patients in routine office care</td>
<td>BILAG Software</td>
<td>Usage of the British Lupus Integrated Program</td>
<td>Presentation of software</td>
</tr>
<tr>
<td>Gordon 2003</td>
<td>2 batches of exercises with real patients (8 patients/8 doctors)</td>
<td>BILAG to define &quot;lupus flare&quot;</td>
<td>Reliability: ICC among physicians for organ/system and degree of agreement between doctors</td>
<td>Good validity for assessing severe flare</td>
</tr>
<tr>
<td>Isenberg 2005</td>
<td>Multicenter study with 569 patients and 14 rheumatologists</td>
<td>BILAG 2004 update from previous versions</td>
<td>Reliability: ICC among physicians for organ/system and degree of agreement between doctors</td>
<td>All evaluators were members of BILAG</td>
</tr>
<tr>
<td>Yee 2001</td>
<td>Multicenter cross-sectional study of 369 patients</td>
<td>BILAG 2004</td>
<td>Criterion validity and construct validity</td>
<td>GS: Treatment change</td>
</tr>
<tr>
<td>Cresswell 2009</td>
<td>Multicenter cross-sectional study of 369 patients</td>
<td>Numerical scale for the BILAG</td>
<td>Comparison between 3 indexes using ROC curves</td>
<td>Cross-sectional study: ignores previous activity</td>
</tr>
<tr>
<td>Yee 2009</td>
<td>Multicenter longitudinal study</td>
<td>BILAG 2004</td>
<td>Sensitivity to change</td>
<td>Difficult to differentiate activity and damage</td>
</tr>
<tr>
<td>Nasiri 2010</td>
<td>Cross-sectional study of 100 patients</td>
<td>BILAG</td>
<td>Construct validity Hypothesis: if ↑ BILAG then ESR + dsDNA and complement</td>
<td>Some patients with SLE have a normal serological profile</td>
</tr>
<tr>
<td>Vitali 1992</td>
<td>Multicenter study with 704 patients</td>
<td>ECLAM</td>
<td>Face and content validity</td>
<td>Initial Validation</td>
</tr>
<tr>
<td>Bencivelli 1992</td>
<td>Multicenter study of 75 patients</td>
<td>ECLAM</td>
<td>Construct validity and criterion (GS: PGA)</td>
<td>The PGA showed unreliability</td>
</tr>
<tr>
<td>Vitali 1999</td>
<td>Comparison: manual/computerized system</td>
<td>ECLAM</td>
<td>Criterion validity</td>
<td>Correlations between the 2 systems</td>
</tr>
<tr>
<td>Fly 2000</td>
<td>Review of 64 medical records</td>
<td>ECLAM</td>
<td>Direct vs calculated ECLAM by reviewing medical records</td>
<td>Doctors knew they were to be reviewed</td>
</tr>
<tr>
<td>Petri 1992</td>
<td>150 patients with SLE. Subgroup of 6 evaluated by 9 doctors (reliability)</td>
<td>UCSF/JHU LAI</td>
<td>Reliability and construct validity (PGA)</td>
<td>Cross-sectional study</td>
</tr>
<tr>
<td>Cirenna 1992</td>
<td>574 cases evaluated by 14 rheumatologists</td>
<td>SLEDAI</td>
<td>Initial development of the index, reliability and face validity</td>
<td>Initial Validation</td>
</tr>
<tr>
<td>Guzman 1992</td>
<td>39 patients in a center (3 views)</td>
<td>MEX-SLEDAI</td>
<td>Comparison with SLEDAI. Reliability and construct validity</td>
<td>No laboratory variables, lower cost</td>
</tr>
<tr>
<td>Chang 2002</td>
<td>Post hoc analysis of a multicenter RCT</td>
<td>SLEDAI and SLAM-R</td>
<td>Sensitivity to change</td>
<td>Relevant changes for doctor and patient</td>
</tr>
<tr>
<td>Gladman 2002</td>
<td>Cohort of 960 patients</td>
<td>SLEDAI-2K</td>
<td>Construct validity</td>
<td>Reference: SLEDAI</td>
</tr>
<tr>
<td>Uribe 2004</td>
<td>93 patients (3 hospitals)</td>
<td>SLAM-R, MEX-SLEDAI and modified SLEDAI</td>
<td>Construct validity and criterion (GS = SLEDAI-2K)</td>
<td>V. construct Ref: PGA (↑ variability)</td>
</tr>
<tr>
<td>Touma 2011</td>
<td>Cross-sectional study of 298 patients</td>
<td>Answer 50% of the SLEDAI</td>
<td>Face validity, content and construct reliability and construct validity</td>
<td>Cross-sectional study</td>
</tr>
<tr>
<td>Bae 2001</td>
<td>30 patients evaluated by 2 physicians/2 visits</td>
<td>SLAM-R</td>
<td>Construct validity</td>
<td>Physicians inexperienced with SLAM-R</td>
</tr>
<tr>
<td>2003 Carlson</td>
<td>93 patients evaluated in a SLE clinic</td>
<td>SLAQ</td>
<td>Asymptomatic patients with abnormal lab.</td>
<td></td>
</tr>
<tr>
<td>Yazdany 2008</td>
<td>Observational cohort</td>
<td>SLAQ</td>
<td>Reliability, construct validity and sensitivity to change</td>
<td>Highly educated cohort vs general population</td>
</tr>
<tr>
<td>Gladman 1996</td>
<td>42 cases of 19 physicians</td>
<td>SDI</td>
<td>Face and content validity</td>
<td>Initial Validation</td>
</tr>
<tr>
<td>Gladman 1997</td>
<td>10 patients for 6–10 physicians from 5 countries</td>
<td>SDI</td>
<td>Interobserver reliability</td>
<td>Variability between observers through ANOVA</td>
</tr>
<tr>
<td>Stoll 1997</td>
<td>141 patients in a single center</td>
<td>SDI</td>
<td>Internal consistency and construct validity</td>
<td>Compare SDI BILAG score and medication</td>
</tr>
<tr>
<td>Gladman 2000</td>
<td>Multicenter study in 1297 patients</td>
<td>SDI</td>
<td>Apparent validity</td>
<td>Evaluates association with mortality</td>
</tr>
<tr>
<td>Costenbader 2010</td>
<td>Multicenter study with 569 patients and 14 rheumatologists</td>
<td>LDQ</td>
<td>Face, content, criterion and construct validity</td>
<td>GS: Response of doctors in the SDI</td>
</tr>
</tbody>
</table>
An updated version was published in BILAG 2005 by Isenberg et al. in an attempt to improve the characteristics of the index.12 In this study, where the BILAG 2004 is described, its viability or ability to be applied by 2 exercises with real patients who were examined by 8 doctors belonging to the BILAG group was evaluated. The ICC was calculated for the total BILAG (0.48, 95% CI: 0.23–0.81) and for each of the 9 systems. The system that showed a higher ICC was the kidney (0.98, 95% CI: 0.96–0.99) and the one that showed a lower ICC was musculoskeletal (0.17, 95% CI: 0.01–0.56). Another measure used in this study to assess feasibility was the ratio between the standard deviation attributable to the physician and the same measure attributable to the patient, showing a high agreement in the mucocutaneous, nervous, renal, ophtalmic and hematological systems. In both exercises, the musculoskeletal system showed low reliability in terms of ICC and level of agreement among physicians.

The BILAG 2004 has been developed from the original index, and includes 9 systems: general, mucocutaneous, nervous, musculoskeletal, cardiovascular/respiratory, gastrointestinal, renal, ophtalmic and hematological system. It consists of several items or questions that are assessed on a scale of 0–4, where 0 represents “not present”, 1 means “improving”, 2 is “similar”, 3 is “worse” and 4 is “new event.” Responses were combined into 5 states of system activity: A, the patient is very active and, therefore, would require treatment with immunosuppressants, corticosteroids at medium or high doses (>20 mg prednisolone or equivalent) or anticoagulation at a high dose; B, certain activity and, therefore, need for treatment with moderate dose corticosteroid (<20 mg prednisolone), antimalarial, anti-depressants, or NSAIDs; C, little activity and need only for symptomatic treatment; D, no activity at the time, although present previously, and E, has never presented activity in this system.

Other attempts to validate the BILAG 2004 were conducted by Yee et al. in 2007. Criterion validity using as a gold standard changes in treatment and the construct validity of trying to corroborate the hypothesis that patients with higher values in the BILAG 2004 would have higher ESR, anti-dsDNA titers and lower values of complement was evaluated.13 BILAG 2004 values indicating high or moderate disease activity (A and B) were significantly associated with an increase in medication (OR=19.3, p<0.01). In addition, the BILAG 2004 had a sensitivity of 81%, a specificity of 91.9%, a positive predictive value of 56.8% and a negative predictive value of 93.6% compared to the change in treatment. Top BILAG 2004 scores were associated with increased ESR of 60 mm/h (OR=2.9), a rise in anti-dsDNA titers (OR=2.7), an elevation in the SLEDAI above 4 (OR=20) and halving of C3 (OR=5) and C4 (OR=4.2).

Although the BILAG is an ordinal scale that provides a general assessment of the disease, some authors have proposed its transformation into a numerical scale to facilitate statistical analysis. Initially, the following scheme was proposed: A=9 points, B=3 points, C=1 point and both D and E were equivalent to 0 points, with a range of 0–72.14 In 2009, Cresswell et al. presented a study aimed to validate a numerical score for the BILAG.15 They described three different scoring methods using logistic regression models and later the ROC curves were compared. The transformation finally selected was A=12, B=5, C=1 and D/E=0.

The BILAG is the only validated instrument that gives an idea of lupus activity in each organ at a glance instead of merging the information into an overall score.16 The clinician should be familiar with the glossary where each item is defined, and follow the guidelines and recommendations established by the BILAG group, especially the specific recommendations for use in RCTs and multicenter studies.5,17

European Consensus Lupus Activity Measurement

The ECLAM was described by the European working group consensus for measuring activity in SLE in 1992.18 It is an index
designed to measure disease activity in the last month in patients with lupus.

This index was developed from a cohort that included a significant number of real patients. In an initial analysis, the index was developed and its feasibility and validity were assessed, both apparent and in the content. To do so, those clinical and laboratory variables that best correlate with PGA as the gold standard, both on a quantitative scale (0–10) and a qualitative scale (inactive to active) were selected. By univariate analysis, a total of 15 variables able to predict disease activity were selected. Subsequently, a multivariate analysis the weight of each of these variables, which were assigned a specific score, was defined. The ECLAM not only allows to properly classify patients according to their level of disease activity, it is also easy to calculate. In a second study, construct validity was assessed, with the gold standard being the PGA and criterion validity, comparing ECLAM with other composite indices. The ECLAM correlation with SLAM, BILAG and SLEDAI was greater than 0.72. These four levels are correlated similarly to the PGA. The ECLAM includes evaluation of 10 organs and/or systems and 2 laboratory values that are ESR and complement levels. It consists of a total of 33 items that are evaluated from 0.5 to 2, depending on the type of involvement, and a combined global score ranging from 0 to 17.5.

The ECLAM is a simple index that can be used in retrospective studies, since there is a good correlation between the ECLAM calculated immediately and that calculated with data collected on clinical history (rho=0.871). Furthermore, it can be calculated by a computerized system with very similar results to the manual calculation (r=0.90–0.92).

**Systemic Lupus Activity Index**

This index measures the degree of activity globally in the last month. The first publication in which it was described appeared in 1986. Two years later, it was revised by the residents of Harvard University, who modified the section on cardiovascular events and “others” to improve clarity and reproducibility, resulting in a new version, the SLAM-R. In 2001, Bae et al. presented a study of 30 patients to evaluate the feasibility and construct validity of this new version. Reliability was estimated by analysis of variance, with 0.78 interobserver reliability and 0.61 intraobserver reliability. Regarding construct validity, Pearson correlations were calculated SLAM-R with PGA (0.87), levels of anti-dsDNA (0.51), C3 (−0.60) and C4 (−0.29).

The SLAM evaluates specific manifestations in 9 organs and includes 7 laboratory measurements. Some items are scored from 0 to 3, depending on the degree of severity, and others are evaluated only 0–1. The maximum score is 84, being part of the laboratory parameters a maximum 21 points. A score of 7 or more is considered relevant, since the patient will require a change in treatment in 50% of cases.

The SLAM is considered by some experts the less appropriate index because it includes subjective measures, such as fatigue and joint pain; however, these variables must be scored by the doctor if considered to be due to lupus activity. Training is required, since it is necessary to reach a consensus to assess the subjective aspects, especially in multicenter studies.

**Systemic Lupus Erythematosus Disease Activity Measure**

The SLEDAI is a global index that was developed by an expert group of Toronto in 1986 and described in detail by Bombardier et al. It was amended by the SLEDAI group for a study to evaluate the use of estrogen and progesterone in women with SLE and later was updated by Gladman et al. In addition, a version developed by Mexican researchers, to reduce the costs of using this index, excluded some laboratory values.

Therefore, there are currently 4 versions of this index: SLEDAI, SELENA-SLEDAI, SLEDAI 2000 and MEX-SLEDAI.

To develop the first version of SLEDAI, 24 variables that might be important factors for evaluating disease activity in SLE patients were identified. With these variables, 574 profiles were generated for potential patients and this information was provided to 14 rheumatologists considered to be SLE experts to assess the extent of disease on a scale of 0–10. Multiple regression models were used to estimate the relative importance of each of these 24 clinical variables as assessed by the experts and thus the overall index was generated.

The SLEDAI, therefore, is a global index that evaluates the activity of the disease in the last 10 days and consists of 24 items that include specific manifestations in 9 organ systems with a maximum score of 105.

In a 1992 study by Guzman et al. various aspects of the validation of MEX-SLEDAI with the original SLEDAI were compared. The correlation between 2 reviewers calculating these indices ranged between 0.87 and 0.89. The agreement on the assessment of the disease among physicians was moderate, with a kappa of 0.43 (p=0.17). Construct validity was assessed using the PGA comparator (on a scale of 0–10), with an almost identical correlation of 0.68 for both versions; the correlation between the two versions was 0.77.

SLEDAI can be used in research and clinical practice. It has proven to be a tool sensitive to change. To assess the sensitivity to change of SLEDAI compared with SLAM-R, Chang et al. conducted a secondary analysis of data from a Canadian multicentric RCT where the effectiveness of methotrexate in patients with SLE were evaluated. Changes were compared in these 2 indices with respect to changes in the physician and patient global evaluations. Regarding the assessment of physician, the average standardized control response (C-SRM) for the SLAM-R and SLEDAI was −0.47 vs −0.42 for improvement and 0.65 vs 0.66 for worsening. Regarding the assessment by the patient, the C-SRM for the SLAM-R and SLEDAI was −0.31 vs −0.18 for improvement and 0.48
Table 4  
Aspects of validation for a Disease Activity and Damage Index.

<table>
<thead>
<tr>
<th>Name of instrument</th>
<th>Viability</th>
<th>Reliability</th>
<th>Validity</th>
<th>Sensitivity to change</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>British Isles Lupus Assessment Group (BILAG) 2004</td>
<td>Accurate and complete history, 5–20 min</td>
<td>α = 0.35&lt;sup&gt;6&lt;/sup&gt;</td>
<td>kappa = 0.79–0.97&lt;sup&gt;5&lt;/sup&gt; CC = 0.48 (95% CI: 0.23–0.81)&lt;sup&gt;1,2&lt;/sup&gt; Kappa evaluation lupus flare&lt;sup&gt;8&lt;/sup&gt; BILAG = 0.54 SELENA = 0.21 PGA = 0.18</td>
<td>NE</td>
<td>NE</td>
</tr>
<tr>
<td>European Consensus Lupus Activity Measurement (ECLAM)&lt;sup&gt;15,20,41&lt;/sup&gt;</td>
<td>Needs clinical examination 5–10 minutes Simple to calculate</td>
<td>NE</td>
<td>NE</td>
<td>Coef. correlation = 0.90–0.93&lt;sup&gt;20&lt;/sup&gt;</td>
<td>NE</td>
</tr>
<tr>
<td>Systemic Lupus Erythematosus Disease Activity Measure (SLEDAI)&lt;sup&gt;24,28&lt;/sup&gt;</td>
<td>Requires examination and analysis. 10–20 min</td>
<td>NE</td>
<td>Analysis of variance coef. correlation =0.33&lt;sup&gt;37&lt;/sup&gt; Intrao. = 0.66–0.99 Intero. = 0.60–0.80&lt;sup&gt;34&lt;/sup&gt; Intero. = 0.47&lt;sup&gt;27&lt;/sup&gt;</td>
<td>NE</td>
<td>NE</td>
</tr>
<tr>
<td>Systemic Lupus Erythematosus Disease Activity Measure (SLEDAI-2K)&lt;sup&gt;26&lt;/sup&gt;</td>
<td>The same</td>
<td>NE</td>
<td>NE</td>
<td>NE</td>
<td>NE</td>
</tr>
<tr>
<td>MEX-SLEDAI&lt;sup&gt;27&lt;/sup&gt;</td>
<td>Is completed in 16.9 min</td>
<td>NE</td>
<td>NE</td>
<td>Intero. Spearman =0.87–0.89</td>
<td>NE</td>
</tr>
</tbody>
</table>
Table 4 (Continued)

<table>
<thead>
<tr>
<th>Name of instrument</th>
<th>Viability</th>
<th>Reliability</th>
<th>Validity</th>
<th>Sensitivity to change</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCSF/JHU Lupus Activity Index (LAI)</td>
<td>1 min</td>
<td>NE</td>
<td>Analysis of variance coeff. intrao. correlation=0.81</td>
<td>Intero. =0.89</td>
<td>NE</td>
</tr>
<tr>
<td>Systemic Lupus Activity Questionnaire* (SLAQ)</td>
<td>Patient completed</td>
<td>α=0.87</td>
<td>NE</td>
<td>NE</td>
<td>NE</td>
</tr>
<tr>
<td>Indices to assess cumulative damage Systemic Lupus International Collaborating Clinics Damage Index (SLICC-DI, SDI)</td>
<td>Requires examination Approx. 15 min (depends on complexity)</td>
<td>α=0.41</td>
<td>ICC=0.553</td>
<td>10 patients and 6–10 physicians from 5 countries</td>
<td>NE</td>
</tr>
<tr>
<td>Lupus Damage Index Questionnaire* (LDIQ)</td>
<td>Patient completed</td>
<td>NE</td>
<td>NE</td>
<td>NE</td>
<td>Variables included in SDI</td>
</tr>
<tr>
<td>Brief Index of Lupus Damage* (BILD)</td>
<td>Patient completed</td>
<td>NE</td>
<td>NE</td>
<td>NE</td>
<td>Variables included in SDI</td>
</tr>
</tbody>
</table>

AUC: area under the curve; ICC: intraclass correlation coefficient; RCT: randomized clinical trial; S: specificity; PE: physical examination; GS: gold standard; SEM: standard error of measurement; MH: medical history; Intero.: interobserver; Intra.: intraobserver; LA: limits of agreement; PLR+: positive likelihood ratio; NLR−: negative likelihood ratio; LS: Likert scale; MDC: minimum detectable change; NE: not evaluated; r: Pearson correlation coefficient; rho: Spearman correlation coefficient; SRM: standardized response mean; S: sensitivity; PGA: physician global assessment; PaGA: global assessment by the patient; NPV: negative predictive value; PPV: positive predictive value.

* Patient completed.
vs 0.05 for worsening. Only SLAM-R maintained 0 out of the 95% CI to detect improvement or deterioration. Both indices exhibit significant sensitivity to change for the physicians but only SLAM-R exhibits a sensitivity to change relevant for patients, possibly by the inclusion of more subjective variables assessed by the patient.

In 2002 a revised and updated version of the SLEDAI, SLEDAI 2000 (SLEDAI-2K), which scored rash, alopecia, ulcers or persistent proteinuria, not only recently introduced but also of new appearance, as occurred in the previous version. The SLEDAI-2K was validated against the original SLEDAI with an evaluation within 10 days prior to the visit and it was validated with a temporal space of 30 days for use in RCTs. The SLEDAI-2K and the previous version had a high correlation ($r=0.97$) and both indices similarly predict mortality.26

The SLEDAI-2K reflects aspects of the disease as present or absent and may not reflect partial improvement, which limits their use in RCT. For this reason, in 2011 Touma et al. developed a response rate of 50% (SRI-50) to document a 50% improvement in SLEDAI. This study evaluated the construct validity using assessment of response to treatment by a physician as an external comparator on a Likert scale (LS) that ranged from 7 (significant improvement) to 1 (much worse). A 50% improvement on this scale would be greater than or equal to a change of 6. In patients with an improvement $\geq 6$ in the LS, there was a decrease in the SRI-50 of 4.15 $\pm$ 3.01 ($P<0.0001$).

In 2004, Uribe et al., in an attempt to dig deeper into the validation of available versions of the SLEDAI, presented a study in which the construct validity of the SLAM-R, the MEX-SLEDAI and SLEDAI-2K was analyzed, using the PGA as the external criterion on a visual analog scale (VAS) from 0 to 10. Spearman correlations of these indices with the PGA ranged from 0.54 to 0.67 for the MEX-SLEDAI and SLEDAI-2K. In an attempt to assess the criterion validity using as a gold standard the SLEDAI-2K, it was found that the sensitivity of the SLAM-R was 73%, while that of the MEX-SLEDAI was 58%. The specificity was 63% for the SLAM-R and 93% for MEX-SLEDAI.

UCSF/JHU Lupus Activity Index

The LAI was initially used in studies of serious infections and renal involvement in patients with SLE. This is an index with 5 domains that reflect the disease activity in the previous 2 weeks and is completed by the physician in about a minute. As in the rest of the indices, only those manifestations attributable to SLE were evaluated. It includes PGA VAS from 0 to 3, and includes 4 VAS regarding fatigue, rash, joint and serositis, and a part that quantifies the involvement in 4 organs: neurological, renal, pulmonary, and hematologic, with a VAS ranging from 0 to 3. Moreover, this last part assigns different scores according to the need for medication and laboratory values.

In a study by Petri et al., the reliability and construct validity of this index were evaluated using SLEDAI compared to the PGA reference. The PGA had a higher correlation with LAI ($r=0.64$) than with SLEDAI ($r=0.55$). As for reliability, LAI showed greater reliability, both intra-and interobserver.

Systemic Lupus Activity Questionnaire

The SLAQ was developed as a measurement tool for the patient to be used in epidemiological studies and large patient cohorts. It was described and validated by Karlson et al. It was developed in a clinical cohort of 93 patients and compared with SLAM (excluding laboratory data), showing a good correlation ($r=0.62, P<0.001$). Each individual item correlation between patients–physicians ranged from 0.06 for the evaluation of lymphadenopathy and vasculitis and 0.7 for evaluation of Raynaud’s syndrome.

Subsequently Yazdany et al. evaluated other aspects of validation in a cohort of 982 patients. The SLAQ showed good internal consistency, with a Cronbach $\alpha=0.87$. To assess construct validity, we examined the correlation of SLAQ with PGA ($r=0.73$) and the SF-36 ($r=0.66$). The index showed a low sensitivity to change with an SRM of 0.12, but when it was compared with the change in other measures, the SLAQ score changed in the right direction.

Studies Comparing Rates

The first study that compares several composite indexes was presented in 1989 by Liang et al. In this study the reliability of the SLAM, SLEDAI and BILAG, and their construct validity is discussed; on the one hand, the correlation between each index (which ranged between 0.81 and 0.97) was calculated and, secondly, calculating the correlation of each of them with a PGA VAS 0–10 ($r=0.76–0.96$). In an attempt to assess not only the construct validity, but also the sensitivity to change, Gladman et al. designed a study in 8 patients with 3 visits per patient and evaluated by 8 rheumatologists. The correlation between these three indices ranged from 0.35 to 0.61. Regarding the sensitivity to change, only the SLEDAI was able to differentiate between visits by analysis of variance. These results are based on the average rate change for each index and therefore the clinical relevance of this change observed in each patient is difficult to assess. A few years later, Fortin et al. conducted a study to assess the sensitivity to change using different methods and using as an external criterion the change assessment made by the physician in a LS of 5 points, ranging from “much better” to “much worse.” The 3 indices were shown to be sensitive to change, SLAM performing best. The authors interpret that this small difference may be due to the SLAM allowing increasing values according to severity while the SLEDAI has a fixed score for each item.

In another study of the same year, Ward et al. evaluated the construct validity and sensitivity to change of the SLAM, SLEDAI, LAI, BILAG and ECLAM. All these indices were valid for measuring disease activity in patients with SLE. The SLAM best captured the patient’s perception ($r=0.22, P<0.0001$). The correlation between the change in each of the indices and the change in the health evaluation were in decreasing order: LAI $r=0.75$; R ECLAM $r=0.65$; R BILAG $r=0.61$; SLAM $r=0.54$ and r SLEDAI=0.52, all with $P<0.0001$. The index that showed greater sensitivity to change measured by the SRM using and using the PGA as reference was the LAI (SRM=0.74) and to a lesser extent, the SLEDAI (SRM=0.48).

In an attempt to develop an index to assess the response to treatment, in 2004, Wollastron et al. compared the change in two validated indices, the BILAG and SLEDAI, with the change in disease activity evaluated by a skilled practitioner in lupus using LS of 7 points (1=important improvement and 7=worsening with respect to baseline). The ICC of the physicians evaluation in the 4 groups of evaluators ranged between 0.25 and 0.46. The change in baseline to 3 months of SLEDAI correlated well with the change in BILAG ($r=0.75, 95\%\, CI: 0.63–0.83$). The conclusion of this study was that, while the composite indices are comparable between themselves, there exists a large variability in the assessment made by physicians, even among experienced physicians with lupus patients.

Indices to Assess Damage in Systemic Lupus Erythematosus

Systemic Lupus International Collaborating Clinics/American College of Rheumatology-Damage Index

It was developed in 1996 by an international collaboration (SLICC group) and adopted by the CAR. In this first study, the index was developed and its content validity was assessed. A list of variables that could reflect damage in patients with SLE was generated and a consensus was reached on which variables should be included in an index generated to assess irreversible damage. In another study of the same year and by the same authors, interobserver
reliability was assessed with 10 patients who rheumatologists from 5 different countries applied this index to.\textsuperscript{43} The authors concluded in this study that doctors in different countries assess damage in patients with SLE very similarly.

The SDI evaluates irreversible damage in lupus patients regardless of the cause that produces it. It includes 42 items measuring involvement in 12 domains, with a maximum score of 46 points. Each item is scored as present or absent with the possibility of scoring a 2 or 3 in case of recurring events, such as may be the case of a stroke.

The definition of damage in patients with SLE is an irreversible change in an organ or system that has occurred since the onset of SLE and is present for at least the last 6 months. At the time of diagnosis, the score should be 0 by definition.

The SDI is completed by the physician at baseline in order to enter an RCT. It has moderate internal consistency (Cronbach $\alpha$ = 0.41). In an attempt to assess construct validity, Stoll et al. compared SDI with BILAG and a patient’s medication score, both with very low correlation (0.19 BILAG and 0.33 for the medication index).\textsuperscript{14}

When SDI is completed by another physician to retrospectively review the patient’s history, it also shows good interobserver reliability.\textsuperscript{44} SDI values increase with disease progression similarly in patients from different countries and is also a tool that predicts mortality in patients with SLE.\textsuperscript{45,46}

**Lupus Damage Index Questionnaire**

The LDIQ was described by Costenbader et al. in an attempt to develop a measure of irreversible damage based on the SDI but completed by the patient to be used in clinical practice or epidemiological studies.\textsuperscript{47} This study is a first step in the development of an index, evaluating the apparent content validity with a group of 37 patients and 7 rheumatologists, with a more elaborate multicentric level validation later. To assess criterion validity, SDI is used as the gold standard, resulting in a sensitivity of 53.3%, a specificity of 94.6% and a degree of agreement between the 2 measures of 93.2%.

To assess construct validity, the correlation between LDIQ and SDI (Spearman correlation coefficient = 0.48, $P$ < 0.001) was calculated.

The LDIQ consists of 56 questions to assess each domain included in the SDI and is designed to be administered as a survey. Versions in Spanish, Portuguese and French have been validated.\textsuperscript{48}

**Brief Index of Lupus Damage**

The BILD was described in 2011 by Yazdany et al. as a measure of damage in SLE patients for use in population studies.\textsuperscript{49} Like LDIQ, it is completed by the patient but is a shorter version that includes only 28 questions. BILD criterion validity was examined through its correlation with other instrument that also measured damage, such as SDI, and can be considered as the gold standard ($r$ = 0.64, $P < 0.001$).

Construct validity was assessed by comparing the demographic and clinical characteristics of the patients divided into the 4 quartiles of BILD. Patients with higher values for the BILD were older, with longer duration of illness and greater disease activity.

**Discussion**

To measure disease activity and irreversible damage in patients with SLE, there is a need for quantitative composite measures that have an acceptable validation. Designing these composite measures is a challenge, given the multisystem involvement in SLE and its extensive variability. At present, there are multiple tools with varying degrees of validation, although none has been accepted as the only measure recommended internationally. Choosing the most appropriate measure in each case mainly depends on the context in which it will be used and the question one wants to answer in terms of disease assessment.

Broadly speaking, there are 2 types of activity indices, those which function as global measures (such as ECLAM, SLAM, SLEDAI and LAI) and those that give a specific score for each organ system (BILAG).

The global indices are considered useful to compare cohorts of patients with SLE, as they are simpler to implement and therefore more useful in population studies. They can also be used to define inclusion criteria for RCTs or response criteria for treatment. But these indices do not give information on the degree of activity in a specific organ, for which indices of organ/system would be most useful.

In terms of implementation, most of these indices are complex and require information from the clinical history and physical examination. The index takes that takes longer to implement is the BILAG, and the LAI and SLAQ are the simplest to apply. Internal consistency was assessed only in the BILAG and SLAQ, being higher in the latter. Regarding the interobserver reliability, the index showed a higher ICC was the SLEDAI (ICC = 0.79). The interobserver reliability was assessed in a very heterogeneous way across studies, making it difficult to compare the indices. This aspect of validation is not required in the indexes answered by the patient.

Construct validity was the most studied part of the validation for most indices, with similar results. It is difficult to establish a gold standard as a reference to assess disease activity. To assess criterion validity, in some studies the gold standard was change of treatment and, in others, the PGA. Use of the PGA as a gold standard has certain problems, since several studies have shown that there is a very low degree of agreement between doctors when assessing the disease activity in patients with SLE.\textsuperscript{1,27} Other studies used as gold standard another index already described in the literature, with which the new index was compared and evaluated. Most indices correlate acceptably with each other.\textsuperscript{22}

Finally, the index that shows a greater sensitivity to change is the ECLAM.

Some important aspects to consider is that the activity indexes should only rate aspects that are directly related to the disease. The SLAM and BILAG collected manifestations that appear in the last month, while others collected manifestations in the last 10 days. Some indexes include laboratory variables and it is important to note that there are patients with active SLE who have no laboratory abnormalities. For example, only 60% of patients have positive anti-dsDNA antibodies during the course of the illness.\textsuperscript{20}

A critique of these indices could be that most of them have been validated in the context of long-term cohorts instead of RCT.\textsuperscript{51} Even so, in 1998, one of the OMERACT objectives was to preliminarily define minimum variables to be included in the evaluation of patients with SLE, both for RCTs and longitudinal cohort of long-term activity, and these were illness, injury, quality of life and toxicity/adverse effects.\textsuperscript{52}

The concept of damage in SLE has become an important outcome measure, as it not only predicts mortality also functional capacity and health resource utilization.\textsuperscript{53–55} In this review, three measures of injury have been identified, one that is completed by the doctor, the SDI and 2 to be completed by the patient LDIQ and BILD. The 2 self-administered measures of damage showed a good correlation with the extent to which the physician measured was considered the gold standard.

In conclusion, most composite indices to assess disease activity in SLE patients have been validated and are comparable. Although complex, when applied in clinical practice, these indices facilitate the collection of relevant clinical information quantitatively and EULAR recommends their use to monitor patients.\textsuperscript{56} The use of composite indices to assess both the activity and structural damage would assist in clinical practice to guide therapeutic decisions as objectively as possible.\textsuperscript{57} Furthermore, the PGA exhibits great variability, even among physicians with extensive experience.
treating patients with SLE, not making it an ideal tool to assess activity.

Both the BILAG and SLEDAI are the indexes that have a more complete validation and are most commonly used in RCTs and cohort studies. The ECLAM has the advantage of being the easiest to calculate and SLAM is the most sensitive to change, since it is the only one that can give a greater score to high severity. The SLAQ is a patient answered index that exhibits good correlation with SLAM, from which it derives, so it may be a good alternative for patient visit in which there is less time to spend per patient or in population studies.

Ethical Responsibilities

Protection of people and animals. The authors declare that no experiments were performed on humans or animals for this investigation.

Data confidentiality. The authors declare that no patient data appears in this article.

Right to privacy and informed consent. The authors have obtained the informed consent of the patients and/or subjects mentioned in the article. The author for correspondence is in possession of this document.

Financing

This study was funded by the Spanish Foundation of Rheumatology.

Conflict of Interest

Iñigo Rúa-Figueroa has had contracts with GSK and has attended courses/conferences sponsored by GSK and MSD.

The rest of the authors have no disclosures to make.

Appendix. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.reuma.2014.01.011.

References


