Measures of Adult Systemic Lupus Erythematosus: Disease Activity and Damage

Shilpa Arora,1 David A. Isenberg,2 and Isabel Castrejon3

INTRODUCTION

Many instruments are available to assess disease activity and capture damage in patients with systemic lupus erythematosus (SLE). These indices are frequently used in clinical trials, and some are used in epidemiological studies and daily practice, because they help to guide clinical decisions and evaluate treatment response. SLE is a heterogeneous disease with variable presentations of different organ systems and an unpredictable disease course in the same and different patients. Because of disease heterogeneity and lack of a gold standard reflecting the complete disease spectrum, a combination of different measures in composite indices is necessary to follow disease activity.

Several indices have been developed over the last 30 years with different levels of validation (1); however, there is no universal consensus about their use. Feasible and sensitive assessment of SLE in clinical trials, observational studies, and clinical settings remains a challenge to the rheumatology community. This review presents a summary of the indices most frequently used in the last 5 years for disease activity and damage measurement in patients with SLE. Selection of these indices was based on evidence assessing their psychometric properties and popularity of use. The updated versions are included here with the original versions reviewed previously (2).

Four disease activity measures have been included, one of which is an organ/system assessment scale that assesses disease activity in individual organs or systems (the British Isles Lupus Assessment Group [BILAG] 2004 index); the other three are global scoring systems that provide an overall measure of activity (the Systemic Lupus Erythematosus Disease Activity Index 2000 [SLEDAI 2K], the Systemic Lupus Activity Questionnaire [SLAQ], and the Lupus Foundation of America Rapid Evaluation of Activity in Lupus [LFA-REAL]). It is inherent in the scoring of these indices that only items attributable to active SLE are recorded and not items attributable to damage or comorbid conditions. The global scoring systems, such as the SLEDAI 2K and SLAQ, are binary (ie, they record clinical features to be present or absent and do not distinguish clinical features that are improving, but not resolved, from those that are unchanged or worse). The BILAG index and LFA-REAL, on the other hand, are able to capture improvement or worsening of persistent clinical features. We have also included the Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI), which grades cutaneous manifestations of SLE, and the Safety of Estrogens in Lupus National Assessment (SELENA)–Systemic Lupus Erythematosus Disease Activity Index (SELENA) Flare Index (SFI), which is useful for better capturing flares and is increasingly used in therapeutic trials in SLE.

Damage is irreversible by definition in patients with SLE and may be driven by diverse factors: inflammation from SLE itself, concomitant comorbidity, side effects of drugs, and infectious complications. Damage is an important outcome in SLE and is known to predict morbidity and mortality (3–5). Two indices have been included in this review to assess cumulative damage: the Systemic Lupus International Collaborating Clinics (SLICC)/American College of Rheumatology (ACR) Damage Index (SDI) and the Brief Index of Lupus Damage (BILD).

Patient-reported outcomes (PROs) have gained popularity for use in SLE because these are unique in capturing patients’ perspectives of their disease. A majority of PRO measures focus on measurement of overall health assessment and quality of life in SLE; however, few are available for measurement of disease activity and damage (6,7). The SLAQ and BILD, included in this review, are PRO measures, and recently a PRO version of the LFA-REAL has been described. These measures offer advantages of being simple to use and applicable in large epidemiological studies. The European League Against Rheumatism (EULAR) Outcome Measures Library is a growing and freely available database of validated PRO measures (http://oml.eular.org/), providing unified access to PRO measures to improve research and enhance the use of these measures in clinical practice (8). When considering the use...
of PROs, it should be recognized that it might be hard for a patient (and sometimes even for a physician) to distinguish whether the cause of a clinical feature is due to activity or damage or even both.

One of the aims of the disease measures is to gauge improvement over time, which is especially relevant in clinical trials to be able to measure the response from interventions effectively. Different responder indices have been developed that are either derivatives or combinations of existent indices. The SLE Responder Index 4 (SRI 4) and the BILAG-based Combined Lupus Assessment (BICLA), the two most commonly used responder indices, have been included and reviewed for this purpose.

Numerous failed SLE therapeutic trials are blamed, in part, on the deficiencies in available disease activity measures, highlighting the importance of the instrument selected to measure disease activity or damage in SLE. Disease activity, damage, and responder indices, with their specific characteristics, are reviewed to help clinicians and researchers in the selection of the most appropriate measure according to the intended purpose of the study or clinical evaluation.

**BRITISH ISLES LUPUS ASSESSMENT GROUP 2004 INDEX**

**Description**

**Purpose.** The BILAG 2004 index is a transitional score to assess organ-/system-based activity due to lupus based on the physician’s intent-to-treat premise. The BILAG 2004 index, published in 2005 (9), is a revision of the classic BILAG index published in 1988 (10).

**Content.** This index comprises specific manifestations across nine organs/systems: constitutional, mucocutaneous, neuropsychiatric, musculoskeletal, cardiorespiratory, gastrointestinal, ophthalmic, renal, and hematological.

**Number of items.** There are 97 items in the 9 different organs/systems and 4 additional items for calculating the glomerular filtration rate.

**Response options/scale.** Each item (assumed to be due to activity) is scored as follows: ND = not done, 0 = not present, 1 = improving, 2 = same, 3 = worse, and 4 = new, yes, or no.

**Recall period for items.** The recall period refers to the assessment of manifestations in the last 4 weeks compared with the previous 4 weeks.

**Cost to use.** There are no additional costs if the measure is performed manually. However, this instrument includes 97 items that need to be scored according to a glossary, adding complexity, even for a trained investigator, to complete manually.

**How to obtain.** The BILAG 2004 index form, glossary, and scoring scheme are available at https://www.ncbi.nlm.nih.gov/ pmc/articles/PMC2919194/.

**Practical application**

**Method of administration.** The BILAG 2004 index is completed by the physician and requires additional training on use of the glossary items and scoring scheme.

**Scoring.** Each item is scored as either ND or from 0 to 4. Scores are then converted into alphabetical scores (A-E) for each system either through a computerized system or manually after training.

**Score interpretation.** Scoring is based on the principle of the physician’s intention to treat, in which category A (for activity) implies severe disease activity requiring systemic high-dose oral glucocorticoids (equivalent to prednisolone at > 20 mg/dl), systemic immunomodulators, or high-dose anticoagulation; category B (for beware) implies moderate disease activity requiring systemic low- dose oral glucocorticoids (equivalent to prednisolone at ≤ 20 mg/dl), intramuscular/intraarticular or soft tissue glucocorticoid injection, or topical glucocorticoids, topical immunomodulators, or antimalarias; and categories C (for contentment), D (for discount), and E (for never, ever active) imply mild disease, inactive disease but previously affected, and system never involved, respectively.

**Respondent time to complete.** Ten to 15 minutes in addition to completing a history and physical examination.

**Administrative burden.** Usually up to 30 to 60 minutes. Some items cannot be scored until laboratory test results are available (usually later that day or the following day). Ophthalmic manifestations usually need to be assessed by an ophthalmologist, and these items would need to be recorded after the response is received from the ophthalmologist.

**Translations/adaptations.** The BILAG 2004 system tally (BST) and simplified BST divide systems into six and three components, respectively, based on the level of activity and have been shown to have good area under the receiver operating characteristic curve in predicting an increase in therapy (>0.8) (11). The BILAG 2004 pregnancy index has been shown to be reliable for assessment of disease activity in pregnant patients with SLE, with a good level of agreement greater than 70% in all systems (12). The British Lupus Integrated Prospective System (BLIPS) can calculate the BILAG 2004 index score and also records other disease activity indices. A web-based program, known as iBLIPS, has been launched and is continuously updated for use in research studies and clinical trials (13). It has been widely used for more than 10 years.
Psychometric information

Floor and ceiling effects. Assessment of the BILAG 2004 index in a sample of 250 patients with SLE revealed a new A or B score in 61.6% of patients over a 12-month period. An A flare was observed in 10.4% of patients, a B flare (in which a B score was preceded by a D or E score) in 26.0% of patients, and 25.2% of patients had a B score in a system in which a C score was previously recorded (14).

Reliability. Isenberg et al (9) conducted two real-patient exercises with eight adult patients with SLE and eight rheumatologists during the development of the BILAG 2004 index and demonstrated good reliability (intraclass correlation coefficient [ICC] >0.60) and high levels of physician-patient agreement (ratio of SD attributable to the physician/the SD attributable to the patient: <0.40) across all organ systems except for the musculoskeletal system (15). Interrater reliability was also assessed in a multicenter study by Yee et al (16) using 97 patients with 2 exercises (E1 and E2). Several issues were identified in the glossary following E1, which was then updated and additional training provided to the raters. The overall ICC determined in E1 was 0.45 (95% confidence interval [CI] 0.31-0.58) and improved to 0.67 (95% CI 0.54-0.76) in E2 (16).

Validity. Construct and criterion validity of the BILAG 2004 index to assess disease activity in SLE was determined in a multicenter cross-sectional study using 369 patients with SLE. Increasing overall scores, using the BILAG 2004 index, were associated with increasing erythrocyte sedimentation rates, decreasing C3 and C4 levels, elevated anti–double-stranded DNA levels, and higher SLEDAI 2K scores (P < 0.01). Scores indicating active disease (overall BILAG 2004 index scores of A and B) were significantly associated with an increase in therapy (odds ratio 19.3; P < 0.01). The BILAG 2004 and classic BILAG indices have comparable sensitivity, specificity, positive predictive value, and negative predictive value (17).

Responsiveness. The ability to detect change was studied in a prospective longitudinal study in which the relationship between change in disease activity and change in therapy between two consecutive visits was analyzed using 1761 assessments from 347 patients with SLE. An increase in the overall score was associated with an increase in therapy (coefficient: 1.35; 95% CI 1.01-1.70), and a decrease in the overall score was associated with a decrease in therapy (coefficient: 0.44; 95% CI 0.16-0.71) (18).

Minimally important differences. By using the BILAG 2004 index, a flare can be defined in terms of the number of systems scoring a new A or B score based on items recorded as 4 (new) or 3 (worse). On this basis, severe flare is defined as a score of A in any system, and moderate flare is defined as two B scores, whereas mild flare is a single new B score (14).

Generalizability. The numerical scoring system facilitates comparisons with global indices by converting the assessments so that A = 12 points, B = 8 points, C = 1 point, and D/E = 0 points (15).

Use in clinical trials. The BILAG 2004 index has been extensively included in randomized clinical trials (RCTs). However, it is important to consider that exploratory RCTs, by their nature, involve selected patient populations and focus on short-term efficacy. For long-term efficacy and safety, registry data are more useful to evaluate biologic agents in real-world practice between other registries, as exemplified by the BILAG Biologics Register (19) and the Registry of Systemic Lupus Erythematosus Patients of the Spanish Society of Rheumatology (RELESSER) (20).

Critical appraisal of overall value to the rheumatology community

Strengths. The BILAG 2004 index is a comprehensive index that has been shown to be valid, reliable, and sensitive to change. The BILAG 2004 index measures disease activity in different organs/systems separately and can also measure partial improvement in a given organ system, in contrast with the SLEDAI.

Caveats. Administrative time burden is a caveat to the BILAG 2004 index, along with the need for special training to the raters.

Clinical usability. The BILAG 2004 index is useful in assessing flares between consecutive visits. The assessment of disease flare using the BILAG 2004 index was studied in 16 patients who were also rated by a panel of 16 rheumatologists, and the rate of complete agreement for any flare versus no flare was 81% (95% CI 55%-94%) for the BILAG 2004 index and 75% (95% CI 49%-90%) for Physician Global Assessment (PGA) (21).

Research usability. The BILAG 2004 index records each item as new, the same, worse, or improving in separate organs, which is relevant when assessing the effect of an intervention in a clinical trial.

SYSTEMIC LUPUS ERYTHEMATOSUS DISEASE ACTIVITY INDEX 2000

Description

Purpose. To measure global disease activity from lupus modeled on clinicians’ global judgment. The SLEDAI 2K (22) is the 2002 version of the original SLEDAI, which was published in 1992 (23).
Content or domains. The SLEDAI 2K includes evaluation of specific manifestations in nine organ systems.

Number of items. The SLEDAI 2K contains 24 items, of which 16 are clinical and 8 are based solely on laboratory test results.

Response options. A manifestation is recorded if it is present, regardless of severity or whether it has improved or worsened. In the original index, the variables rash, alopecia, mucus membrane lesions, and proteinuria were counted as active only if they represented their first occurrence or recurrence (to distinguish from chronic lesions). In contrast, the SLEDAI 2K scores the presence of any rash, alopecia, and mucosal ulcers and a new, recurrent, or persistent proteinuria (urine protein levels >0.5 gm/24 hours) to capture persistent disease activity.

Recall period. This index records disease manifestations occurring within the 10 days preceding assessment.

Cost to use. No additional costs.

How to obtain. The SLEDAI 2K can be obtained by contacting Drs. Dafna Gladman, MD, and Murray Urowitz, MD, at Toronto Western Hospital, 399 Bathurst Street, JE 410B, Toronto, Ontario, M5T 2S8, Canada.

Practical application

Method of administration. The SLEDAI 2K is completed by the physician.

Scoring. Weighting is used, resulting in individual item scores ranging from 1 to 8, which are simply added to give a global score ranging from 0 to 105.

Score interpretation. Remission, low disease activity (LDA), and high disease activity (HDA) have been defined for SLEDAI 2K scores. Remission has been defined as no clinical manifestation (with or without serologic manifestations), although patients could be taking antimalarial medications only. LDA has been defined as a clinical SLEDAI 2K score less than 3 (with or without positive serology results) based on the presence of only one clinical manifestation, with a score range of 1 to 2 for patients who could be taking antimalarial medications but not glucocorticoids or immunosuppressive drugs. HDA has been defined as an SLEDAI 2K score greater than 6 (24). Similarly, lupus LDA state has been defined using the following criteria, the attainment of which is associated with improved outcomes in SLE: 1) SLEDAI 2K score less than or equal to 4, with no activity in major organ systems; 2) no new lupus disease activity compared with the previous assessment; 3) PGA less than or equal to 1; 4) a current prednisolone (or equivalent) dose less than or equal to 7.5 mg daily; and 5) well-tolerated standard maintenance doses of immunosuppressive drugs and approved biological agents (25).

Respondent time to complete. The SLEDAI 2K takes approximately 10 minutes to complete.

Administrative burden. A history and physical examination are needed. The SLEDAI 2K cannot be scored until laboratory test results are available, which may take a few days; however, clinical SLEDAI 2K scores without laboratory test results could be reported.

Translations/adaptations. There are simplified English and Spanish versions of the SLEDAI without the immunological tests, which makes the index cheaper to administer, such as the MEX-SLEDAI (26). The SELENA-SLEDAI was devised for use in the SELENA study, in which the definitions of some of the descriptors in the SLEDAI were modified to ensure that the descriptors captured ongoing disease activity (27). A modified version of the SLEDAI has been developed for use in pregnancy (28). The SLEDAI 2K score measured over the last 30 days has been described and found to be comparable to the SLEDAI 2K score measured over 10 days (29). Recently, the SLEDAI 2KG, a modification of the SLEDAI 2K, has been suggested to describe disease activity while accounting for glucocorticoid dose category, and to identify more responders compared with the SLEDAI 2K (30).

Psychometric information

Floor and ceiling effects. The mean SLEDAI 2K score in the initial validation cohort of 960 patients with SLE was 10.4 (SD 8.7). On studying outcomes of remission and LDA of SLE using a cohort of 1783 patients with lupus observed between 1970 and 2015, 620 were found to fulfill the criteria for one of the three groups (LDA: 80 patients [12.9%]; remission: 191 patients [30.8%]; HDA: 349 patients [56.3%]) (24).

Reliability. The reliability of the SLEDAI 2K was demonstrated in a multicenter and multiethnic study of 93 patients with SLE, with an agreement for each of the items between 81.7% and 100%.

Validity. Gladman et al (22) showed a high correlation between SLEDAI 2K and SLEDAI scores across all visits (r = 0.97; P = 0.0001), and both indices similarly predicted mortality. The SLEDAI 2K demonstrated a similar progression of the same magnitude over a range of disease activity as the SLEDAI when compared against independent clinician’s assessment of disease activity (22). In another study of 92 patients with SLE, evaluation of construct validity of multiple indices showed a correlation coefficient of 0.677 between the SLEDAI 2K and PGA (P < 0.0001) (31).
**Responsiveness.** Sensitivity analyses of four SLE disease activity indices in a longitudinal study of 102 patients with SLE showed that the BILAG 2004 index and the SLEDAI 2K have the highest association with treatment change (32).

**Minimally important differences.** The SLEDAI 2K score increased by more than 3 when the clinician assessed that the patient was experiencing a flare (33). In a recent longitudinal cohort of 334 patients with SLE with 36 months of follow-up, adjusted mean PGA and SLEDAI 2K scores demonstrated a high correlation ($\rho = 0.824; P < 0.0005$), although the SLEDAI 2K had limited sensitivity to evaluate for meaningful change in disease activity, defined by a PGA change greater than or equal to 0.3 points from baseline (34).

**Generalizability.** Activity categories have been defined, and a score greater than 5 is associated with a probability of initiating therapy in more than 50% of patients.

**Use in clinical trials.** Although the SLEDAI does not capture improving or worsening in a given organ system and is, thus, less sensitive to change, it has been included in multiple RCTs, for example, in the recently published protocol for the Belimumab After B cell Depletion Therapy in patients with systemic lupus erythematosus (BEAT Lupus) RCT, together with the BILAG 2004 index, to evaluate the use of belimumab after B cell depletion therapy in patients with SLE (http://www.isrctn.com/ISRCTN47873003).

**Critical appraisal of overall value to the rheumatology community**

**Strengths.** The SLEDAI 2K is completed rapidly (apart from the laboratory test results) and requires minimal training, which accounts for its wide acceptance in clinical and research settings.

**Caveats.** Neither version of the SLEDAI captures worsening of an already existing feature or detects partial improvement, limiting their role in RCTs.

**Clinical usability.** Adjusted mean SLEDAI 2K scores, which measure change in the SLEDAI 2K score over time, have been shown to correlate with survival and other major outcomes in SLE, including presence of damage and coronary artery disease (35).

**Research usability.** The SLEDAI 2K is one of the most commonly used global disease activity measures in longitudinal observational studies and clinical trials.

**SYSTEMIC LUPUS ACTIVITY QUESTIONNAIRE**

**Description**

**Purpose.** The SLAQ is a patient-reported assessment of disease activity in SLE that is useful for tracking and screening large groups of patients in epidemiological studies.

**Content.** It is modeled on the physician-reported disease activity index Systemic Lupus Activity Measure (SLAM) using items amenable to patient self-report (36).

**Number of items.** The SLAQ consists of questions on 24 specific symptoms of disease activity, patient assessment of flare, and patient global assessment (PaGA) of their disease activity.

**Response options.** The SLAQ has Likert responses with four response categories (no problem, mild, moderate, and severe) for specific symptom questions and patient assessment of flare and has a numerical rating score of 0 to 10 for PaGA.

**Recall period.** The previous 3 months was chosen because it may be burdensome for the patients in an epidemiological study to fill questionnaires monthly.

**Cost to use.** No additional costs.

**How to obtain.** Available in Appendix A of the article by Karlson et al (36).

**Practical application**

**Method of administration.** Paper or electronic copy is available and can be mailed or administered through telephone interviews.

**Scoring.** Items are weighted and aggregated in a manner analogous to the scoring system used in the SLAM, and scores can range from 0 to 44 for the disease-specific questions (2).

**Score interpretation.** Using a cutoff score of 9 points or more on the SLAQ results in a sensitivity of 83%, specificity of 70%, positive predictive value of 74%, and negative predictive value of 80% to predict a score of 3 points on the SLAM-no lab (36).

**Respondent time to complete.** Up to 10 minutes.

**Administrative burden.** Up to 10 minutes because it does not require any laboratory data.
Translations/adaptations. Translation and cross-cultural validation is available in Italian, Japanese, and German languages (37–39).

Psychometric information

Floor and ceiling effects. In a large-sample validation study of the SLAQ using the Lupus Outcome Study (LOS) sample from University of California, San Francisco, SLAQ scores were seen to be fairly normally distributed, although a slight rightward skew was noted, with 0.9% of the sample having scores greater than 35 (40).

Reliability. The SLAQ demonstrated excellent internal consistency, with a Cronbach’s $\alpha$ of 0.87 in the previously mentioned longitudinal cohort (40).

Validity. Karlson et al (36), during development of the SLAQ, studied it in 93 patients with SLE and showed moderately high correlation between the total SLAQ score and the SLAM-no lab score ($r = 0.62; P < 0.0001$). However, correlations between patient-clinician matched pairs of items ranged from 0.06 to 0.71.

Responsiveness. In the LOS study, the SLAQ demonstrated a small to moderate degree of responsiveness; standardized response means were 0.66 and −0.37 for those reporting clinical worsening and improvement, respectively (40).

Minimally important differences. Positive predictive values for the SLAQ range from 56% to 89% for detecting clinically significant disease activity (36).

Generalizability. The SLAQ is useful to identify patients with new or increased disease activity who need further evaluation by a physician, with a positive predictive value ranging from 56% to 89% for detecting clinically significant disease activity.

Use in clinical trials. It is less frequently used in RCTs and more frequently used in epidemiological studies.

Critical appraisal of overall value to the rheumatology community

Strengths. The SLAQ is the first patient-reported disease activity measure developed specifically for patients with SLE. It has been shown to correlate with SLAM and other PRO measures, such as the Routine Assessment of Patient Index Data 3 (RAPID3), but it poorly correlates with other physician-reported disease activity measures (41).

Caveats. In the SLAQ, the patient is asked to rate the presence and severity of lupus symptoms, but attribution to SLE cannot be taken at face value. Indicators of disease activity, such as laboratory findings of hematologic or renal disease, might not be associated with symptoms and, hence, are not captured.

Clinical usability. The SLAQ is intended to be used as an initial screen to identify patients with new or increased disease activity who may need further evaluation by a physician.

Research usability. The SLAQ is useful in conducting research studies in which the numbers of patients and distance from the center are barriers to face-to-face assessment by physicians.

LUPUS FOUNDATION OF AMERICA RAPID EVALUATION OF ACTIVITY IN LUPUS

Description

Purpose. The LFA-REAL is an efficient and scalable SLE disease activity measure for use as an outcome measure in clinical trials as well as in real-world clinical practices (42).

Content. It is constructed as an expanded version of PGA (based on completing a separate PGA for each active organ system).

Number of items. The LFA-REAL has six or more anchored visual analog scales (VASs); the first six assess the most commonly affected organs, and other scales can be added to record features that do not fit the fixed six categories or to separately score each descriptor in organs with two or more manifestations.

Response options. Each VAS is 100 mm in length with landmarks: 0 = none, 1 = mild, 2 = moderate, and 3 = severe.

Recall period. The rater needs to score the symptoms that are active in a given patient on the given day of scoring.

Cost to use. No additional costs.

How to obtain. Available as a downloadable figure in the article by Askanase et al (42).

Practical application

Method of administration. Physician completed.

Scoring. Scoring requires simple addition of VASs of individual items, and the score ranges from 0 to 700 mm.

Score interpretation. An anchor at 1.5 serves as a potential cutoff for the initiation of immunosuppressants.

Respondent time to complete. Up to 5 minutes.
Administrative burden. Up to 5 minutes because it does not require any additional laboratory measures.

Translations/adaptations. Recently, a PRO version of the LFA-REAL was proposed, which, with a recall period of 4 weeks, asks patients to evaluate the severity and progression of symptoms associated with their current lupus activity through a series of anchored VASs (43). Further studies on assessment of the quantitative performance of LFA-REAL PRO scores, including tests of reliability, validity, and responsiveness, are under way.

Psychometric information

Floor and ceiling effects. Usually two to four scales in total are scored for an individual patient.

Reliability. LFA-REAL scores were compared when scored by trained lupus clinical investigators or clinicians at two separate visits. The ICC between the scores of investigators and clinicians was 0.79 for visit 1 ($P < 0.001$) and 0.86 for visit 2 ($P < 0.001$); higher correlations were seen in trained users.

Validity. In the initial validation study of 91 consecutive patients with SLE, the total LFA-REAL score (sum of each organ-based PGA) correlated with the PGA, SLEDAI, and BILAG index scores, with Spearman’s rank correlation coefficient (SRCC) values of 0.903, 0.816, and 0.933, respectively ($P < 0.001$ for all analyses) (42).

Responsiveness. Responsiveness of the LFA-REAL was evaluated and compared with standard SLE measures and presented recently to EULAR (44). Change in LFA-REAL scores at each visit correlated with change in PGA ($r = 0.857$), change in SLEDAI scores ($r = 0.701$), and change in BILAG index scores ($r = 0.700$) ($P < 0.0001$). Change in LFA-REAL scores also strongly correlated with dichotomous scores of the responder indices SRI 4 and BICLA ($r = 0.876$ and 0.852, respectively; $P < 0.0001$).

Minimally important differences. Evaluation is pending.

Generalizability. Evaluation is pending.

Use in clinical trials. The LFA-REAL is not routinely included in RCTs so far.

Critical appraisal of overall value to the rheumatology community

Strengths. The LFA-REAL can be learned by a clinician skilled in the care of patients with lupus within minutes; in practice, it can be scored rapidly for most patients.

Caveats. Although it is possible to detect small changes in organ-specific activity using the LFA-REAL, these may not reflect clinically significant change, nor are they necessarily the most discriminatory end points for clinical trials. There also may be measurement bias in using VASs due to potential variations in how physicians interpret these scales. So far, no formal data on minimally important differences or generalizability are available.

Clinical usability. The data from the LFA-REAL can be evaluated on the basis of individual features, organ-based scoring, or overall disease activity, and it is easy to distinguish between different combinations of mild, moderate, or severe manifestations, allowing each to contribute appropriately to an overall score.

Research usability. It is suitable to track the course of disease accurately over time in response to therapy, both in trials and in general clinical practice.

CUTANEOUS LUPUS ERYTHEMATOSUS DISEASE AREA AND SEVERITY INDEX

Description

Purpose. The CLASI is an organ-specific outcome measure used to assess and grade cutaneous manifestations in both SLE and cutaneous lupus erythematosus (CLE) (45,46).

Content or domains. The CLASI has two scores. The activity (A) score describes activity of the disease, whereas the damage score (D) describes damage caused by the disease. It is designed as a table in which rows denote anatomical areas and columns include scores for major clinical symptoms.

Number of items. Items for activity include erythema, scale/hypertrophy, mucous membrane involvement, acute hair loss, and no scarring alopecia. The extent of erythema and scale/hypertrophy is recorded for different body parts. Items for damage include depigmentation and scarring in different body parts and scarring alopecia.

Response options/scale. Response options are recorded as dichotomous (presence/absence of symptoms) and vary from 0 to 6, depending on the extent of severity. The severity of involvement for each skin area is documented by the worst affected lesion within each specific area.

Recall period for items. Assessment on the given day.

Cost to use. No additional cost.
How to obtain. Available as a downloadable figure in the articles by Albrecht et al and Albrecht and Werth (45,46).

Practical application

Method of administration. Physician administered.

Scoring. Scores are summated in each of the two domains of activity and damage. To increase the score’s responsiveness, more weight has been given to photosensitive areas, such as the face, neck, and hands. Patients are asked whether depigmentation has occurred in CLE lesions and has remained visible for more than 12 months, which is taken to be permanent. If so, the depigmentation score is doubled.

Score interpretation. Mild, moderate, and severe disease, based on the physician’s subjective assessment of severity, correspond with CLASI A score ranges of 0 to 9, 10 to 20, and 21 to 70, respectively (47).

Respondent time to complete. Between 2 and 10 minutes.

Administrative burden. Up to 10 minutes because no laboratory results are required.

Translation/adaptations. A revised version of the CLASI was reported in 2010, with additional items describing the morphology of the rashes added to the original score (edema/infiltration, subcutaneous nodule/plaque, and “lupus hair”) (48). The CLASI exhibits a good intra- and interrater reliability in pediatric CLE (49).

Psychometric information

Floor and ceiling effects. The mean activity score in the initial validation study ranged from 10.1 to 40, and the mean damage score ranged from 0.7 to 41.6 (45).

Reliability. In the previously mentioned study, Spearman’s coefficient for intra-rater reliability was 0.96 for the activity score (95% CI 0.89-1.00) and 0.99 for the damage score (95% CI 0.97-1.00). The ICC for interrater reliability was 0.86 for the activity score (95% CI 0.73-0.99) and 0.92 for the damage score (95% CI 0.85-1.00) (45).

Validity. Content and face validity have been assessed among dermatologists, rheumatologists, and an ACR response criteria committee on SLE using standardized interviews, with appropriate changes in the score made based on their suggestions (45). The CLASI activity and damage scores have been shown to correlate with physician-assessed cutaneous activity and damage in SLE (total CLASI activity and SLEDAI-rash \( r = 0.42; P = 0.02 \); CLASI-mucosal and SLEDAI-mucosal \( r = 0.65; P = 0.001 \); and CLASI-recent hair loss and SLEDAI-alopecia \( r = 0.61; P = 0.001 \)) (50).

Responsiveness. Eight patients with CLE were followed up for 56 days after change in therapy. Changes in the CLASI A score correlated with physicians’ assessment of skin health \( r = 0.97; P = 0.005; n = 8 \), patients’ global health score \( r = 0.85; P = 0.004; n = 9 \), and pain \( r = 0.98; P = 0.004; n = 5 \). Those patients who failed to improve had stable scores throughout their course (51).

Minimally important differences. Clinical improvement has been shown to be associated with a mean 3- to 4-point or 18% to 20% decrease in the CLASI A score (47).

Generalizability. Specific for skin involvement in SLE.

Use in clinical trials. Extensively used in therapeutic clinical trials of CLE and cutaneous manifestation of SLE (52–54).

Critical appraisal of overall value to the rheumatology community

Strengths. The CLASI score is a specific score for measuring cutaneous disease activity and damage in SLE. Measurement of erythema as a marker of activity, more weightage to the scalp and face (which have larger impacts on PROs), and differentiation into activity and damage scores are major strengths of this instrument.

Caveats. The CLASI must be rated by physicians trained in the cutaneous manifestations of SLE. Untrained observers may be easily confused by comorbidities.

Clinical usability. The CLASI is useful for detailed measurement of the extent and severity of skin involvement, allowing for assessment and comparison of patients in terms of extent and acuity of the disease as well as observation of their reaction to therapy.

Research usability. The CLASI measures skin changes from SLE reliably, and activity scores are specifically useful in therapeutic trials in which improvement in the score can capture efficacy of treatment.

SELENA-SLEDAI FLARE INDEX

Description

Purpose. The SFI was developed for use in clinical trials by the SELENA group with the intention of distinguishing severe flares from those that are only mild or moderate (55,56).
Content or domains. It includes change in the SELENA-SLEDAI score, PGA, a few additional items, and treatment modifications.

Number of items. The SFI requires scoring of the SELENA-SLEDAI. Additional items not captured by the disease activity measure include the following: new/worse skin, stomatitis, serositis, arthritis or fever for mild/moderate flare and new/worse central nervous system SLE, vasculitis, nephritis, myositis, platelet count less than 60,000, and hemolytic anemia with hemoglobin less than 7 mg/dl for severe flare.

Response options/scale. A SLEDAI score greater than or equal to 3 points and a greater than or equal to 1-point increase in PGA (range 0-3) is considered mild/moderate flare. A SLEDAI score greater than or equal to 12 points, and a greater than or equal to 2.5-point increase in PGA is included for severe flare. Treatment criteria include the following: prednisone less than 0.5 mg/kg/d or addition of non-steroidal anti-inflammatory drugs with hydroxychloroquine in mild/moderate flare and prednisone greater than 0.5 mg/kg/dl, new immunosuppressive, or hospitalization for SLE in severe flare.

Recall period for items. Ten days preceding assessment.

Cost to use. No additional costs to use.

How to obtain. Downloadable from the supplemental material in the article by Petri et al (56).

Practical application

Method of administration. Administered by physicians.

Scoring and interpretation. Presence of any of the abovementioned criteria places patients into either the mild/moderate flare or the severe flare category.

Respondent time to complete. Up to 5 to 10 minutes.

Administrative burden. Laboratory measures are included in the scoring of the SELENA-SLEDAI.

Translation/adaptations. The SFI has been revised to provide evaluation of flares in different organs similarly to the BILAG index. The revised SFI (rSFI) is independent of the SLEDAI, and includes treatment data, and distinguishes between mild, moderate, and severe flares. Specific manifestations in different organ systems (constitutional, mucocutaneous, neurological, musculoskeletal, cardiopulmonary, hematological, renal, and gastrointestinal) are recorded in the rSFI. Another modification of the SFI with exclusion of medication scores showed that after exclusion of medication criteria, 42 of 55 SFI severe flares and 15 of 49 mild/moderate flares were downgraded in severity, and those retained in original categories had higher disease activity measured by several indices, thus suggesting the role of the deconstructed flare index in recognizing true clinical worsening for each category of flare (57).

Psychometric information

Floor and ceiling effects. Researchers used the SFI in a prospective French cohort of 331 patients with SLE and found that 40.2% had a mild/moderate flare and 5.7% had a severe flare at 6 months (58).

Reliability. The SFI has shown high reliability for severe flare classification ($\kappa = 0.65$) and is modest for mild/moderate flare classification ($\kappa = 0.45$) (59).

Validity. The SFI and rSFI have shown fair agreement with the BILAG flare index ($\kappa = 0.30$ [weighted $\kappa = 0.39$] and $\kappa = 0.31$ [weighted $\kappa = 0.51$], respectively). Comparison of assessment of no flare, mild/moderate flare, and severe flare between the SFI and rSFI showed substantial agreement between the SFI and rSFI ($\kappa = 0.70$; weighted $\kappa = 0.73$). In another study, assessment of any flare versus no flare showed an agreement rate of 84% (73%-91%) between the BILAG 2004 index and rSFI among different rheumatologists and a precise agreement rate of 52% (39%-63%) if type of flare was taken into account. A similar comparison between PGA and the rSFI gave agreement rates of 98% (90%-99%) and 82% (67%-92%), respectively (21). No flare and severe flare were easily distinguished; mild and moderate flare, less easily so.

Responsiveness. In a large study presenting patient exercise of 451 cases of SLE patients, 3 physician assessments of flare was compared with the BILAG 2004 index, SFI, and rSFI, with corresponding rates of 67%, 72%, and 70%, respectively (60). The corresponding weighted $\kappa$ coefficients for each instrument were as follows: BILAG 2004 index, $\kappa = 0.82$; SFI, $\kappa = 0.59$; and rSFI, $\kappa = 0.74$.

Minimally important differences. Not formally assessed.

Generalizability. Not formally assessed.

Use in clinical trials. Both versions are used in clinical studies (58) as well as trials (59).
Critical appraisal of overall value to the rheumatology community

Strengths. Both of the indices (SFI and rSFI) are unique in specifically capturing flare in SLE. Inclusion of treatment modifications increases the sensitivity to identify any flares.

Caveats. These instruments measure worsening and flare in SLE as intended but not improvement in the disease. The SFI does not distinguish between mild and moderate flares, which is a limitation in trials assessing clinical efficacy because a moderate flare is likely to be more clinically significant and better able to discriminate treatment efficacy than a mild flare.

Clinical usability. The SFI and rSFI can be incorporated in clinical practice to observe patients clinically for documentation of flares.

Research usability. Ability to quantify flares using multiple items provides for sensitive tools to assess for flares in descriptive clinical studies and therapeutic trials.

SYSTEMIC LUPUS INTERNATIONAL COLLABORATING CLINICS/AMERICAN COLLEGE OF RHEUMATOLOGY DAMAGE INDEX

Description

Purpose. To assess the accumulated damage in patients with SLE; the clinical index was developed by the SLICC group in 1996 and endorsed by the ACR (61).

Content. The SDI records damage occurring in patients with lupus regardless of cause. Damage may result from previous disease activity and may lead to organ dysfunction (e.g., renal failure or neurocognitive abnormalities), adverse effects from medications (e.g., diabetes or avascular necrosis as side effects of the use of glucocorticoids), or concomitant disease (e.g., myositis or rheumatoid arthritis).

Number of items. The index has 41 items covering 12 systems.

Response options. Each item is recorded as present or absent. Manifestations should be recorded as damage only if they develop after the onset of lupus, provided they fulfill the definition in the glossary and irrespective of attribution.

Recall period. To be scored, items should persist continuously for 6 months or be associated with a pathological scar (e.g., myocardial infarction).

Cost to use. No additional costs.

How to obtain. Contact Dr. Dafna D. Gladman, Krembil Research Institute, Toronto Western Hospital, University Health Network, 399 Bathurst Street, IE-410B, Toronto, Ontario M5T 2S8, Canada. The SDI is downloadable from Appendix 1 of the article by Gladman et al (62).

Practical application

Method of administration. Completed by physician.

Scoring. The SDI is scored from 0 to 47. Some items can score 2 for recurrent events, such as repeated strokes and avascular necrosis at two sites. End-stage renal disease (ESRD) is given 3 points.

Score interpretation. The SDI score is 0 at diagnosis by definition. Damage is considered if the score is greater than or equal to 1.

Respondent time to complete. Up to 15 minutes.

Administrative burden. Complete history and a physical examination are needed. A chart review for patients with longer disease duration is the most time-consuming step. Some items require imaging, such as a shrinking lung and avascular necrosis, and ocular items require an ophthalmological examination.

Translations/adaptations. The SDI is available in English only. The Lupus Damage Index Questionnaire is a self-administered version of the SDI that has been validated in English, Spanish, Portuguese, and French (63,64).

Psychometric information

Floor and ceiling effects. The maximum score is 47, but patients rarely score above 10 points. In most lupus cohorts, approximately 50% of patients will have at least one item of damage.

Reliability. Interrater reliability was assessed in 10 actual patients with SLE, and the SDI detected differences among patients (\(P < 0.001\)) but not by physician (observer effect, \(P = 0.933\)) (62). Concordance was demonstrated in the SDI among observers despite a wide spectrum of disease activity.

Validity. During the development of the index, 19 physicians scored the index on 42 case scenarios, and analysis showed that it could measure change in damage seen in patients with both active and inactive disease. Patients who had active disease had
a higher increase in damage, and for those with stable disease, the damage index score did not change significantly (61). In a prospective study of 141 patients with SLE, components of the SDI were shown to be valid and associated with the immunosuppressive medication score. SDI scores showed a weak relationship with the BILAG activity index (SRCC 0.25-0.28; \( P < 0.01 \)). Scores for the musculoskeletal system, which are associated with limitation in physical functioning and measured with the 20-item Medical Outcomes Study Short form, as well as for renal damage varied inversely with fatigue (SRCC −0.23; \( P < 0.01 \)). Renal and neuropsychiatric damages were shown to be significantly associated with the medication score (SRCC 0.27 and 0.23, respectively; \( P < 0.01 \)) (65).

**Responsiveness.** In a large-sample multicentric study, damage scores were shown to increase with time, irrespective of the initial scores, and higher scores early in the course of the disease correlated with mortality (1.56 vs 0.99; \( P = 0.0003 \)) (66).

**Minimally important differences.** A cohort study of 80 patients with SLE showed that an SDI score of 2 or more at 5 years increased the relative risk for fatality by 3.4 (95% CI 1.5-14.4) and had a predictive value of 38% (67).

**Generalizability.** The SDI is the most widely used index for measuring damage in patients with lupus.

**Use in clinical trials.** The SDI is rarely used in RCTs but is an important outcome measure of interest in SLE. The recently published open-label extension of belimumab showed stable SDI scores in an 8-year follow-up (68).

**Critical appraisal of overall value to the rheumatology community**

**Strengths.** The SDI is a valid and reliable measure for assessing damage in patients with SLE, which has been shown to be associated with adverse outcomes, including mortality.

**Caveats.** Component scores, rather than the total score, are more relevant in defining prognosis. The renal component of the SDI has been shown to predict renal failure, and the pulmonary component predicts mortality; hence, instead of a focus on the total score, component scores should be recorded and followed (69). The index does not work as well for children and young adults, in whom growth retardation needs to be considered.

**Clinical usability.** Measurement of damage in patients with SLE is a distinct measure (different and separate from disease activity) and provides valuable and additional information in determining clinical prognosis.

**Research usability.** The SDI is useful both as a descriptor of patient populations included in studies and as an outcome measure for therapeutic trials and studies of prognosis.

**BRIEF INDEX OF LUPUS DAMAGE**

**Description**

**Purpose.** The BILD was described in 2011 by Yazdany et al (70) as a PRO measure of damage in patients with SLE for use in population studies.

**Content.** The BILD is a proxy measure of the SDI to distinguish between degrees of damage in SLE in which the existing SDI items have been modified to be comprehensible by patients. Items thought to be too rare or not likely to be interpreted with enough specificity, such as alopecia, were removed.

**Number of items.** The BILD contains 26 of the original SDI items. The initial questionnaire had 28 questions, of which 2 were dropped after initial validation.

**Response options.** Response options are “yes” and “no.” Responses that are not marked are recorded as “do not know.”

**Cost to use.** No additional costs.

**How to obtain.** Downloadable from supplementary appendix of the article by Yazdany et al (70).

**Practical application**

**Method of administration.** The BILD is administered by an interviewer either in person or by telephone.

**Scoring.** The BILD is scored from 0 to 31. Items are scored if present, with an additional score for recurrent events, such as multiple strokes. ESRD or being on dialysis is given a score of 3.

**Score interpretation.** A score of 0 implies no damage, and higher scores represent more accrued damage.

**Respondent time to complete.** Up to 10 minutes.

**Administrative burden.** The original version is administered by an interviewer, but modifications with self-administered versions have also been validated and have low administrative burden.

**Translations/adaptations.** The self-administered version of the BILD has been described and validated in a predominantly African American cohort of US patients with SLE (71).
self-administered version has also been translated into German and has shown comparable validity and high correlation with physician-reported damage (72).

**Psychometric information**

**Floor and ceiling effects.** In the LOS, the median BILD score was 1, with an interquartile range of 0 to 3 and a maximum score of 6 (70).

**Reliability.** No formal studies to demonstrate reliability of the interviewer-administered BILD have been performed; however, the self-administered version has been shown to be highly reliable (test-retest correlation score 0.93; \( P < 0.0001 \)) (71).

**Validity.** Criterion validity was shown in the initial validation study using 81 patients with SLE, which showed 75% to 100% agreement between BILD and SDI items. Prevalence-adjusted, bias-adjusted \( \kappa \) (PABAK) ranged from 0.70 to 1.00, except for deforming or erosive arthritis (PABAK 0.68) and extensive scarring/panniculus (PABAK 0.51), which were eventually dropped from the questionnaire (70). The BILD and SDI also had a moderately high SRCC of 0.64 (\( P < 0.001 \)). Construct validity was demonstrated in the LOS sample, with higher BILD scores seen among older individuals, those with longer disease duration, and those with a higher mean disease activity in the preceding 4 years. In addition, higher BILD scores were shown to be associated with poorer self-rated health and functional status, greater unemployment and work disability, and increased health care use (70). The self-administered version of the BILD has been shown to have good criterion validity when compared with the SDI (SRCC 0.59; \( P < 0.0001 \)), with significant associations in the expected directions with age, disease duration, disease activity, overall health, comorbidity index, and physician visits (71).

**Responsiveness.** The BILD scores were shown to increase in a large longitudinal cohort of patients with SLE (\( n = 958 \)), which showed that the mean annual change in BILD scores was 0.21 points. Changes in BILD scores ranged from 0 to 13 with a median of 1 (interquartile range 0-1). Half of the sample reported at least one additional manifestation of damage in the second BILD administration (5 years later) (73).

**Minimally important differences.** BILD scores of 2 (hazard ratio [HR] 6.1; 95% CI 1.3-30.0) and 3 or more (HR 10.8; 95% CI 2.5-46.2) have been shown to be associated with higher risk of death (73).

**Generalizability.** Generalizability has not been assessed, and the BILD does not seem to be widely used.

**Use in clinical trials.** The BILD is not routinely included in RCTs.

**Critical appraisal of overall value to the rheumatology community**

**Strengths.** The BILD is a cost-effective tool to quantify organ damage in patients with SLE to better understand the burden of the disease at the population level.

**Caveats.** The original version is administered by an interviewer; the application of which may be limited depending on resource availability; self-administered versions may be used in such scenarios. Both versions have the potential of self-report bias.

**Clinical usability.** The BILD is not time consuming and is easily applicable to routine care. It may help patients to have a better understanding of their disease. In general, when PROs are used, patients may be more actively involved in the processes of care.

**Research usability.** It is feasible to use the BILD for measuring damage in SLE in large-scale epidemiological studies in which the SDI has limited utility.

**SYSTEMIC LUPUS ERYTHEMATOSUS RESPONDER INDEX 4**

**Description**

**Purpose.** Composite responder indices have been developed as outcome measures to be used in clinical trials of therapeutic agents. The SRI is the initial responder index that was developed based on data from the belimumab phase II trials (74). The SRI and its modified version, SRI 4, are anchored to the SLEDAI, in contrast to the BICLA, which is anchored to the BILAG index.

**Content.** Requires improvement of disease activity without worsening of the overall condition or specific organ systems.

**Number of items.** Requires scoring of the SLEDAI 2K, BILAG index, and PGA (75).

**Response options.** Response options are based on the individual scales scored (see above for the SLEDAI 2K and BILAG index). PGA is a VAS with three benchmarks for assessing disease activity over the last 2 weeks. Mild flare is scored as 1 point, moderate flares are scored as 2 to 2.5 points, and severe flares are scored as 3 points on the 0 to 3 analog scale (55).
**Cost to use.** No additional costs.

**How to obtain.** Each individual scale needs to be obtained to derive the SRI 4.

**Practical application**

**Method of administration.** All components are administered by physicians.

**Scoring.** See scoring of the SLEDAI 2K and BILAG 2004 index in previous sections. An increase of 0.3 points or more (greater than 10% on the 3-point VAS) on PGA is considered clinically significant worsening of disease.

**Score interpretation.** The SRI 4 requires an improvement of 4 points or more in the SLEDAI 2K score, no new BILAG 2004 index A score or more than one new B domain score, and no clinically significant worsening (0.3 or more) in PGA.

**Respondent time to complete.** The SRI 4 is dependent on the time to complete individual scores and, hence, is time consuming at 45 to 60 minutes.

**Administrative burden.** Laboratory data are also needed to complete the SLEDAI 2K and BILAG 2004 index.

**Translations/adaptations.** The original SRI includes an improvement of 4 points or more in the SELENA-SLEDAI score and no worsening on the BILAG index and PGA (74). The SRI 5 is another modification of the SRI, with an improvement of 5 points or more in the SELENA-SLEDAI score and the same parameters for the BILAG index and PGA (76).

**Psychometric information**

**Floor and ceiling effects.** Not assessed.

**Reliability.** Not assessed.

**Validity and responsiveness.** The SRI 4 response has been shown to be associated with global clinical benefit based on a post hoc analysis of 736 patients with SLE from two RCTs of belimumab in the treatment of nonrenal SLE (77). In the study, SRI 4 responders achieved significantly greater improvements in clinical outcome measures (including percentages of patients with a greater than or equal to 7-point reduction in the SLEDAI 2K score, a BILAG 2004 index A or 2B flare rate, an oral corticosteroid reduction to less than or equal to 7.5 mg/dl, a change from baseline in PGA, and numbers of SLEDAI 2K organ domains with improvement) as well as in PRO measures (PaGA, the Functional Assessment of Chronic Illness Therapy−Fatigue, and the 36-item Short Form Health Survey [SF-36] physical component summary, mental component summary, vitality domain scores) compared with nonresponders. In patients with abnormal serology, SRI 4 responders had numerically greater improvements (baseline to week 52) in anti-double-stranded DNA concentrations compared with nonresponders ($P = 0.051$).

**Minimally important differences.** Not assessed.

**Generalizability.** Not assessed.

**Use in clinical trials.** The SRI 4 is increasingly used in RCTs (75).

**Critical appraisal of overall value to the rheumatology community**

**Strengths.** The SRI 4 is a composite index that offers comprehensive assessment of SLE disease because it combines advantages from three validated measurement tools. The SLEDAI 2K covers global disease improvement, the BILAG 2004 index covers organ-specific disease worsening or improvement, and PGA incorporates the physician global perspective.

**Caveats.** A major caveat of use of the SRI 4 is the time spent in scoring multiple individual indices to get a composite score. An additional caveat is that the SRI 4, being anchored to the SLEDAI, cannot capture partial improvement in a given organ system. This is an important limitation to consider for a responder index, which may explain some of the discrepant findings observed in the anifrolumab trial (TULIP I) (78). In this study, the primary end point was not reached according to the SRI, but it was reached according to the BICLA.

**Clinical usability.** Although it detects improvements in both clinical disease manifestations and SLE-related laboratory abnormalities, given its time-consuming aspects, it is less applicable in daily clinical practice.

**Research usability.** Despite its limitations, high threshold set by the SRI 4 is relevant in clinical trials to detect a clinically meaningful improvement at the group level in assessing a new therapeutic agent.

**BILAG-BASED COMPOSITE LUPUS ASSESSMENT**

**Description**

**Purpose.** The BICLA is a BILAG-based composite index originally derived by expert consensus of disease activity indices. The BICLA response was the primary end point in the EMBLEM
Study (Phase IIb Study of Epratuzumab in Serologically Positive Lupus Patients; ClinicalTrials.gov Identifier: NCT00624351) (79).

**Content.** The BICLA requires improvement in organ system scores and no worsening in organ-based or overall disease activity scores.

**Number of items.** The BICLA combines the BILAG 2004 index, SLEDAI 2K, and PGA but is driven primarily by the BILAG 2004 index.

**Response options.** Based on the individual index (see above).

**Cost to use.** No additional costs.

**How to obtain.** Each individual measure needs to be obtained.

**Practical application**

**Method of administration.** All scores are completed by the physician.

**Scoring.** Separate scoring for each index.

**Score interpretation.** Requirements for a response are 1) improvement of the BILAG 2004 index score from A to B, C, or D or improvement from B to C or D; 2) no worsening (no single new BILAG 2004 index A score or two new BILAG 2004 index B scores, no worsening of the baseline SLEDAI total score, and no worsening in PGA [less than 10% worsening relative to baseline]); and 3) no treatment failure (treatment failure defined by the use of nonprotocol treatment in the original study).

**Respondent time to complete.** The BICLA is time consuming, requiring 45 to 60 minutes (similar to the SRI 4).

**Administrative burden.** Laboratory data are needed to complete individual measures.

**Translations/adaptations.** No adaptations.

**Psychometric information**

**Floor and ceiling effects.** Not assessed.

**Reliability.** Not assessed.

**Validity.** The BICLA response was the primary end point in the EMBLEM study. In this study, the BICLA showed meaningful and sustained improvements in PaGA and PGA of disease activity, in the SF-36, and in quality of life and reductions in the corticosteroid doses (79).

**Responsiveness.** The BICLA and SRI 4 were compared against physician-rated improvement (PRI) in a cohort of 91 patients with SLE. Not accounting for medication changes, the SRI 4 had better agreement with PRI than the BICLA (85% vs 76%). Most patients with improvement by PRI also had improvement by the SRI 4 but not by the BICLA. The BICLA had slightly better specificity compared with the SRI 4 (78% vs 74%) (80).

**Minimally important differences.** Not assessed.

**Generalizability.** A post hoc analysis of data from the EMBLEM study (79) using the SRI 4 as the responder index showed higher rates of responses in the placebo group compared with BICLA responders, again suggesting better specificity of the BICLA to ascertain drug response (75,79). In that study, the SRI 4 was driven by baseline distribution of items with high SLEDAI weights. It is notable, as previously mentioned, that in the phase III TULIP I trial of anifrolumab using the SRI 4, the primary end point was not reached (78). However, a post hoc analysis showed that it would have been successful if the BICLA had been chosen. The BICLA was thus the preferred primary end point for TULIP 2, which did meet its primary end point.

**Use in clinical trials.** Previously reviewed (EMBLEM and TULIP 2).

**Critical appraisal of overall value to the rheumatology community**

**Strengths.** The BICLA is anchored on the BILAG index and thus captures partial improvement in each of the nine organ systems.

**Caveats.** The BICLA requires improvement in every active organ system, which may limit its sensitivity, especially when more organs are involved at baseline (80).

**Clinical usability.** Time-consuming aspects to score each component of the BICLA make it less useful for clinical practice.

**Research usability.** Use of composite end points may yield greater power than individual items to identify differences between treatment groups, which are particularly useful in clinical trials in SLE. However, some researchers in the lupus community may prefer to use organ-specific end points, such as the CLASI and joint counts.
CONCLUSIONS

In summary, most composite indices to assess disease activity and damage in patients with SLE have been extensively validated and are comparable. Most of them are complex and require training. However, when applied in research and clinical practice, these indices facilitate the collection of relevant clinical information quantitatively. Composite indices to assess both disease activity and cumulative damage may assist clinicians in guiding therapeutic decisions because PGAs exhibit great variability across physicians. This variability across lupus expert clinicians has been illustrated in a study of 54 patients with SLE in which significant difference was noted between pre- and postlaboratory PGA (81). Although greater correlation between postlaboratory PGA and the SELENA-SLEDAI score was observed, significant variability in PGA was present even by the same provider.

Both the BILAG-2004 index and SLEDAI-2K are the indexes with the most complete validation and are the most commonly used in RCTs and cohort studies. The SLAQ has the advantage of being the easiest to calculate because it is a patient self-reported index. It is mainly used in population studies and may be feasible for use in clinical practice; however, the SLAQ reflects the patient’s perspective, and distinguishing clinical features due to activity rather than damage or concomitant disease, such as chronic pain syndrome, fibromyalgia, or depression, may not be easy. Thus, it shows low correlation with minimized PGA, BILAG index, and SLEDAI scores (41). Organ-specific indices, such as the CLASI, joint counts, and the SFI (which measures flares), find their use in therapeutic clinical trials in SLE.

The SDI and its brief version, the BILD, measure irreversible damage from onset of lupus, regardless of the cause. The SDI shows good reliability when completed by a different physician, based on a retrospective medical history of the patient, and the BILD shows moderate correlation with the SDI ($r = 0.54$). SLE experts have suggested that the failure of some important clinical trials in SLE may be attributed, in part, to the insensitivity of available indices to detect a response to therapy. This consideration has led to development of the responder indices described in this review.

In conclusion, choosing the most appropriate measure in each situation mainly depends on the context in which it will be used and the question to be answered in terms of disease and damage assessment. A better knowledge of the measurement properties for these indices may help an investigator or clinician decide which one would be more useful in a specific scenario. For a summary of each measure’s practical applications and psychometric information, see Tables 1 and 2.

AUTHOR CONTRIBUTIONS

All authors drafted the article, revised it critically for important intellectual content, and approved the final version to be published.

REFERENCES

1. Castrejón I, Tani C, Jolly M, Huang A, Mosca M. Indices to assess patients with systemic lupus erythematosus in clinical trials, long-term observational studies, and clinical care. Clin Exp Rheumatol 2014;32 Suppl 85:S95–95.

2. Romero-Diaz J, Isenberg D, Ramsey-Goldman R. Measures of adult systemic lupus erythematosus: updated version of British Isles Lupus Assessment Group (BILAG) 2004, European Consensus Lupus Activity Measurements (ECLAM), Systemic Lupus Activity Measure, Revised (SLAM-R), Systemic Lupus Activity Questionnaire for Population Studies (SLAQ), Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K), and Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S37–46.

3. Alarcon GS, Mcqwin G Jr, Bastian HM, Roseman J, Lisse J, Fessler BJ, et al, and the LUMINA Study Group. Systemic lupus erythematosus in three ethnic groups. VII. Predictors of early mortality in the LUMINA cohort [published erratum appears in Arthritis Rheum 2001;45:306]. Arthritis Rheum 2001;45:191–202.

4. Bertoil AM, Fernandez M, Alarcon GS, Vila LM, Reveille JD. Systemic lupus erythematosus in a multiethnic US cohort LUMINA (XLI): factors predictive of self-reported work disability. Ann Rheum Dis 2007;66:6–7.

5. Rahman P, Gladman DD, Urowitz MB, Hallett D, Tam LS. Early damage as measured by the SLICC/ACR damage index is a predictor of mortality in systemic lupus erythematosus. Lupus 2001;10:93–6.

6. Castrejón I, Carmona L, Agrinier N, Andres M, Briot K, Caron M, et al. The EULAR Outcome Measures Library: development and an example from a systematic review for systemic lupus erythematosus instruments. Clin Exp Rheumatol 2015;33:910–6.

7. Mahieu M, Yount S, Ramsey-Goldman R. Patient-reported outcomes in systemic lupus erythematosus. Rheum Dis Clin North Am 2016;42:253–63.

8. Castrejón I, Gosses L, Carmona L. The EULAR Outcome Measures Library: an evolutional database of validated patient-reported instruments. Ann Rheum Dis 2015;74:475–6.

9. Isenberg DA, Rahman A, Allen E, Farewell V, Akil M, Bruce IN, et al. BILAG 2004: development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. Rheumatology (Oxford) 2005;44:902–6.

10. 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.

11. Yee CS, Gordon C, Isenberg DA, Griffiths B, Teh LS, Bruce IN, et al. The BILAG-2004 systems tally: a novel way of representing systemic lupus erythematosus in three ethnic groups. VII. Predictors of early mortality in systemic lupus erythematosus: updated version of British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. Rheumatology (Oxford) 2012;51:1877–80.

12. Murphy CL, Yue CS, Gordon C, Isenberg DA. From BILAG to BILAG-based combined lupus assessment: 30 years on. Rheumatology (Oxford) 2016;55:1357–63.

13. Gordon C, Sutcliffe N, Skan J, Stoll T, Isenberg DA. Definition and treatment of lupus flares measured by the BILAG index. Rheumatology (Oxford) 2003;42:1372–9.
16. Yee CS, Farewell V, Isenberg DA, Prabu A, Sokoli K, Teh LS, 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.

17. Yee CS, Farewell V, Isenberg DA, Rahman A, Teh LS, Griffiths B, 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.

18. Yee CS, Farewell V, Isenberg DA, Griffiths B, Teh LS, Bruce IN, et al. The BILAG-2004 index is sensitive to change for assessment of SLE disease activity. Rheumatology (Oxford) 2009;48:691–5.

19. McCarthy EM, Sutton E, Nesbit S, White J, Parker B, Jayne D, et al. Short-term efficacy and safety of rituximab therapy in refractory systemic lupus erythematosus: results from the British Isles Lupus Assessment Group Biologics Register. Rheumatology (Oxford) 2018;57:470–9.

20. Rua-Figueroa I, Lopez-Longo FJ, Calvo-Alen J, Galindo-Izquierdo M, Loza E, Garcia de Yebenes MJ, et al. National registry of patients with systemic lupus erythematosus of the Spanish Society of Rheumatology: objectives and methodology. Reumatol Clin 2014;10:17–24.

21. Isenberg DA, Allen E, Farewell V, D’Cruz D, Alarcon GS, Aranow C, et al. An assessment of disease flare in patients with systemic lupus erythematosus: a comparison of BILAG 2004 and the flare version of SELENA. Ann Rheum Dis 2011;70:54–9.

22. Gladman DD, Ibanez D, Urowitz MB. Systemic Lupus Erythematosus Disease Activity Index 2000. J Rheumatol 2002;29:288–91.

23. Bombardier C, Gladman DD, Urowitz MB, Caron D, Chang CH, et al. Systemic Lupus Erythematosus Disease Activity Index: a disease activity index for lupus patients. Arthritis Rheum 1992;35:630–40.

24. Polachek A, Gladman DD, Su J, Urowitz MB. Defining low disease activity in systemic lupus erythematosus. J Rheumatol 2004;31:1934–40.

25. Franklyn K, Lau CS, Navarra SV, Louthrenoo W, Lateef A, Hamijoyo GS. The Systemic Lupus Activity Measure- revised, the Mexican Systemic Lupus Erythematosus Disease Activity Index (SLEDAI): a disease activity index for lupus patients. Arthritis Rheum 2011;65:165–71.

26. Segovia D. Measurement of disease activity in systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2017;69:997–1003.

27. Franklyn K, Lau OS, Navarra SV, Louthrenoo W, Lateef A, Hamijoyo L, et al. Definition and initial validation of a Lupus Low Disease Activity State (LLDAS). Ann Rheum Dis 2016;75:1615–21.

28. Guzman J, Cardiel MH, Arce-Salinas A, Sanchez-Guerrero J, Alarcon-Segovia D. Measurement of disease activity in systemic lupus erythematosus: prospective validation of 3 clinical indices. J Rheumatol 1992;19:1551–8.

29. Petri M, Hellmann D, Hochberg M. Validity and reliability of lupus activity measures in the routine clinic setting. J Rheumatol 1992;19:53–9.

30. Buyon JP, Kalunian KC, Ramsay-Goldman R, Petri MA, Lockshin MD, Ruiz-Indartza G, et al. Assessing disease activity in SLE patients during pregnancy. Lupus 1999;8:677–84.

31. Touma Z, Urowitz MB, Ibanez D, Gladman DD. SLEDAI-2K 10 days versus SLEDAI-2K 30 days in a longitudinal evaluation. Lupus 2011;20:67–70.

32. Touma Z, Urowitz MB, Gladman DD. SLEDAI-2K for a 30-day window. Lupus 2010;19:49–51.

33. Uribe AG, Vila LM, McGavin G Jr, Sanchez ML, Reveille JD, Alarcon GS. The Systemic Lupus Erythematosus Activity Measure-revised, the Mexican Systemic Lupus Erythematosus Disease Activity Index (SLEDAI), and a modified SLEDAI-2K are adequate instruments to measure disease activity in systemic lupus erythematosus. J Rheumatol 2004;31:1934–40.

34. Fatemi A, Raesi A, Sayedbonakadar Z, Smiley A. Sensitivity analyses of four systemic lupus erythematosus disease activity indices in predicting the treatment changes in consecutive visits: a longitudinal study. Clin Rheumatol 2018;37:955–62.

35. Gladman DD, Urowitz MB, Kagal A, Hallett D. Accurately describing changes in disease activity in systemic lupus erythematosus. J Rheumatol 2000;27:377–9.
50. Jolly M, Kazmi N, Mikolaitis RA, Sequiera W, Block JA. Validation of the Cutaneous Lupus Disease Area and Severity Index (CLASI) using physician- and patient-assessed health outcome measures. J Am Acad Dermatol 2013;68:618–23.

51. Bonilla-Martinez ZL, Albrecht J, Troxel AB, Taylor L, Okawa J, Dulay S, et al. The cutaneous lupus erythematosus disease area and severity index: a responsive instrument to measure activity and damage in patients with cutaneous lupus erythematosus. Arch Dermatol 2008;144:173–23.

52. Merrill JT, Furie R, Werth VP, Khamashta M, Drappa J, Wang L, et al. Anifrolumab effects on rash and arthritis: impact of the type I interferon gene signature in the phase IIb MUSE study in patients with systemic lupus erythematosus. Lupus Sci Med 2018;5:e000284.

53. Iaccarino L, Bettio S, Reggia R, Zen M, Frassi M, Andreoli L, et al. Effects of belimumab on flare rate and expected damage progression in patients with active systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2017;69:115–23.

54. Ugarte A, Porta S, Rios R, Martinez-Zapico A, Ortego-Centeno N, Agesta N, et al. Combined mepracine-hydroxychloroquine treatment in patients with systemic lupus erythematosus and refractory cutaneous and articular activity. Lupus 2018;27:1718–22.

55. Petri M, Buyon J, Kim M. Classification and definition of major flares in SLE clinical trials. Lupus 1999;8:685–91.

56. Petri M, Kim MY, Kalunian KC, Grossman J, Hahn BH, Sammaritano LR, et al. Combined oral contraceptives in women with systemic lupus erythematosus. N Engl J Med 2005;353:2550–8.

57. Thanou A, Chakravarty E, James JA, Merrill JT. How should lupus flares be measured? Deconstruction of the safety of estrogen in lupus erythematosus national assessment-systemic lupus erythematosus disease activity index flare index. Rheumatology (Oxford) 2014;53:2175–81.

58. Corneloup M, Maurier F, Wahl D, Muller G, Aumaitre O, Seve P, et al. Disease-specific quality of life following a flare in systemic lupus erythematosus: an item response theory analysis of the French EQUAL cohort. Rheumatology (Oxford) 2020;59:1398–1406.

59. Mathian A, Pha M, Haroche J, Cohen-Aubart F, Hie M, Pineton de Chambrun M, et al. Withdrawal of low-dose prednisone in SLE patients with a clinically quiescent disease for more than 1 year: a randomised clinical trial. Ann Rheum Dis 2020;79:339–46.

60. Iaccarino L, Bettio S, Reggia R, Zen M, Frassi M, Andreoli L, et al. Effects of belimumab on flare rate and expected damage progression in patients with active systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2017;69:115–23.

61. Gladman D, Ginzler E, Fortin P, Liang M, Urowitz M, et al. The development and initial validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index for systemic lupus erythematosus. Arthritis Rheum 1996;39:363–9.

62. Gladman DD, Urowitz MB, Goldsmith CH, Fortin P, Ginzler E, Gordon C, et al. The reliability of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index in patients with systemic lupus erythematosus. Arthritis Rheum 1997;40:809–13.

63. Costenbader KH, Khamashta M, Ruiz-Garcia S, Perez-Rodriguez MT, Petri M, Elliott J, et al. Development and initial validation of a self-assessed lupus organ damage instrument. Arthritis Care Res (Hoboken) 2010;62:559–68.

64. Pons-Estel BA, Sanchez-Guerrero J, Romero-Diaz J, Iglesias-Gamarr A, Bonfa E, Borba EF, et al. Validation of the Spanish, Portuguese and French versions of the Lupus Disease Damage Questionnaire: data from North and South America, Spain and Portugal. Lupus 2009;18:1033–52.

65. Stoll T, Stucki G, Malik J, Pyke S, Isenberg DA. Association of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index with measures of disease activity and health status in patients with systemic lupus erythematosus. J Rheumatol 1997;24:309–13.

66. Gladman DD, Goldsmith CH, Urowitz MB, Bacon P, Fortin P, Ginzler E, et al. The Systemic Lupus International Collaborating Clinics/American College of Rheumatology (SLICC/ACR) Damage Index for systemic lupus erythematosus international comparison. J Rheumatol 2000;27:373–6.

67. Nived O, Jonsen A, Bengtsson AA, Bengtsson C, Sturfelt G. High predictive value of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index for survival in systemic lupus erythematosus. J Rheumatol 2002;29:1398–400.

68. Van Vollenhoven RF, Navarra SV, Levy RA, Thomas M, Heath A, Ruste T, et al. Long-term safety and limited organ damage in patients with systemic lupus erythematosus treated with belimumab: a phase IIIb study extension. Rheumatology (Oxford) 2020;59:281–91.

69. Stoll T, Seifert B, Isenberg DA. SLICC/ACR Damage Index is valid, and renal and pulmonary organ scores are predictors of severe outcome in patients with systemic lupus erythematosus. Br J Rheumatol 1996;35:248–54.

70. Yazdany J, Trupin L, Gansky SA, Dall’era M, Yelin EH, Criswell LA, et al. Brief index of lupus damage: a patient-reported measure of damage in systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2011;63:1170–7.

71. Drankard C, Yazdany J, Trupin L, Katz PP, Duniop-Thoms C, Bao G, et al. Validity of a self-administered version of the brief index of lupus damage in a predominantly African American systemic lupus erythematosus cohort. Arthritis Care Res (Hoboken) 2014;66:889–96.

72. Chehab G, Sander O, Richter J, Acar H, Vordenbaumen S, Brinks R, et al. Evaluation and validation of the German Brief Index of Lupus Damage (BILD): a self-reported instrument to record damage in systemic lupus erythematosus. Lupus 2013;22:1050–5.

73. Katz P, Trupin L, Rush S, Yazdany J. Longitudinal validation of the Brief Index of Lupus Damage. Arthritis Care Res (Hoboken) 2014;66:1057–62.

74. Furie RA, Petri MA, Wallace DJ, Ginzler EM, Merrill JT, Stohl W, et al. Novel evidence-based systemic lupus erythematosus responder index. Arthritis Rheum 2009;61:1143–51.

75. Furie R, Khamashta M, Merrill JT, Werth VP, Kalunian K, Brohawn P, et al. Anifrolumab, an anti-interferon-α receptor monoclonal antibody, in moderate-to-severe systemic lupus erythematosus. Arthritis Rheumatol 2017;69:376–86.

76. Furie RA, Leon G, Thomas M, Petri MA, Chu AD, Hsiop C, et al. A phase 2, randomised, placebo-controlled clinical trial of blisibimod, an inhibitor of B cell activating factor, in patients with moderate-to-severe systemic lupus erythematosus, the PEARL-SC study. Ann Rheum Dis 2015;74:1667–75.

77. Furie R, Wang L, Illei G, Drappa J. Systemic Lupus Erythematosus (SLE) Responder Index response is associated with global benefit for patients with SLE. Lupus 2018;27:955–62.

78. Morand EF, Furie R, Tanaka Y, Bruce IN, Askasian AD, Richez C, et al. Trial of anifrolumab in active systemic lupus erythematosus. N Engl J Med 2020;382:211–21.

79. Wallace DJ, Kalunian K, Petri MA, Strand V, Houssiau FA, Pike M, et al. Efficacy and safety of spratuzumab in patients with moderate/severe active systemic lupus erythematosus: results from EMBLEM, a phase IIb, randomised, double-blind, placebo-controlled, multi-centre study. Ann Rheum Dis 2014;73:183–90.

80. Thanou A, Chakravarty E, James JA, Merrill JT. Which outcome measures in SLE clinical trials best reflect medical judgment? [original manuscript]. Lupus Sci Med 2014;1:e000005.

81. Aranow C. A pilot study to determine the optimal timing of the Physician Global Assessment (PGA) in patients with systemic lupus erythematosus. Immunol Res 2015;63:167–9.
| Table 1. Practical applications* |
|--------------------------------|
| **Measure**                  | **Number of Items** | **Content/Domains** | **Method of Administration** | **Recall Period** | **Response Format** | **Range of Scores** | **Score Interpretation** | **Availability of Normative Data** | **Cross-Cultural Validation** |
| Disease activity indices     |                    |                    |                              |                   |                    |                    |                               |                               |                        |
| BILAG 2004 Index             | 97                 | 9                  | Physician                    | 4 wk              | 0-4                | 0-81               | A-E                           | Yes                            | Yes                        |
| SLEDAI 2K                    | 24                 | 9                  | Physician                    | 10 d              | Present/absent     | 0-105              | Remission 0, LDA >3, HDA >6 | Yes                            | Yes                        |
| SLAQ                         | 24                 | 3                  | Patient reported             | 3 mo              | 0-3                | 0-44               | SLAQ score of 9 = SLAM score of 3 | Yes                            | Yes                        |
| LFA-REAL                     | 7                  | Organ-based PGA    | Physician                    | Given day         | VAS landmark: 0 = none, 1 = mild, 2 = moderate, 3 = severe | 0-100 mm |                               | Yes                            | No                         |
| CLASI                        | 5 (activity), 3 (damage) | Cutaneous activity and damage score | Physician | Given day; 12 mo (dyspigmentation) | 0/1 (absence/presence); 0-2, 3, or 6 (severity) | 0-70 (activity), 0-84 (damage) | Mild: 0-9; moderate: 10-20; severe: 21-70 | Yes                            | No                         |
| SFI                          | 4                  | Change in SELENA-SLEDAI, PGA, additional items, and treatment modifications | Physician | 10 d | 3 SLEDAI or 1 PGA; mild/moderate flare: ≥12 SLEDAI or ≥2.5 PGA; severe flare | Present/absent | No, mild-moderate, or severe flare | Yes                            | Yes                        |
| Cumulative damage indices    |                    |                    |                              |                   |                    |                    |                               |                               |                        |
| SDI BILD                     | 41                 | 12                 | 11 organ systems             | Physician         | 6 mo               | –                  | 0-3 | Yes, no, do not know (0-3) | 0-47 | 0-31 | Damage ≥ 1 | Damage ≥ 1 | Yes | Yes | Yes |                        |
| Responder indices            |                    |                    |                              |                   |                    |                    |                               |                               |                        |
| SRI 4                        | Composite index (3) | ≥4 SLEDAI 2K, no new A or B, PGA change < 0.3 | Physician | Variable | Variable | Response: + or − | Responder, nonresponder | Yes                            | No                         |
| BICLA                        | Composite index A to B/C/D, B to C/D, no new A or B, no SLEDAI 2K/PGA worsening, no treatment failure | Physician | Variable | Variable | Response: + or − | Responder, nonresponder | Yes                            | No                         |

* BICLA = BILAG-based Combined Lupus Assessment; BILAG 2004 = British Isles Lupus Assessment Group; BILD = Brief Index of Lupus Damage; CLASI = Cutaneous Lupus Erythematosus Disease Area and Severity Index; HDA = high disease activity; ISM = immunosuppressive medication; LDA = low disease activity; LFA-REAL = Lupus Foundation of America Rapid Evaluation of Activity in Lupus; PGA = Physician Global Assessment; SELENA = Safety of Estrogens in Lupus National Assessment; SFI = SELENA-SLEDAI Flare Index; SLAM = Systemic Lupus Activity Measure; SLAQ = Systemic Lupus Activity Questionnaire; SLEDAI = Systemic Lupus Erythematosus Disease Activity Index; SLEDAI 2K = Systemic Lupus Erythematosus Disease Activity Index 2000; SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SRI 4 = Systemic Lupus Erythematosus Responder Index 4; VAS = visual analog scale.
### Table 2. Psychometrics*

| Measure               | Floor, Ceiling Effects | Reliability | Validity | Responsiveness | Minimally Important Differences | Generalizability | Used in RCTs |
|-----------------------|------------------------|-------------|----------|----------------|----------------------------------|-----------------|-------------|
| BILAG-2004 Index      | New A or B score       | ICC > 0.60  | OR 2.9 (ESR 60), OR 2.7 (d/dNA rise), OR 5 (C3 decrease), OR 4.2 (C4 decrease), OR 19.3 (increase in medication) | Coefficient 1.35 (increase in therapy), coefficient 0.44 (decrease in therapy) | Severe flare: any A; moderate flare: new B or 2B | Yes            |
|                       |                        |             |          |                | A = 12, B = 8, C = 1, D = 0     |                 |
| SLEDAI 2K             | Mean (SD) 10.4, remission 30.8%, LDA 12.9%, HDA 56.3% | 81.7%-100% agreements | Correlation coefficient 0.677 (PGA) | AUC = 0.771 (treatment change) | AUC 0.697, 5 28.8% (improvement); AUC 0.877, 5 35.3% (worsening) | Yes            |
|                       |                        |             |          |                | Generalizable                   |                 |
| SLAQ                  | Almost normal distribution with a slight right skew | Cronbach α 0.87 | Correlation 0.62 | SRM 0.66 (worsening), SRM −0.37 (improvement) | PPV 56%-89% Epidemiological studies | No             |
|                       |                        |             |          |                |                                  |                 |
| LFA-REAL              | 2-4 scales             | ICC 0.79 and 0.86 | SRCC 0.903 (PGA), SRCC 0.816 (SLEDAI), SRCC 0.933 (BILAG) | SRCC 0.857 (ΔLFA-REAL and ΔPGA), SRCC 0.701 (ΔLFA-REAL and ΔSLEDAI), SRCC 0.700 (ΔLFA-REAL and ΔBILAG) | – | – | No |
|                       |                        |             |          |                |                                  |                 |
| CLASI                 | Activity: 10.1-40; damage: 0.7-41.6 | ICC 0.86 (activity), ICC 0.92 (damage) | SRCC 0.42 (CLASI activity and SLEDAI-rash), SRCC 0.51 (CLASI damage and SDI scarring) | SRCC 0.97 (CLASI A: change in physician assessment of skin health), SRCC 0.85 (CLASI A: change in patient GA), SRCC 0.98 (CLASI A: change in pain) | Mean 3- or 4-point decrease or 20% improvement in activity score | – | Yes |
|                       |                        |             |          |                |                                  |                 |
| SFI                   | 40.2%; mild/moderate flare; 5.7%; severe flare | $\kappa = 0.65$ (severe flare); $\kappa = 0.45$ (mild/moderate flare) | SRFI (BILAG): $\kappa = 0.30$, weighted $\kappa = 0.39$; rSFI (BILAG) $\kappa = 0.31$, weighted $\kappa = 0.51$; rSFI (BILAG) $\kappa = 0.31$, weighted $\kappa = 0.51$; rSFI (BILAG) $\kappa = 0.31$, weighted $\kappa = 0.51$ | Flare agreement with 3 physician assessment: 72% (SFI), 70% (rSFI) | – | – | Yes |
|                       | Rarely >12             | Observer effect ($P = 0.933$) | SRCC 0.28 (BILAG), SRCC −0.30 (MSK: MOS SF 20 PF); SRCC −0.23 (renal: MOS SF 20 fatigue); SRCC 0.27 (renal: medication score), SRCC 0.23 (NP: medication score) | OR 1.56 (high scores: mortality) | ≥2, RR (fatality) 3.4 | Widely used | Rare |
|                       |                        |             |          |                |                                  |                 |
| SDI                   | Rarely >12             | Observer effect ($P = 0.933$) | SRCC 0.28 (BILAG), SRCC −0.30 (MSK: MOS SF 20 PF); SRCC −0.23 (renal: MOS SF 20 fatigue); SRCC 0.27 (renal: medication score), SRCC 0.23 (NP: medication score) | Flare agreement with 3 physician assessment: 72% (SFI), 70% (rSFI) | – | – | Yes |
|                       |                        |             |          |                |                                  |                 | (Continued)
Table 2. (Cont’d)

| Measure | Floor, Ceiling Effects | Reliability | Validity | Responsiveness | Minimally Important Differences | Generalizability | Used in RCTs |
|---------|------------------------|-------------|----------|----------------|----------------------------------|------------------|-------------|
| BILD    | Median 1 (IQR 0-3)     | SA version  | SRCC 0.64 (SDI), SRCC 0.59 (SA version and SDI) | Annual change 0.21 | 2, HR 6.1 (risk of death), >3 or ≥3, HR 10.8 (risk of death) | –                | No          |
| SRI 4   | –                      | –           | ≥7 SLEDAI reduction 61.9 (R) vs 0.3 (NR), BILAG A or 2B flare rate 5.1 (R) vs 21.9 (NR), PGA change 71.7 (R) vs –13.6 (NR), FACIT-F change 6.5 (R) vs 0.7 (NR), SF-36 PCS change 6.3 (R) vs 1.3 (NR), SF-36 MCS change 4.3 (R) vs –0.1 (NR) | SRCC 0.563 (physician-rated improvement) | –                | –           | Yes         |
| BICLA   | –                      | –           | OR (response in drug vs placebo) 2.9 | –                | –                | Yes          |

* AUC = area under the curve; BICLA = BILAG-based Combined Lupus Assessment; BILAG-2004 = British Isles Lupus Assessment Group; BILD = Brief Index of Lupus Damage; CLASI = Cutaneous Lupus Erythematosus Disease Area and Severity Index; dsDNA = double-stranded DNA; ESR = erythrocyte sedimentation rate; FACIT-F = Functional Assessment of Chronic Illness Therapy–Fatigue; GA = Global Assessment; ICC = intraclass coefficient; IQR = interquartile range; HDA = high disease activity; HR = hazard ratio; LDA = low disease activity; LFA-REAL = Lupus Foundation of America Rapid Evaluation of Activity in Lupus; MCS = Mental Component Summary; MOS SF = 20-item Medical Outcomes Study Short Form; MSK = musculoskeletal; NR = nonresponder; NP = neuropsychiatric; PCS = Physical Component Summary; PF = physical function; PGA = Physician Global Assessment; PPV = positive predictive value; R = responder; RCTs = randomized clinical trials; RR = relative risk; rSFI = revised SFI; S = sensitivity; SA = self-administered; SELENA = Safety of Estrogens in Lupus National Assessment; SF-36 = 36-item Short Form Health Survey; SFI = SELENA-SLEDAI Flare Index; SLAM = Systemic Lupus Activity Measure; SLAQ = Systemic Lupus Activity Questionnaire; SLEDAI = Systemic Lupus Erythematosus Disease Activity Index; SLEDAI-2K = Systemic Lupus Erythematosus Disease Activity Index 2000; SRCC = Spearman rank correlation coefficient; SRM = standardized response mean; OR = odds ratio; RCT = randomized clinical trial; SDI = Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SRI 4 = Systemic Lupus Erythematosus Responder Index 4.