Defining Response in Systemic Lupus Erythematosus: A Study by the Systemic Lupus International Collaborating Clinics Group

SEAN J. WOLLASTON, VERNON T. FAREWELL, DAVID A. ISENBERG, CAROLINE GORDON, JOAN T. MERRILL, MICHELLE A. PETRI, and KENNETH C. KALUNIAN, for the Systemic Lupus International Collaborating Clinics (SLICC)

ABSTRACT. ***Objective.*** In a preliminary attempt to develop a drug responder index for patients with systemic lupus erythematosus (SLE), 2 validated disease activity instruments were studied for their responsiveness and compared to a physician visual analog scale (VAS) assessment of disease activity. We attempted to determine whether these validated instruments were useful components in characterizing response in the setting of a clinical trial.

Methods. Eighty paper patients were assessed using the British Isles Lupus Assessment Group (BILAG) and Systemic Lupus Disease Activity Index (SLEDAI) and by physician's assessment of global activity. The cases were arranged in random order and divided into groups of 20 patients and each group was assessed by 20 lupus experts; change in disease activity was recorded at 3 and 6 months compared to baseline using a physician VAS.

Results. Four different lupus experts assessed disease activity in all 80 patients at baseline and 3 and 6 months after initiation of therapy using the BILAG and SLEDAI instruments. BILAG and SLEDAI scores correlated well over time; however, in a regression analysis where average physician VAS were chosen as the outcome variable, a significant amount of variation in the average physician VAS not related to the SLEDAI and BILAG scores was noted.

Conclusion. The physician VAS may be too blunt to assess response in SLE, because even among experienced lupus assessors, there were considerable differences in what influenced scoring decisions. (J Rheumatol 2004;31:2390–4)

Key Indexing Terms:
SYSTEMIC LUPUS ERYTHEMATOSUS RESPONSE INDEX

Systemic lupus erythematosus (SLE) is an autoimmune rheumatic disease characterized by a remarkable diversity of clinical and serological features. It causes significant morbidity in most, and the immunosuppressive drugs used to control the disease cause many side effects including infection, bone marrow toxicity, and osteoporosis. More effective therapy based upon an improved understanding of its etiopathogenesis is needed with drugs less prone to cause major complications.

Agreement is needed on the methods of assessment to determine the benefits, if any, of an individual drug or combination of drugs. A consensus is needed as exciting new ideas about the treatment of SLE are increasingly being brought from the bench to the bedside. In a condition as protean in its manifestations as SLE, it is essential that investigators use comparable tools to assess the disease so that others may meaningfully attempt to reproduce claims made on behalf of these new therapies.

The Systemic Lupus International Collaborating Clinics group (SLICC), which represents 30 practicing rheumatologists from 25 different lupus centers in 10 different countries, has focused for the past few years on the task of devising a drug responder index for patients with lupus[1]. We believe that such an index should include the following elements: (1) a measurement that assesses disease activity (i.e., to record active, ongoing inflammation and active vasculopathy); (2) an assessment of damage (representing permanent organ dysfunction that has developed since the onset of the disease regardless of causality); (3) the patient's own perspective of her or his health status; and (4) a measure-

From the Division of Rheumatology, David Geffen School of Medicine at UCLA, Los Angeles, California; MRC Biostatistics Unit, University of Cambridge, Cambridge, UK; Centre for Rheumatology, University College London, London, UK; University of Birmingham, Birmingham, UK; Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma; and Division of Rheumatology, Johns Hopkins University, Baltimore, Maryland, USA.

Supported by a grant from the American College of Rheumatology.

S.J. Wollaston, MD, Division of Rheumatology, David Geffen School of Medicine at UCLA; V.T. Farewell, PhD, MRC Biostatistics Unit, University of Cambridge; D.A. Isenberg, MD Centre for Rheumatology, University College London; C. Gordon, MD, University of Birmingham; J.T. Merrill, MD, Oklahoma Medical Research Foundation; M.A. Petri, MD, MPH, Division of Rheumatology, Johns Hopkins University; K.C. Kalunian, MD, UCSD School of Medicine, La Jolla, CA.

Address reprint requests to Dr. K.C. Kalunian, UCSD Center for Innovative Therapy, 9320 Campus Point Drive #227, La Jolla, CA 92037.

Submitted April 28, 2003; revision accepted June 29, 2004.

ment of drug related side effects[2]. The OMERACT Committee is in agreement with this view[3].

In a preliminary attempt to develop a drug responder index for patients with SLE, we studied the responsiveness of 2 validated disease activity instruments in characterizing change in SLE patients undergoing different interventions compared to a physician assessment of disease activity. The study was undertaken to determine whether these validated instruments were useful components in characterizing response in the setting of a clinical trial. We analyzed a set of 80 paper patients in which the British Isles Lupus Assessment Group (BILAG) and Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) were used to characterize disease activity and a separate scale was used that was intended to determine an overall impression of disease activity from the perspective of physicians with extensive expertise in caring for patients with lupus. Although previous successful attempts have been made to compare BILAG (when converted into a global score) and SLEDAI in both real and paper patients[4,5], the numbers of patients studied were small and little effort had been made to compare the use of physician assessment scores and scores from the validated activity instruments in response to a change in a particular medical intervention.

MATERIALS AND METHODS

Eighty detailed patient histories were assembled by 2 of us (SJW, KCK) from several sources, but invariably from patients involved in different drug trials. The cases were arranged in random order and divided into 4 groups, each group consisting of 20 patients. During a workshop meeting in Barcelona prior to the International Lupus Conference in March 2001, 20 members of the SLICC group were also divided into 4 equal-size groups. Each group was asked to assess the change in disease activity in a group of 20 patients at 3 and 6 months after starting treatment compared to baseline. These assessments of disease activity were made using a 7-point Likert visual analog assessment scale (VAS). The 7 points on this Likert scale assessed disease activity as (1) much improved, (2) moderately improved, (3) slightly improved, (4) unchanged, (5) slightly worse, (6) moderately worse, and (7) much worse compared to baseline. Four different members of the SLICC group were asked to assess the disease activity in all 80 patients at baseline and at 3 and 6 months after initiation of therapy using the BILAG and SLEDAI instruments of disease activity; these raters were chosen by the entire SLICC group because they were felt to have the most clinical experience using the BILAG and SLEDAI instruments.

The degree of improvement/deterioration in VAS scores was then compared with the change (if any) in the BILAG and SLEDAI activity indices.

The BILAG activity index divides lupus activity into 8 organs/systems and is based on the principle of the physician's intention to treat[6], assessing activity in the previous one month. Each organ or system is given a score of A to E, where A denotes disease thought to be sufficiently active to require disease modifying treatment (i.e., prednisone equivalent > 20 mg/day or immunosuppressants); a B score identifies problems requiring symptomatic treatment such as antimalarials or nonsteroidal antiinflammatory drugs or prednisolone < 20 mg/day; C indicates stable mild disease; D indicates a previously affected but currently inactive system; and E indicates that the system or organ has never been involved. The BILAG score has been converted into a numerical scoring system by the following assignments: A = 9, B = 3, C = 1, D/E = 0[7]. The index measures the degree of disease activity in the 30-day period prior to the assessment.

The SLEDAI index, originally developed following a meeting in Toronto in 1985, is a global score index based upon an assessment of 24 items. It is a descriptive index that has been shown to be reliable in naive observers[8] and in routine use[9]. The index measures the degree of disease activity in the 10-day period prior to the assessment.

Statistics. The experiment design used to collect the data was a plaid square with 4 separate groups comprising 20 patients and 5 physician assessors each. Agreement in VAS scores among physicians was examined using intraclass correlation measures defined within each group. Correlations are reported along with associated confidence intervals between average VAS scores, SLEDAI scores, and BILAG scores. Regression models were used to examine in more detail the link between physician ratings and SLEDAI and BILAG scores. Usual normal theory regression was used with the average VAS score as the response variable to calculate multiple correlation coefficients. In addition, additional analyses used logistic regression, with a binary outcome indicating consensus improvement.

RESULTS

Consistency of VAS scores. In the 4 patient/clinician groups (20 patients, 5 clinicians), there was considerable variation in the VAS scores. The intraclass correlations for the clinician ratings in the 4 groups were 0.25, 0.39, 0.46, and 0.29 for the 3-month assessments. For the 6-month assessments the corresponding intraclass correlations were 0.18, 0.34, 0.34, and 0.20.

The 5 scores for each patient were used in 2 ways. A simple average of the scores was calculated and a majority score on a 3-point scale was defined based on the criterion that 3 scores in the upper 2 categories would correspond to improvement, 3 scores in the lower range would correspond to deterioration, and otherwise the classification would be no change.

3-month visit assessments. The change in the 3-month SLEDAI and BILAG scores from baseline were highly correlated, with r = 0.751 (CI 0.636, 0.833). A slightly lower correlation between average VAS scores of change from baseline and change in SLEDAI scores from baseline was observed, with r = –0.592 (CI –0.718, –0.428). The estimated correlation between the average VAS scores and change in BILAG scores from baseline was r = –0.684 (CI –0.786, –0.546). No substantial improvement in the relationship between the VAS scores and the SLEDAI and BILAG scores could be achieved through the use of combination of scores or separate BILAG organ system classifications. The estimated multiple correlation coefficient baseline and 3-month values for both SLEDAI and BILAG was 0.744.

Four patients were classified by the majority of reviewers as deteriorating, 43 were judged to have had no or little change, and 33 were classified as improved. A binary logistic regression analysis relating an improvement/no improvement classification to SLEDAI and BILAG scores led to models with characteristics similar to the normal theory linear regression models discussed above. There was, however, more evidence that SLEDAI scores added to the information provided by BILAG scores. The probability of improving estimated by the fitted model was over 0.5 for only 4 of the patients classified as having no improvement,

and was less than 0.5 for only 5 of the patients classified as having improved.

Table 1 provides descriptive statistics for the difference in SLEDAI and BILAG scores stratified on the improvement/no improvement classification. Table 2 provides a tabulation of the improvement classification by SLEDAI and BILAG scores, using as cutoff points the values of their respective 25th percentiles in the not-improved group. These cutoff points are thus defined so that 75% of the not-improved patients have changes larger than the value chosen. It can be seen that only 5 patients of the 47 classed as having no improvement have both change in SLEDAI and change in BILAG below the cutoff points, 15 have one of them below a cutoff, and 27 have both above the cutoff. For the 33 patients classified as improved, only 2 have both changes above the cutoff, 8 have one change above the cutoff, and 23 have both changes below the cutoff.

6-month visit assessments. The change in the 6-month SLEDAI and BILAG scores from baseline were significantly correlated, with r = 0.713 (CI 0.585, 0.806). A lower correlation between average VAS scores of change from baseline and change in SLEDAI scores from baseline was observed, with r = –0.523 (CI –0.666, –0.343). The estimated correlation between the average VAS scores and change in BILAG scores from baseline was r = –0.573 (CI –0.704, –0.404). As for the 3-month scores, little gain is achieved through use of the combination of scores. The estimated multiple correlation coefficient associated with separate baseline and 6-month scores for both SLEDAI and BILAG was 0.67.

Seven patients were classified by the majority of reviewers as deteriorating, 36 were judged to have had no or little change, and 37 were classified as improved. A binary logistic regression analysis relating an improvement/no improvement classification to SLEDAI and BILAG scores led to models with characteristics similar to the normal theory linear regression models discussed above. As for the 3-month assessments, there was more evidence that SLEDAI scores added to the information provided by BILAG scores. The probability of improving estimated by the fitted model was over 0.5 for 8 of the patients classified as having no improvement and was less than 0.5 for 7 of the patients classified as having improved.

Table 2. Improvement classification by SLEDAI and BILAG scores at 3 months.

	BILAG > –3	BILAG ≤ –3
Not improved at 3 months		
SLEDAI > –4	27	8
SLEDAI ≤ –4	7	5
Improved at 3 months		
SLEDAI > –4	2	5
SLEDAI ≤ –4	3	23

Table 3 provides a tabulation of the improvement classification by SLEDAI and BILAG scores from the 3-month assessments. Table 4 provides descriptive statistics for the difference in SLEDAI and BILAG scores stratified on the improvement/no improvement classification. Six patients of the 43 classed as having no improvement have both the change in SLEDAI and the change in BILAG below the cutoff points, 13 have one of them below a cutoff, and 24 have both above the cutoff. For the 37 patients classified as improved, only 2 have both changes above the cutoff, 11 have one change above the cutoff, and 24 have both changes below the cutoff.

DISCUSSION

In the original studies carried out by the SLICC group[4,5], it was shown using relatively small numbers of patients in both paper and real-patient exercises that the BILAG, SLEDAI, and Systemic Lupus Activity Measure (SLAM) disease activity indices were comparable even though derived by different groups from a different philosophical standpoint. As we now report, the correlations reported in earlier work between BILAG and SLEDAI scores are maintained if the change in the BILAG and SLEDAI scores over time is examined. This remains true having converted the BILAG letter scores into an overall global score to facilitate the comparison. However, in a regression analysis in which the average physician VAS score was chosen as the response or outcome variable, we noted a significant amount of variation in the average physician VAS score that is not related to the SLEDAI and BILAG scores.

In general, the agreement in physician VAS scores, as indicated by intraclass correlation measures, was not particularly tight. It was of interest, therefore, to look closely at

Table 1. Difference in SLEDAI scores and BILAG scores at 3 months.

	Mean	Median	Minimum	Maximum	25th Percentile	75th Percentile
Difference in SLEDAI scores						
Not improved	–0.72	0	–8	12	–4	0
Improved	–7.94	–6	–36	2	–10	–4
Difference in BILAG scores						
Not improved	–0.06	0	–8	15	–3	2
Improved	–9.00	–8	–35	1	–12	–3.5

Table 3. Difference in SLEDAI scores and BILAG scores at 6 months.

	Mean	Median	Minimum	Maximum	25th Percentile	75th Percentile
Difference in SLEDAI scores						
Not improved	–1.07	0	–16	8	–4	2
Improved	–7.70	–6	–38	0	–11	–3.5
Difference in BILAG scores						
Not improved	–1.58	–1	–12	5	–3	0
Improved	–8.30	–6	–37	0	–11.5	–3

Table 4. Improvement classification by SLEDAI and BILAG scores at 6 months.

	BILAG > –3	BILAG ≤ –3
Not improved at 6 months		
SLEDAI > –4	24	6
SLEDAI ≤ –4	7	6
Improved at 6 months		
SLEDAI > –4	2	7
SLEDAI ≤ –4	4	24

those cases where members of the group gave the widest diversity of opinion. It became clear, with respect to the physician VAS scores, that different members of the group varied markedly in their weighting of the importance of one organ/system getting worse while others got better. In addition, in patients who were clinically stable, improvement or deterioration in serological markers was more influential with regard to disease activity for some than for others. Similarly, the importance of patients remaining well (or getting worse) as their drug therapy was altered also affected the raters' scores overall. It is readily acknowledged that some of these difficulties are inherent in the use of paper patients. There is a clear difference between seeing a patient in the clinic and simply having a reasonably detailed history and examination findings provided.

Nevertheless, we feel that the physician VAS scale may be too blunt to assess response to an intervention in SLE, because even among experienced assessors of lupus patients there were considerable differences in what influenced scoring decisions. We are of the view that response must include detailed assessments of disease, which is achieved by the use of both SLEDAI and BILAG. We also believe that it is important to include other aspects of clinical change in a response instrument in addition to assessments of change in disease activity, since we noted a significant amount of variation in the average physician VAS score that was not related to the SLEDAI and BILAG scores. To quantify this variation better than by simply measuring the physician assessment score, the following concepts may be considered for inclusion in a responder index: change in quality of life (assessed by the Medical Outcomes Study Short-Form-36); change in irreversible damage due to SLE (assessed by the SLICC/ACR damage index); measurement of adverse events that are attributable to the therapeutic intervention that is the subject of observation; and change in baseline therapeutic agents (i.e., corticosteroid reduction), if consistent with the study protocol.

We have examined the use of particular cutoff points in SLEDAI and BILAG scores and tried to link these to a consensus measure of response. However, the definition of formal rules for a response index, if this proves both useful and appropriate, is better based on data from longitudinal followup of patients seen first-hand rather than based on data from medical records. These data may suggest that a meaningful change in both SLEDAI and BILAG or to equivalent instruments may define response, whereas a partial response may be characterized by a meaningful change in only one of these instruments; this may warrant further investigation. It is possible that some investigational therapies may cause improvement of one organ system and worsening or no change in others. This may differentially affect identifiable subsets of patients. Such agents can be utilized wisely by experienced clinicians to have optimal influence on the quality of life of lupus patients once these clinical benefit/risk factors are properly understood. Use of the BILAG or a similar instrument that uses a comprehensive scoring system to assess degrees of activity in each organ may be useful in this assessment. On the other hand, a more centrally-acting agent might have the potential to abrogate disease activity in multiple organs. The SLEDAI may provide a more accurate and sensitive instrument for this type of assessment. Further study of these issues as ancillary investigation in real-time trials may lead to judicious selection and improved interpretation of the instruments used in lupus clinical studies. We conclude from these initial data, however, that it would be premature at this time to limit trials of new agents for this complex, multifactorial disease to one type of outcome measurement.

ACKNOWLEDGMENT

The Systemic Lupus International Collaborating Clinics (SLICC) group at the time of this study consisted of the following members: Graciela Alarcon, University of Alabama at Birmingham, Birmingham, AL, USA; Cynthia Aranow, Columbia Presbyterian Medical Center, New York, NY; Sang-Cheol Bae, Hanyang University, Seoul, Korea; Paul Bacon, University of Birmingham, Birmingham, UK; Jill Buyon, Hospital for Joint Diseases, New York, USA: Ann Clarke, McGill University, Montreal, Canada; Michael Corzillius, Christian Albrechts University, Kiel,

Germany; Vernon Farewell, Cambridge University, Cambridge, UK; Ellen Ginzler, SUNY Downstate, Brooklyn, New York USA; Dafna Gladman, University of Toronto, Toronto, Canada; Mary Anne Dooley, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; Paul Fortin, University of Toronto, Toronto, Canada; Caroline Gordon, University of Birmingham, Birmingham, UK; John Hanly, Dalhousie Unversity, Halifax, Canada; David Isenberg, University College London, London, UK; Kenneth Kalunian, UCSD School of Medicine, La Jolla, CA, USA; Munther Khamashta, St. Thomas Hospital, London, UK; Peter Maddison, North West Wales National Health Service Trust, Bangor, Wales, UK; Susan Manzi, University of Pittsburgh, Pittsburgh, USA; Joan Merrill, Oklahoma Medical Research Foundation, Oklahoma City, USA; Ola Nived, University Hospital, Lund, Sweden; Michelle Petri, Johns Hopkins University, Baltimore, USA; Rosalind Ramsey-Goldman, Northwestern University, Chicago, USA; Jorge Sanchez-Guerrero, Instituto Nacional Natricion, Mexico City, Mexico; Gunnar Sturfelt, University Hospital, Lund, Sweden; Thomas Stoll, Rheumaklinik und Institut fur Physikalische Medizin, Universitatsspital, Zurich, Switzerland; Lori Tucker, British Columbia Children's Hospital, Vancouver, Canada; Murray Urowitz, University of Toronto, Toronto, Canada; Asad Zoma, Stonehouse Hospital, Stonehouse, Lanarkshire, UK.

REFERENCES

1. Isenberg DA, Gladman D. The Systemic Lupus International Collaborating Clinics Group — origins and outcomes. Lupus 2001;10:1-3.
2. Isenberg DA, Ramsey–Goldman R. Assessing patients with lupus: towards a drug responder index. Rheumatology 1999;38:1045-9.
3. Strand V, Gladman D, Isenberg D, Petrie M, Smolen J, Tugwell P. Outcome measures to be used in clinical trials in systemic lupus erythematosus. J Rheumatol 1999;26:490-7.
4. Gladman DD, Goldsmith CH, Urowitz MB, et al. Cross cultural validation of three disease activity indices in systemic lupus erythematosus. J Rheumatol 1992;19:608-11.
5. Gladman DD, Goldsmith CH, Urowitz MB, et al. Sensitivity to change of 3 systemic lupus erythematosus disease activity indices: international validation. J Rheumatol 1994;21:1468-71.
6. Hay E, Bacon P, Gordon C, 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.
7. Stoll T, Stucki G, Malik J, Pyke S, Isenberg DA. Further validation of the BILAG disease activity index in patients with systemic lupus erythematosus. Ann Rheum Dis 1996;55:756-60.
8. Hawker G, Gabriel S, Bombardier C, et al. A reliability study of SLEDAI: a disease activity index for systemic lupus erythematosus. J Rheumatol 1993;20:657-60.
9. Petri M, Hellmann D, Hochberg M. Validity and reliability study of lupus activity measures in the routine clinical setting. J Rheumatol 1992;19:53-9.