A Comparison of Prospective and Retrospective Evaluations of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index for Systemic Lupus Erythematosus

SASHA BERNATSKY, ANN CLARKE, MICHAL ABRAHAMOWICZ, CAROLYN NEVILLE, IGOR KARP, and CHRISTIAN A. PINEAU

ABSTRACT. ***Objective.*** To evaluate the comparability of prospective and retrospective evaluations of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SLICC/ACR DI).

Methods. Consecutive patients meeting ACR criteria for SLE were enrolled prospectively in our cohort. Prospective SLICC/ACR DI scores were collected on the 134 cohort members who were observed in the cohort between 1993–1999. The last available prospective SLICC/ACR DI scores were compared to scores that were retrospectively assigned (for the corresponding time point) from chart review by a research nurse blinded to the prospective values. Intra- and inter-observer agreement was assessed. Kappa coefficients with 95% confidence intervals (CI) were determined.

Results. The kappa correlation coefficient for agreement between prospective versus retrospective total damage scores was 0.68 (95% CI 0.54–0.81). Moderate to very good agreement was also observed with respect to the 12 individual organ systems itemized in this damage index. Substantial agreement was found between assessments done by different research nurses and for repeat assessments done by the same research nurse.

Conclusion. These data suggest good agreement between prospective and retrospective evaluations of the SLICC/ACR DI scores. (J Rheumatol 2005;32:820–3)

Key Indexing Terms:
SYSTEMIC LUPUS ERYTHEMATOSUS DAMAGE INDEX
SLICC/ACR DI PROSPECTIVE RETROSPECTIVE

The Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SLICC/ACR DI) is currently the standard tool used to prospectively measure non-reversible damage in patients with systemic lupus erythematosus (SLE)[1-4]. The SLICC/ACR DI records irreversible damage occurring in patients with SLE regardless of etiology. Damage may arise from active disease, medication, or intercurrent illness. To be included in the score, a damage item must be present for at least 6 months unless otherwise stated. In its initial form, the SLICC/ACR DI records damage in the following organ systems: ocular, neuropsychiatric, renal, pulmonary, cardiovascular, peripheral vascular, gastrointestinal, musculoskeletal, cutaneous, gonadal and endocrine, with one item also for malignancy (see Table 1 for range of scores possible for each system). The total damage score is the sum of these items (i.e., a maximum score of 46). The SLICC/ACR DI is used extensively in outcome assessments in clinical SLE research[5,6].

Since the SLICC/ACR DI was developed only fairly recently, application of this tool to assess damage over time in longstanding SLE cohorts often requires retrospective assessments. However, there are limited data evaluating retrospective construction of scores, compared to the scoring of damage prospectively. In particular, information regarding inter- and intra-observer variability are lacking. Our aim therefore was to evaluate the comparability of prospective and retrospective evaluations of the SLICC/ACR DI in a large group of patients.

From the Division of Clinical Epidemiology, Montreal General Hospital Research Institute, Montreal, Quebec, Canada.

Supported by the Singer Family Fund for Lupus Research and the Canadian Institutes for Health Research (CIHR) Grant #82349. Dr. Bernatsky received fellowship support from the Canadian Arthritis Network, Lupus Canada, and the CIHR. Dr. Clarke is a CIHR Investigator. Dr. Abrahamowicz is the James McGill Professor at McGill University.

S. Bernatsky, MD, FRCPC, MSc; A. Clarke, MD, FRCPC, MSc; M. Abrahamowicz, PhD; C. Neville, BScN, I. Karp, MD; C.A. Pineau, MD, FRCPC.

Address reprint requests to Dr. S. Bernatsky, 1650 Cedar Ave., Room L10-520, Montreal, Quebec, H3G 1A4, Canada.

Accepted for publication December 13, 2004.

MATERIALS AND METHODS

Data collection. The Montreal General Hospital (MGH) lupus cohort enrolls consecutive patients with ACR criteria for SLE[7,8] at the time they present for their first clinic visit.

Prospective SLICC/ACR DI scores have been collected for the cohort members since 1993. In our analyses, the last available SLICC/ACR DI scores (as of 1999) that had been obtained prospectively for each patient were compared to the scores that were retrospectively assigned (for the corresponding time point) from chart review by a research nurse blinded to the prospective values. The prospective scores had been assigned by the attending lupus specialists. The research nurse who completed the retrospective scores was provided with the scoring materials that define each item on the SLICC/ACR DI, but no specific training, other than her clinical background. Medical records reviewed by the research nurse included both charts used by the lupus clinic physicians and hospital medical records.

Data analysis. Kappa coefficients with 95% confidence intervals (CI) were determined for the correlation of prospective and retrospective evaluations with respect to the binary outcome of presence/absence of damage. To interpret the kappa correlation point estimates, we used the Landis and Koch kappa interpretation scale[9].

Kappa Value	Correlation
< 0.00	Poor
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

To calculate inter-observer variability, a random sample of 10 patients was generated and a second research nurse then retrospectively assigned (for the corresponding time point) the SLICC/ACR DI scores from chart review for this sample. The second research nurse was blinded to both the prospective values and the values generated by the first research nurse. The first and second research nurse evaluations were then compared.

To calculate intra-observer variability, the first nurse provided repeat scores on the random sample of 10 patients using the same approach as her initial approach. For this exercise, she was blinded to both the prospective values and the values she had generated earlier. Her first and second evaluations were then compared.

Ethics approval. The research was approved by our institutional ethics review board.

RESULTS

The number of patients who had been part of the cohort at any time during the interval 1993–1999 was 134. Of this number, 93.3% were female. The mean age at cohort entry was 36.3 years [standard deviation (SD) 13.1] and the median duration of SLE at cohort entry was 1.8 years (mean 4.2, SD 5.6).

The median value of the cohort SLICC/ACR DI total scores was somewhat lower (1.0) for the retrospective evaluations compared to the prospective assessments (2.0). The mean for absolute differences between prospective versus retrospective total scores was 1.0 (SD 1.2). The percentage of patients sustaining any damage was estimated to be 66.4% (95% CI 57.7–74.2) and 75.4% (95% CI 67.0–82.2) by the prospective and retrospective methods respectively. The kappa correlation coefficient for agreement for the presence or absence of any damage (comparing prospective versus retrospective total damage scores) was 0.68 (95% CI 0.54–0.81). The point estimate corresponds to very good or substantial agreement according to the Landis and Koch kappa interpretation scale[9]. Moderate to very good agreement was also observed with respect to the 12 individual organ systems itemized in this damage index, according to the references outlined by Landis and Koch above[9]. Kappa coefficients for individual organ systems are presented in Table 1.

Regarding the comparability of the actual scores, there was perfect agreement in 46.3% of the SLICC/ACR DI total scores. Within organ systems, agreement tended to be high, ranging from 79.9% (for neuropsychiatric damage) to 96.3% (for malignancy); detailed results are provided in Table 1.

Regarding inter-observer variability, the kappa coefficient for agreement between the retrospective assessments of the 2 research nurses (for the presence or absence of damage) was 0.61 (95% CI 0.15–1.0). The point estimate corresponds to good or substantial agreement according to the Landis and Koch kappa interpretation scale[9], although the 95% CI is wide, reflecting the modest sample size.

With respect to intra-observer variability, the kappa coefficient for agreement for the initial and second ratings (for the presence or absence of damage) of the first research nurse was 1.0, indicating perfect agreement. However, in 2 cases the actual scores differed. In one case, a neuropsychiatric damage item (cranial or peripheral neuropathy) was missed on the first assessment but noted on the second assessment. For the second case, 2 damage items (a cerebral vascular accident and an avascular necrosis event) were missed on the first assessment but noted on the second assessment.

When looking at the 12 individual organ systems evaluated by the SLICC/ACR DI (Table 1) in our sample, the category with the best correlation between prospective and retrospective assessments was the malignancy damage item (kappa coefficient 0.90, 95% CI 0.77–1.0). The category with the lowest correlation was the premature gonadal failure damage item (kappa coefficient 0.47, 95% CI 0.16–0.78). In all categories except for diabetes, the retrospective scores tended to be slightly lower than the prospective scores.

DISCUSSION

The retrospective construction of the SLICC/ACR DI has been used for some time, but data on the validity of prospective versus retrospective assessments using this tool are sparse. One small study compared prospective scoring of the SLICC/ACR DI to retrospective completion of damage scores from medical records[10]. This showed good correlation of the prospective versus retrospective assessments with respect to the presence or absence of total damage

Table 1. Kappa correlation coefficients for prospective versus retrospective evaluation of the SLICC/ACR DI by individual categories.

SLICC/ACR DI Category (possible score range)	Mean Difference (SD) Prospective vs Retrospective Scores*	Kappa Coefficient** (95% CI)	% Agreement Prospective vs Retrospective Scores***
Ocular	0.06	0.74	89.6
(0–2)	(0.37)	(0.59, 0.89)	
Neuropsychiatric	0.08	0.66	79.9
(0–6)	(0.48)	(0.52, 0.81)	
Renal	0.05	0.72	91.0
(0–3)	(0.26)	(0.53, 0.908)	
Pulmonary	0.06	0.61	92.5
(0–5)	(0.24)	(0.37, 0.85)	
Cardiovascular	0.07	0.72	88.8
(0–6)	(0.33)	(0.55, 0.89)	
Peripheral vascular disease	0.02	0.64	91.8
(0–5)	(0.30)	(0.39, 0.89)	
Gastrointestinal	0.04	0.64	89.6
(0–6)	(0.34)	(0.44, 0.83)	
Musculoskeletal	0.14	0.67	82.8
(0–6)	(0.61)	(0.52, 0.83)	
Cutaneous	0.11	0.57	83.6
(0–3)	(0.38)	(0.41, 0.73)	
Gonadal (premature failure)	0.02	0.47	92.5
(0–1)	(0.25)	(0.16, 0.78)	
Endocrine (diabetes)	–0.01	0.66	96.3
(0–1)	(0.15)	(0.29, 1.0)	
Malignancy	0.01	0.90	96.3
(0–1)	(0.15)	(0.77, 1.0)	

* Mean difference for retrospective versus prospective across all subjects per category. ** Kappa coefficient indicates correlation (between retrospective and prospective assessments) for the presence or absence of damage; this measure takes into account the possibility of agreement by chance alone. *** Represents percentage of cases where there was perfect agreement for actual SLICC/ACR DI score.

(kappa 0.61, 95% CI 0.40–0.82), and there appeared to be good agreement for specific organ system damage as well. The authors concluded that there was no statistically significant difference between direct scoring and assessments from medical records for total or organ system scores, although admittedly their study was done only in a small group of patients (n = 60).

In a much larger patient group, we found good agreement between prospective versus retrospective evaluations of the total SLICC/ACR DI scores. Moderate to very good agreement was also observed with respect to the 12 individual organ systems itemized in this damage index, according to the references outlined by Landis and Koch above[9].

Regarding inter-observer and intra-observer variability, we found good agreement for assessments done by different research nurses, and for repeat assessments done by the same research nurse. Ours is the first study to provide data on inter-observer and intra-observer variability for retrospective scoring of the SLICC/ACR DI.

Differences between prospective and retrospective scoring may arise for several reasons. One difficulty may be related to incomplete medical records for certain damage items; this may explain our observation of slightly lower mean scores for retrospectively completed (versus prospective) damage scores.

We conclude that our data suggest good agreement between prospective and retrospective evaluations of the SLICC/ACR DI scores. Retrospective evaluation of the SLICC/ACR DI should be considered a validated methodology.

REFERENCES

1. Bootsma H, Derksen R, Jaegers S, et al. Usefulness of the SLICC/ACR Damage Index in patients with systemic lupus erythematosus [abstract]. Arthritis Rheum 1997;40 Suppl:S160.
2. Gladman DD, Ginzler EM, Goldsmith CH, 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.
3. Gladman D, Urowitz M, Goldsmith 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.
4. Gladman DD, Ginzler EM, Goldsmith CH, et al. The development and initial validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index for Systemic Lupus Erythematosus. J Rheumatol 1992;19:1820-1.
5. Alarcon GS, Roseman JM, McGwin G, Jr., et al. Systemic lupus

erythematosus in three ethnic groups. XX. Damage as a predictor of further damage. Rheumatology Oxford 2004;43:202-5.
6. Stoll T, Sutcliffe N, Klaghofer R, Isenberg DA. Do present damage and health perception in patients with systemic lupus erythematosus predict extent of future damage?: a prospective study. Ann Rheum Dis 2000;59:832-5.
7. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus [letter]. Arthritis Rheum 1997;40:1725.
8. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF. The 1982 revised criteria for the classification of SLE. Arthritis Rheum 1982;25:1271-7.
9. Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
10. Thumboo J, Lee HY, Fong KY, et al. Accuracy of medical record scoring of the SLICC/ACR damage index for systemic lupus erythematosus. Lupus 2000;9:358-62.