Defining Lupus Cases for Clinical Studies: The Boston Weighted Criteria for the Classification of Systemic Lupus Erythematosus

KAREN H. COSTENBADER, ELIZABETH W. KARLSON, MATTHEW H. LIANG, and LISA A. MANDL

ABSTRACT. Objective. The 1982 American College of Rheumatology (ACR) revised criteria for the classification of systemic lupus erythematosus (SLE), updated in 1997, have become the standard for establishing eligibility of subjects for epidemiologic and clinical lupus studies. These criteria may exclude patients with limited disease, restricting the generalizability of research findings. We developed and evaluated the ability of a weighted classification system to identify a broader spectrum of patients with lupus.

Methods. We constructed the Boston Weighted Criteria system for the classification of SLE, updating that developed in 1984. Using a hospital billing database, we identified 271 patients seen in our rheumatology clinic for possible SLE and reviewed medical records for all ACR criteria and the treating rheumatologist’s diagnosis. We compared both the Boston Criteria and the treating rheumatologist’s diagnosis to the updated 1982 ACR criteria; we also compared the Boston Criteria to the treating rheumatologist’s diagnosis.

Results. The Boston Criteria identified 190/271 patients as having SLE, the rheumatologist’s diagnosis identified 179/271, and the ACR criteria identified 171/271. The Boston Criteria had a sensitivity of 93% and specificity of 69% compared to the ACR criteria, and would identify 7% more patients.

Conclusion. The Boston Criteria identify a larger number of patients compared with the current ACR criteria, while retaining face validity. This reflects the inclusion of patients with objective findings of SLE but less than 4 ACR criteria. Our Boston Criteria system could minimize selection bias and increase the generalizability of clinical SLE studies. (J Rheumatol 2002;29:2545–50)

Key Indexing Terms: LUPUS WEIGHTED CRITERIA CLINICAL STUDIES
based upon the optimal combination of sensitivity and specificity\textsuperscript{10}. In addition, they considered the decreased likelihood of disease given the absence of certain signs or symptoms. Their criteria gave increased weight to cytopenias, renal disease, and discoid skin disease and decreased importance to seizures or psychoses, arthritis, oral ulcers, and photosensitivity. Negative weight was given to a persistently negative antinuclear antibody (ANA). This system was developed using 161 patients seen in an academic rheumatology clinic, 87 of whom had SLE as determined by their rheumatologist, and 73 controls with other rheumatic diseases. It has also been validated in other rheumatic disease populations\textsuperscript{11}.

While Clough and colleagues’ system effectively weights criteria, there remain a number of drawbacks to its more widespread use. It was proposed as a method for diagnosing SLE, not for identifying patients for clinical studies, and has not been tested in that context. It used controls never considered to have SLE, and was therefore not designed to identify, within a group of possible patients with SLE, those most likely to have disease. In addition, it employed criteria, such as alopecia and Raynaud’s phenomenon, which were dropped from the 1982 ACR classification criteria because of low sensitivity and specificity, and it was created before the more recent 1997 addition of antiphospholipid antibodies\textsuperscript{12,13}.

We devised the Boston Weighted Criteria for classification of SLE based on the work of Clough and colleagues, but updated to reflect current knowledge. We included antiphospholipid antibodies, anti-β\textsubscript{2}-glycoprotein antibodies, and renal pathology showing World Health Organization (WHO) class 3–6 glomerulonephritis as evidence of SLE. We calculated the operating characteristics of the Boston Weighted Criteria, using the updated 1982 ACR classification criteria as the gold standard, in a population of patients seen for SLE at an academic rheumatology center. We included only patients seen for possible SLE, rather than a mixed group of rheumatic disease patients, to better approximate the population in which the weighted algorithm would actually be used: potential SLE patients being considered for clinical studies. We also compared how well the diagnostic opinion of “definite lupus” by a rheumatologist compared with the Boston Weighted Criteria and the updated ACR criteria in this patient population.

**MATERIALS AND METHODS**

**Subject identification.** We used a hospital billing database to identify all patients seen at the Brigham and Women’s Hospital outpatient general rheumatology clinic between January 1, 1997, and December 31, 2000, and who were given the International Classification of Disease (ICD-9) code of 710.0 (SLE). Of 600 patients with at least 2 office visits coded as 710.0, we randomly selected 271 for this study. We reviewed at least 2 visits to the clinic in order to ensure adequate clinical data available for review. (Patients who were seen only once were thought more likely to have missing data, not to have had laboratory tests performed at the Brigham and Women’s Hospital laboratories, and to be one-time “rule out SLE” visits, and are not necessarily representative of patients with SLE undergoing routine followup.) Quality control was also maintained by choosing patients with 2 visits, both coded as 710.0, preventing inadvertent inclusion of miscoded patients.

**Data collection.** Hospital internal review board approval was obtained for complete review of all patient records. All data present in the Brigham and Women’s medical record were reviewed, including referral letters and outside laboratories, reports from diagnostic tests, and specialist consultations. Three ACR member physicians (LAN and EWK, board certified rheumatologists, and KHC, second year fellow in rheumatology) performed the chart reviews. For each patient, one of these physicians recorded the presence or absence of each of the 11 ACR criteria, including antinuclear antibodies and lupus anticoagulant (LAC), as well as anti-β\textsubscript{2}-glycoprotein. Renal biopsy results were documented, and the treating rheumatologist’s clinical diagnosis was recorded as “SLE” or “not SLE.” (Statements by the treating rheumatologists of “possible SLE,” “lupus-like syndrome,” or “suspicious for SLE,” were regarded as “not SLE.”)

Criteria were considered positive if documented as present at any time since symptom onset. Criteria were considered negative if documented as such, which time since symptom onset was noted on any medical record. If the records were ambiguous, reviewers met to reach consensus. ANA tests were performed at the Brigham and Women’s Hospital clinical immunology laboratory, using immunofluorescence on HEp-2 cells.

**Construction of the Boston Weighted SLE classification criteria.** Our Boston Weighted Criteria classification system, designed to identify patients with SLE for inclusion in clinical studies, was built upon Clough and colleagues’ algorithm\textsuperscript{10} (Table 1). The system devised by Clough and colleagues used sensitivity and specificity data for each individual SLE manifestation to arrive at a Bayesian weighting scheme. Our system also gives each sign or symptom of SLE a certain point value, and these are then summed to create a total score for each patient.

A number of modifications to Clough and colleagues’ original criteria have been made. We made modifications in content to update the criteria and bring them into alignment with current ACR criteria. For example, we included antiphospholipid antibodies, and excluded Raynaud’s phenomenon, alopecia, and false-positive syphilis serology as evidence of SLE. We have also incorporated the fact that cellular casts and a lower level of proteinuria (0.5 g/day vs 3.5 g/day) are accepted as evidence of renal disease. We also made a number of clinically based modifications. We included anti-β\textsubscript{2}-glycoprotein antibodies, as recent data show that some patients with SLE and thrombosis have anti-β\textsubscript{2}-glycoprotein antibodies with negative LAC and negative anticardiolipin antibodies\textsuperscript{12,13}. A renal biopsy showing WHO class 3–6 nephritis is such strong objective evidence of SLE renal disease that we also included it as a criterion. We did not include WHO class 1 or 2 nephritis, as these nonspecific findings can occur in a number of other diseases including rheumatoid arthritis, idiopathic nephrotic syndrome, and acquired immune deficiency syndrome\textsuperscript{14–18}. To avoid excessive weighting of renal disease, the points given for each of these signs of kidney dysfunction are not additive.

Given the low rate of false positive anti-dsDNA, anti-Sm, and antinuclear antibodies in this population, each SLE associated antibody is given equal weight. In addition, to address the poor specificity of arthritis in previous analyses of the ACR criteria, we adopted a more specific definition, which was objective synovitis documented by a physician, in contrast to the ACR definition of “tenderness, swelling or effusion.” We then increased the weight given to arthritis from 0.1 to 0.5.

We estimated the sensitivity and specificity of our criteria using several different threshold values, and then chose the optimal value for differentiating SLE from non-SLE patients. As our goal was to improve the sensitivity of the updated 1982 ACR criteria in identifying SLE subjects for inclusion in clinical studies, we chose 2.0 points as a cutoff that maximized sensitivity, identifying more patients, while retaining good specificity (Table 2). This is the same cutoff point used by Clough and colleagues.

**Statistical analysis.** All data were analyzed using SAS software (SAS...
RESULTS

Two hundred seventy-one patients with an ICD-9 billing code for SLE (710.0) were identified. The mean age was 45 years, 97% of subjects were female, and > 70% were Caucasian. Patients were seen by a total of 17 different board certified rheumatologists. Missing data precluded a more detailed analysis of race (Table 3). One hundred patients had 3 or fewer ACR criteria for SLE.

One hundred seventy-one patients (63%) were classified as having SLE according to the updated 1982 ACR classification criteria (≥ 4 of 11 criteria), 179 patients (66%) were diagnosed as having “definite SLE” by their treating rheumatologists, and 190 patients (70%) met the Boston Weighted Criteria definition of SLE.

When compared to the ACR criteria, the Boston Weighted Criteria had a sensitivity of 93%, a specificity of 69%, a positive predictive value (PPV) of 84%, and a negative predictive value (NPV) of 85% (Table 4). Agreement between fulfilling ACR criteria and the Boston Weighted Criteria was good, with \( \kappa = 0.65 \). Compared with the treating rheumatologist’s diagnosis of “definite SLE,” the Boston Weighted Criteria had a sensitivity of 88%, specificity 65%, PPV 83%, and NPV 74% (Table 5). Compared to the ACR criteria, the clinical diagnosis of SLE by a
rheumatologist had a sensitivity of 84%, specificity 77%, PPV 88%, and NPV 77% (Table 6). Overall, 98% of those classified as having SLE by the Boston Weighted Criteria had at least 3 of 11 ACR criteria.

Thirty-one patients who fulfilled the Boston Weighted Criteria did not meet the updated ACR criteria (Table 7). All these patients were ANA positive, 87% had 3 or more ACR criteria, and their most common signs and symptoms were cytopenias (81%) and arthritis (48%). Two patients (6%) had documented renal disease and 6 (19%) had anti-dsDNA, anti-Sm, or antiphospholipid antibodies.

Twelve patients did not fulfill the Boston Weighted Criteria but met the ACR criteria (Table 7). These patients were most likely to have findings such as oral ulcerations (92%) and photosensitivity (75%). Four patients in this group were ANA negative, although meeting criteria for classification by the ACR criteria, with mucocutaneous symptoms, cytopenias, serositis, and arthritis. Three of these patients were also anti-Ro and anti-La antibody negative (these antibodies were not checked in the 4th patient). The treating rheumatologists of all these patients were unconvinced of the diagnosis, labeling them “possible lupus” or “lupus-like syndrome.”

Of the 32 patients who did not have “definite SLE” according to their treating rheumatologist, but met the Boston Weighted Criteria (Table 5), all were ANA positive and 28% had anti-dsDNA, anti-Sm, or antiphospholipid antibodies. Fifty percent met classification by the updated 1982 ACR criteria and another 38% had 3 of 11 criteria. The most common manifestations were cytopenias (69%), mucocutaneous symptoms (60%), and arthritis (40%).

Twenty-one patients, however, were diagnosed by their treating rheumatologist as having SLE, but did not meet criteria for inclusion by the Boston Weighted Criteria (Table 5). Of these patients, 19% were not ANA positive and 86% did not have more specific serologies (anti-dsDNA, anti-Sm, and/or antiphospholipid antibodies). Thirty-eight percent had arthritis and a positive ANA and another 38% had the combination of mucocutaneous symptoms and arthritis. No patient in this group had evidence of renal involvement.

**DISCUSSION**

SLE is a clinical syndrome with a diverse and often variable phenotype in the individual patient over time. The ACR criteria for SLE were developed to ensure standardization of patients enrolled in clinical studies, and they have accomplished that goal. However, due to the heterogeneity of initial presentations of SLE, these criteria were not intended to be used for the diagnosis of individual patients. The ACR criteria have been tested in various populations, and 3–69% of patients being treated for SLE do not fulfill these criteria. Therefore, while standardization has occurred, many patients being treated for lupus are systematically excluded from clinical trials.

**Table 5.** Boston Weighted Criteria compared with rheumatologist’s diagnosis. Sensitivity 88%, specificity 65%, PPV 83%, NPV 74%.

<table>
<thead>
<tr>
<th>Rheumatologist’s Diagnosis</th>
<th>SLE Yes</th>
<th>SLE No</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston Weighted Criteria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLE Yes</td>
<td>158</td>
<td>32</td>
<td>190</td>
</tr>
<tr>
<td>SLE No</td>
<td>21</td>
<td>60</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>179</td>
<td>92</td>
<td>271</td>
</tr>
</tbody>
</table>

**Table 6.** Updated ACR 1982 criteria compared with treating rheumatologist’s diagnosis. Sensitivity 84%, specificity 77%, PPV 88%, NPV 71%.

<table>
<thead>
<tr>
<th>Updated 1982 ACR Criteria</th>
<th>SLE Yes</th>
<th>SLE No</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rheumatologist’s Diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLE Yes</td>
<td>150</td>
<td>29</td>
<td>179</td>
</tr>
<tr>
<td>SLE No</td>
<td>21</td>
<td>71</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>171</td>
<td>100</td>
<td>271</td>
</tr>
</tbody>
</table>

**Table 7.** Characteristics of patients meeting Boston Weighted Criteria alone, updated 1982 ACR criteria alone, or both.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>31 Meeting Boston Criteria (not 1982 ACR), n (%)</th>
<th>12 Meeting 1982 ACR (not Boston Criteria), n (%)</th>
<th>159 Meeting Both, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANA positivity</td>
<td>31 (100)</td>
<td>8 (67)</td>
<td>157 (99)</td>
</tr>
<tr>
<td>Cytopenias</td>
<td>25 (81)</td>
<td>2 (17)</td>
<td>128 (80)</td>
</tr>
<tr>
<td>Arthritis*</td>
<td>5 (48)</td>
<td>8 (67)</td>
<td>114 (71)</td>
</tr>
<tr>
<td>Anti-dsDNA/Sm/aPL</td>
<td>6 (19)</td>
<td>0 (0)</td>
<td>97 (61)</td>
</tr>
<tr>
<td>Malar rash</td>
<td>4 (13)</td>
<td>2 (17)</td>
<td>68 (43)</td>
</tr>
<tr>
<td>Photosensitivity</td>
<td>4 (13)</td>
<td>9 (75)</td>
<td>76 (48)</td>
</tr>
<tr>
<td>Renal disease</td>
<td>2 (6)</td>
<td>2 (17)</td>
<td>39 (25)</td>
</tr>
<tr>
<td>Discoid rash</td>
<td>1 (3)</td>
<td>2 (17)</td>
<td>23 (14)</td>
</tr>
<tr>
<td>Neurologic disease</td>
<td>1 (3)</td>
<td>1 (8)</td>
<td>21 (13)</td>
</tr>
<tr>
<td>Oral ulcers</td>
<td>0 (0)</td>
<td>11 (92)</td>
<td>48 (30)</td>
</tr>
<tr>
<td>Serositis</td>
<td>0 (0)</td>
<td>5 (42)</td>
<td>68 (43)</td>
</tr>
</tbody>
</table>

* Objective synovitis documented by a physician. aPL: antiphospholipid antibodies.
Those patients who are diagnosed clinically but who do not meet the ACR criteria are different from those who meet the criteria; they have been described as having “incomplete” or “latent” lupus, or perhaps are seen earlier in their course of disease. For example, Lom-Orta and colleagues compared 31 patients who did and 31 patients who did not fulfill ≥ 4 criteria for the older ARA classification of SLE at the time of clinical diagnosis. Twenty-one of those originally not fulfilling criteria later developed other manifestations and could be classified as having SLE, while the remaining 10 did not, after a mean followup period of 41 months. Calvo-Alen and colleagues studied patients presenting with an “undiifferentiated connective tissue disease” and found an increased likelihood of evolution into SLE if discoid lupus, positive anti-dsDNA, and anti-Smith antibodies were among the presenting features. This suggests these signs may carry more weight in the diagnosis of SLE.

That individual criteria are differentially important in separating out patients with SLE was also was recognized in 1982, when the new ACR criteria gave more weight to positive serologies and less weight to renal disease, compared with the 1971 criteria. Edworthy, et al quantified these differences by applying recursive partitioning to the same group of patients used to derive the 1982 ACR criteria. They found that ANA, anti-dsDNA, serum complement, malar rash, pleurisy, and discoid rash most accurately separated patients with and those without SLE. Using this same group of 339 patients, Manu calculated the receiver operating characteristic curve for each sign and symptom in the 1982 criteria. Using this methodology, positive serologies, malar rash, and renal dysfunction were found to be the most important.

Because the ACR criteria may exclude patients with early or mild disease manifesting with few symptoms, or with severe disease limited to a few organs, various other methods for classifying patients have been suggested. Hughes has proposed the “St. Thomas’ alternative criteria,” a 14 criteria scheme based on his clinical experience, which included teenage growing pains, teenage migraines, recurrent miscarriages, and severe reactions to insect bites as signs of SLE. Schur suggests subdividing patients by the number of positive criteria, and designating them as possible, probable, definite, or classic SLE. Neither approach has been validated or used.

The weighted criteria system of Clough, et al is an appealing alternative. It used Bayes’ theorem to calculate weighted scores from sensitivity and specificity data in SLE patients compared to a rheumatology control population. In this analysis, cytopenias and malar rash in particular were powerful discriminators of SLE from other conditions. Clough and colleagues’ weighted system as a whole yielded a sensitivity of 92% and specificity of 96% based on an expert rheumatologist’s diagnosis, compared to 91% and 89%, respectively for the 1982 ACR criteria.

Our aim was to update the Clough group’s weighted set of criteria and to use it in a different setting: to verify SLE cases among many patients presenting with possible SLE to a rheumatology clinic. The Boston Weighted Criteria increase the spectrum of patients who would be included in a clinical trial or observational study of SLE by including 7% more patients than the ACR criteria. When screening large numbers of potential study subjects, this increase could be consequential. Additional patients are ANA positive and have objective findings such as cytopenias or documented synovitis (Table 7). The Boston Weighted Criteria decrease the importance given to photosensitivity and oral ulcers, minimizing inclusion of patients with mucocutaneous symptoms alone, prone to misclassification.

Dissatisfaction with the existing ACR criteria is evident in that the Boston Weighted Criteria have already been independently validated in another large university based rheumatology practice using a physician diagnosis of SLE as the gold standard. The Boston Weighted Criteria’s operating characteristics were similar in their population, with sensitivity of 90.3% and specificity of 60.4%, compared to 86.4% and 71.9% using the ACR criteria.

One concern with more inclusive criteria is that they could be too sensitive, identifying an overly broad range of patients, at the expense of lowered specificity. However, it is important to remember that the current ACR classification criteria are an arbitrary “gold standard” derived from rheumatologists’ clinical opinions, and thus using operating characteristics as the sole means of evaluating the Boston Weighted Criteria is misleading. Unlike Clough and colleagues, who employed their weighted criteria to distinguish patients with SLE from those with other rheumatic diseases, our aim was to verify those most likely to have SLE among patients already identified as having possible SLE. Our low specificity is not surprising, given our a priori goal of increasing the number of lupus patients eligible for clinical studies. Specificity decreases as patients previously identified as “healthy” are now classified as “diseased.”

The most transparent way of evaluating our criteria is to examine the patients it includes and excludes. The Boston Weighted Criteria have good face validity; they include more patients with objective signs and symptoms of SLE and exclude patients with less specific manifestations (Table 7). Moreover, Alarcon and colleagues have recently reported that the Boston Weighted Criteria at the time of diagnosis are an excellent predictor of SLE organ damage (as measured by the Systemic Lupus International Collaborating Clinics damage index) in their SLE cohort.

We have also shown that the treating rheumatologist’s clinical diagnosis performs well in identifying patients with SLE, with a PPV of 88% compared with both the ACR criteria and the Boston Weighted Criteria. An expert rheumatologist’s opinion could potentially function as a
simple validated screen in future population based SLE clinical studies.

Our algorithm was tested in a hospital based rheumatology clinic, which may limit generalizability. However, many patients are self-referred, and the clinic sees a diverse rheumatic disease population. The Boston Criteria may not perform as well in other clinical settings, such as a dermatology clinic or nephrology clinic, where patients may present with a different spectrum of symptoms; the Boston Criteria should be tested in these alternative lupus populations. In addition, given that we had few patients with disease duration of less than 2 years in our sample population, we have not been able to test whether our criteria perform equally well in early onset SLE.

As with other classification criteria systems, our system is not intended for daily clinical use or office based diagnosis. Rather, it is a fairly simple computer algorithm, designed to screen patients with lupus for inclusion in clinical studies, and would identify a more representative clinical sample than previous criteria.

REFERENCES