Comparison between three systems of classification criteria in juvenile systemic lupus erythematos

Adriana R. Fonseca¹, Maria Ignez C. Gaspar-Elsas², Marcelo G. P. Land¹ and Sheila K. F. de Oliveira¹

Abstract

Objective. The most widely used classification criteria for SLE are those derived and validated in adult patients by the ACR. Alternatives include the Boston weighted (BW) and SLICC criteria. The aim of this study was to compare the performance of BW and SLICC criteria with the 1997 ACR criteria in a JSLE cohort.

Methods. Cases were JSLE patients and controls were patients with other rheumatic diseases attending a tertiary centre in the past 10 years. Data were retrospectively collected to establish the ACR, BW and SLICC criteria fulfilled at the first visit and within the first year of follow-up. A consensus diagnosis of JSLE established by the same group of highly experienced paediatric rheumatologists was chosen as the standard of reference.

Results. One hundred and seventy-three patients were included: 81 JSLE and 92 controls. There was a sharp increase in sensitivity and prevalence of all criteria within the first year of follow-up. The BW criteria had higher sensitivity than the ACR criteria (81.5% vs 58%, \( P < 0.001 \)) at the first visit, but lower specificity in both periods. SLICC criteria had higher sensitivity (82.7% vs 58%, \( P < 0.001 \)) at the first visit, but similar specificity in both periods.

Conclusion. In this JSLE population, the SLICC criteria performed best in terms of sensitivity and accuracy at the first visit and within the first year of follow-up.

Key words: systemic lupus erythematos, childhood, adolescence, classification criteria.

Introduction

SLE is a multisystem autoimmune disease with a wide spectrum of clinical patterns that affects all ages and ethnicities [1]. JSLE represents ~20% of SLE cases [2] and displays a higher frequency of atypical manifestations as well as more severe presentation and evolution, leading to significantly higher morbidity and mortality than that reported for adult-onset disease [3–7].

Regardless of the age of disease onset, the diagnosis relies upon a combination of clinical and laboratory findings. The classification criteria of the ACR were developed to standardize the definition of SLE and have become the standard to establish the eligibility of patients for clinical trials and epidemiological studies [8, 9]. However, some limitations of the ACR classification raise concerns, including inclusion of the most typical features only, possible duplication of highly correlated cutaneous manifestations, omission of many neurological manifestations, inadequate quantification of urine protein by dipstick, omission of low complement levels, lack of standardization for autoantibody detection and classification as SLE for patients who do not satisfy the immunological disorder criterion [10–12].

The ACR classification criteria may have further limitations in JSLE, since they were primarily developed and validated in adults with little validation in paediatric populations.

Alternative methods of SLE classification have been developed, such as the Boston weighted (BW) and SLICC criteria. In 2002, Costenbader et al. [13] formulated the BW criteria, showing a sensitivity of 93%, specificity of
69% and identification of 7% more patients than the 1997 ACR criteria.

In 2012, Petri and Magder [14] published a study about the derivation and validation of SLICC criteria for SLE classification. Under this proposal the patient must meet at least 4 of 17 criteria, including at least one clinical and one immunological criterion or have documented LN with ANA and/or anti-dsDNA. When compared with the 1997 ACR criteria, the SLICC criteria resulted in fewer misclassifications, higher sensitivity and lower specificity [14].

The primary aim of this study was to evaluate and compare the ability of the 1997 ACR criteria, BW criteria and SLICC criteria to identify patients with JSLE both at the first visit and within the first year of follow-up.

**Patients and methods**

**Inclusion criteria**

The study included children and adolescents with JSLE and other rheumatic diseases attending the Pediatric Rheumatology Unit of the Instituto de Puericultura e Pediatria Martagão Gesteira/Universidade Federal do Rio de Janeiro in the past 10 years. All baseline and evolutionary information (physical examination, laboratory parameters, imaging) regarding each patient is routinely discussed and re-evaluated at each visit by all members of the team of attending pediatric rheumatologists, which establishes the diagnosis based on continuous follow-up of all clinical scenarios, in accordance with internationally accepted criteria [15–20].

**Exclusion criteria**

Patients were excluded if they were followed up for < 1 year (regarded as an indicator of continuous follow-up).

**Data collection**

A set of precisely defined variables to be retrieved from the medical records was defined. Data collection was retrospective (extracted by M.G.P.L. and M.I.G.E and confirmed by A.R.F. and S.K.F.O.). If the records were ambiguous, the authors met to reach a consensus. Each 1997 ACR, BW and SLICC criterion was recorded as being either present or absent at baseline (the first visit) and within the first year of follow-up. Baseline data were those obtained from clinical history, physical examination and laboratory tests requested by the attending paediatric rheumatologists at the first visit.

**Criteria definitions**

Definitions of malar rash, discoid rash, photosensitivity, oral ulcers, arthritis, serositis, thrombocytopenia, leucopenia and lymphopenia were those provided by the 1997 ACR criteria [9]. Renal and neurological involvement definitions were those provided by the 1997 ACR or SLICC criteria [9, 14]. Haemolytic anaemia was defined as a decrease of >3 g/dl in haemoglobin levels with a reticulocyte count of >5%.

All autoantibody tests were performed at a single reference laboratory. ANA and anti-dsDNA tests were determined by indirect immunofluorescence [with human cell epithelioma (HEp-2) cells and *Crithidia luciliae* as substrate, respectively] and were defined as positive if staining reactivity was seen at >1:160 and >1:10 serum dilution, respectively. The anti-Smith antigen (anti-Sm) was determined by ELISA and was considered positive if it was above the laboratory reference range. *aCL* isotypes IgM and IgG were determined by ELISA with a cut-off value of 20 MPL or GPL, respectively. Lupus anticoagulant (LAC) was determined by the dilute Russell’s viper venom test with confirmatory testing.

The following judgements and adaptations were needed to apply the three criteria sets: (i) renal biopsy was not considered as missing data if it was not performed in patients without clinical or laboratory signs of renal involvement and (ii) anti-β2-glycoprotein I antibodies were measured only in those patients with APS-suggestive manifestations.

**BW criteria**

This classification system assigns separate point values for each SLE finding, summed to generate a total score. In addition, negative ANAs are subtracted from the total score. Biopsy-proven SLE nephritis [World Health Organization (WHO) classes 3–6] was included. Each autoantibody is an individual criterion [13]. Those authors estimated the sensitivity and specificity of weighted criteria using several different threshold values and then chose the optimal value for differentiating SLE from non-SLE patients.

**SLICC criteria**

The main changes proposed in relation to the 1997 ACR criteria were a redefinition of the four cutaneous criteria, inclusion of the urine protein:creatinine ratio, expansion of neurological criteria, separation of cytopenias and autoantibodies in individual criteria and the inclusion of alopecia and hypocomplementaemia [14]. To be classified as SLE, the patient must meet at least 4 of 17 criteria, including at least 1 clinical and 1 immunological criterion or documented LN with ANA and/or anti-dsDNA [14].

**Statistical analysis**

Statistical analyses were performed with SPSS 20.0 for Windows (IBM, Armonk, NY, USA). Differences between quantitative variables were analysed by the Mann–Whitney test. Differences between proportions were analysed by Fisher’s exact test.

Since most rheumatology affections do not have pathognomonic features and have evolutionary manifestations, we decided that it would be more reassuring and productive to use as our standard of reference the diagnosis consolidated during continuous follow-up of all patients through discussion and consensus among our team of paediatric rheumatologists. Receiver-operating characteristic (ROC) analysis was used to select the optimal cut-off point for the weighted criteria total score, considering the original methodology adopted by Clough et al. [21] and subsequently refined by Costenbader et al. [13].
McNemar’s test was applied to assess differences in sensitivity and specificity between the ACR and BW systems and the ACR and SLICC systems. The difference was regarded as statistically significant when the P-value was < 0.05. The $\kappa$ statistic was used to quantify the chance-adjusted degree of agreement between the ACR and BW criteria and also between the ACR and SLICC criteria. The agreement was considered moderate for $\kappa$ values between 0.41 and 0.60, substantial for those between 0.61 and 0.80 and excellent when $>0.81$ [22]. Accuracy was measured through the area under the ROC curve.

The study was approved by the institutional review board of the Instituto de Puericultura e Pedia-tra Martagão Gesteira and it conforms to standards currently applied nationwide in Brazil. Nothing was done that was outside of routine clinical practice, so patient consent was not required.

**Results**

**Sample description**

Some of the charts, 6.9% for the case group and 12% for the control group, had missing information about some of the defined variables. Thus this study evaluated 81 JSLE cases and 92 controls [37 systemic JIA (SJIA), 33 JDM, 5 juvenile SS (JSS), 3 mixed CTD (MCTD), 3 SS, 3 primary APS (PAPS) and 8 primary vasculitides], for a total of 173 patients. There was no statistically significant difference between the groups regarding time to clinical diagnosis, regardless of whether the time was counted from symptom onset ($P = 0.067$) or from enrolment in the outpatient paediatric rheumatology clinic ($P = 0.614$). Table 1 summarizes general characteristics of the two groups.

**1997 ACR criteria**

Among JSLE cases, 46 (58%) patients met the ACR criteria at first visit and 28 (35%) met the criteria within the first year of follow-up. There were three remaining patients diagnosed as JSLE by their treating rheumatologists that did not fulfil the ACR criteria even in a mean follow-up time of 9.5 years. The first patient had malar rash, oral ulcers, alopecia, cutaneous vasculitis and low complement at first visit and positive ANA only in the first year. The second patient had alopecia, proteinuria and biopsy-proven class IV nephritis and anti-dsDNA, but persistently negative ANA. The third patient had arthritis, autoimmune haemolytic anaemia, thrombocytopenia and ANA at first visit. Six controls were misclassified at first visit: four JDM, one MCTD and one primary vasculitis. Two additional controls were also classified as JSLE in the first year: one JDM and one MCTD.

The most commonly observed ACR criteria in JSLE cases, both at first visit and within 1 year of follow-up, were ANA (90.1% and 93.8%), arthritis (71.6% and 84%), haematological (58% and 76.5%), immunological (54.3% and 75.3%) and malar rash (32.1% and 55.6%), as shown in Table 2. ANA titres ranged from 1:80 to 1:5120, with a median of 1:320. Three patients (3.7%) with JSLE had persistently negative ANA, all with class IV nephritis, anti-dsDNA and malar rash.

Arthritis, ANA and malar rash were the most observed ACR criteria in the control group. ANA titres ranged from 1:40 to 1:5120, with a median of 1:80. At first visit, the mean number of ACR criteria was 3.94 (S.D. 1.65) for cases and 1.74 (S.D. 1.1) for controls ($P < 0.001$). Within the first year of follow-up this average was 5.58 (S.D. 1.6) for cases and 2.15 (S.D. 1.1) for controls ($P < 0.001$).

**BW criteria**

As our goal was to identify JSLE patients earlier (high sensitivity), minimizing misdiagnosis (high specificity), we selected a total score of 1.0 at the first visit and a total score of 2.0 within the first year of follow-up.

Considering a cut-off point of 1.0 at first visit, 35 (38%) controls were misclassified as JSLE (6 SJIA, 16 JDM, 4 JSS, 3 MCTD, 2 PAPS, 1 SS and 3 primary vasculitis). Moreover, this cut-off classified earlier than the ACR criteria, with 26 JSLE patients, including 2 incomplete ones by the ACR criteria. Considering a cut-off point of 2.0 within the first year of follow-up, 19 (20.6%) controls had an inadequate classification (1 SJIA, 13 JDM, 1 JSS, 2 MCTD, 1 SS and 1 primary vasculitis).

**SLICC criteria**

Among JSLE cases, 67 (83%) patients met the SLICC criteria at first visit and 11 (13.6%) within the first year of follow-up. No patients with SJIA or JSS were misclassified. Six controls were misclassified at first visit (4 JDM, 1 MCTD and 1 primary vasculitis). Five additional controls were also classified as JSLE in the first year (2 JDM, 1 MCTD and 2 primary vasculitis).

The most frequent SLICC criteria in JSLE, both at the first visit and within 1 year, were ANA (90.1% and 93.8%), arthritis (71.6% and 84%), alopecia (48.1% and 49.4%), hypocomplementaemia (45.7% and 50.6%) and anti-dsDNA (35.8% and 53.1%), acute cutaneous lupus (32.1% and 65.4%) and haemolytic anaemia (32.1% and

---

**Table 1** General characteristics of case and control groups

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex ratio, female:male</td>
<td>5.8:1</td>
<td>2.1</td>
</tr>
<tr>
<td>Age at onset, mean (S.D.), years</td>
<td>9.9 (2.4)</td>
<td>6.6 (3.5)</td>
</tr>
<tr>
<td>Age at diagnosis, mean (S.D.), years</td>
<td>10.4 (2.2)</td>
<td>7.4 (3.5)</td>
</tr>
<tr>
<td>Time to diagnosis from first visit, mean (S.D.), months</td>
<td>1.3 (3.4)</td>
<td>1.5 (2.4)</td>
</tr>
<tr>
<td>Follow-up time, mean (S.D.), years</td>
<td>7.5 (2.8)</td>
<td>6.3 (3.1)</td>
</tr>
</tbody>
</table>

Differences between quantitative variables were analysed by the Mann-Whitney test. Those variables were evaluated in 81 JSLE cases and 92 controls. Differences between proportions were analysed by Fisher’s exact test. The difference was regarded as statistically significant when the P-value was < 0.05. Mean (S.D.) reported considering normal distribution data.
Differences between proportions were analysed by Fisher’s exact test. The difference was regarded as statistically significant when the P-value was <0.05. Anti-Sm: anti-Smith antigen. Prevalence of 1997 ACR and SLICC individual criteria were evaluated in 81 JSLE cases and in 92 controls.

46.9%). Considering the expanded neurological criterion, there was an increase in its prevalence at first visit and within the first year of follow-up to 5% and 20%, respectively.

At first visit, JSLE cases fulfilled a mean of 5.01 (s.d. 1.15) SLICC criteria and controls fulfilled 1.8 (s.d. 1.14) (P < 0.001). At first visit, cases fulfilled a mean of 6.89 (s.d. 2.09) SLICC criteria and controls fulfilled 2.25 (s.d. 1.15) (P < 0.001). Table 3 summarizes the performance of the 1997 ACR, BW and SLICC criteria in terms of sensitivity, specificity, predictive values and accuracy.

Comparison of 1997 ACR and BW criteria performance

The weighted criteria had greater sensitivity (81.5% vs 58%, P < 0.001) at first visit, but similar sensitivity within the first year of follow-up (P = 0.063). However, specificity was lower at first visit (82.6% vs 93.4%, P = 0.001) and within the first year (89.1% vs 91.3%, P = 0.001). If the original cut-off of 2 points was chosen at first visit, a 61.7% sensitivity and a 93.3% specificity would result. The application of weighted criteria resulted in a greater number of misclassifications both at first visit (35 vs 6, P < 0.001) and within the first year of follow-up (19 vs 8, P = 0.001). The agreement between the ACR and BW criteria was moderate at first visit (k = 0.6) and substantial (k = 0.76) within the first year of follow-up.

Comparison of 1997 ACR and SLICC criteria performance

The SLICC criteria had greater sensitivity at first visit (82.7% vs 58%, P < 0.001) and within the first year of follow-up (96.3% vs 91.3%, P = 0.125). There was no statistically significant difference regarding specificity (P = 1 and P = 0.250) and the number of misdiagnoses (P = 1 and P = 0.8), either at first visit or within the first year of follow-up. Accuracy was higher for the SLICC criteria in both periods. The agreement between ACR and SLICC criteria was substantial (k = 0.74) at first visit and excellent within the first year of follow-up (k = 0.91).

Two-by-two tables showing comparisons of 1997 ACR criteria (see supplementary Tables S1 and S2), BW criteria (see supplementary Tables S3–S5) and SLICC criteria (see supplementary Tables S6 and S7) with the standard of reference, at baseline and within the first year of follow-up are available at Rheumatology Online.
Discussion

One of the strengths of this research was the assessment and comparison between three classification systems for SLE (1997 ACR, BW and SLICC) applied exclusively to JSLE patients in different moments of follow-up.

Most previous studies that evaluated the clinical and laboratory findings in JSLE did not include different time points, with the notable exception of Hiraki et al. [3], who analysed the manifestations at diagnosis (time to fulfil ACR criteria), within the first year of follow-up and at the last visit [mean follow-up 3.5 years (s.d. 3)].

In our study, the findings of ANA, arthritis, haematological, immunological and malar rash as the most frequently observed ACR criteria were consistent with those reported by Hiraki et al. [3], Gomez et al. [5], Hoffman et al. [6] and Ferraz et al. [23]. However, in disagreement with the above-cited research, haemolytic anaemia and leucopenia were the most common haematological disorders within the first year of follow-up. In the studies of Hiraki et al. [3], Gomez et al. [5] and Hoffman et al. [6], lymphopenia was the most commonly detected cytopenia.

When the immunological disorder criteria were analysed further, anti-dsDNA followed by anti-Sm were the most common, as published in other series. Hiraki et al. [3] and Gomez et al. [5] found a higher frequency of aCL (40–52%) and LAC (13–34%) than in our study. This may be due to the different cut-off points adopted and laboratory interchangeability of APLs.

The occurrence of alopecia and hypocomplementaemia in about half of the cases is in agreement with the findings of Hoffman et al. [6], supporting the inclusion of these variables in the classification system as suggested by the SLICC criteria. Those findings were rarely observed in the control group ($P < 0.001$).

The performance of the 1982 ACR criteria in JSLE was evaluated in two retrospective studies, one of them being a multicentre Brazilian study [23, 24]. The population in these studies consisted of JSLE cases and controls with other rheumatic diseases (mainly JIA and rheumatic fever). The treating physician diagnosis was the standard of reference. The conclusion was that the 1982 ACR criteria might be applied for the classification of JSLE, obtaining 96% sensitivity and 100% specificity [23]. Our study evaluated the latest ACR version, had a different control group composition and evaluated the performance of classification systems at two moments. Our research also adopted different antigenic substrates and higher cut-offs for ANA (1:160 vs 1:20 in Ferraz et al.'s study [23]). Regarding the performance of weighted criteria, Costenbader et al. [13] showed sensitivity of 93% and specificity of 69% in a population of adults with SLE, considering a total score of 2.0. Our study was the first to evaluate the performance of the BW criteria in JSLE. The cut-off for the total score that obtained the best combination of sensitivity and specificity was different in the two periods of follow-up: 1.0 at first visit and 2.0 within the first year. The weighted criteria had a greater initial sensitivity (81.5% vs 78%, $P < 0.001$), but similar sensitivity within 1 year ($P = 0.063$), compared with the ACR criteria.

Specificity was lower at the first visit (82.6% vs 93.4%, $P = 0.001$) and within the first year (89.1% vs 91.3%, $P = 0.001$). However, when using a cut-off of 2 points at baseline, the sensitivity and specificity of the BW criteria were virtually identical to those of the 1997 ACR criteria. The application of weighted criteria also resulted in a statistically higher number of misdiagnoses, both at first visit and within the first year of follow-up ($P < 0.001$).

The SLICC criteria had higher sensitivity at first visit (82.7% vs 58%, $P < 0.001$) and within the first year of follow-up (96.3% vs 91.3%, $P = 0.125$) compared with the 1997 ACR criteria, as demonstrated by Petri and Magder [14] in adult patients. There was no statistically significant difference in specificity at both observation periods and in the number of misdiagnoses. Some reasons that explain the discrepancy between the lower specificity of the SLICC vs ACR criteria observed in adults and the comparable specificity seen at both time points in our study may be a different control group composition and the definitions of some variables (i.e. haemolytic anaemia, different cut-off values for autoantibodies).
Some reasons that may explain the low sensitivity of ACR criteria at first visit compared with the BW and SLICC criteria are inclusion of autoantibodies as a single criterion, duplication and equal weighting for highly correlated terms relating to cutaneous lupus and the assignment of equal importance to manifestations that differ in their relevance to diagnosis.

Although the SLICC criteria performed best in terms of sensitivity at the first visit, it was still only 82.7%, which means that, on average, one in five cases of JSLE would be missed. Considering the peculiarities of JSLE (higher frequency of atypical presentation and constitutional manifestations) and taking into account that these three sets of criteria were developed and validated in adults, modifications should be considered to increase early diagnostic sensitivity in JSLE. Some suggestions include the appropriateness of some variables for children and adolescents, such as haematological parameters and thresholds for proteinuria levels according to weight or body surface area, as well as inclusion of a composite criterion involving constitutional manifestations.

Several intrinsic limitations and potential biases are present in this study. First, interpretation of the results is limited by the retrospective chart review. Some measures were taken to minimize this limitation, such as the extraction and collection of data by more than one author, as well as adjustments for some variables. Furthermore, those charts with missing information were considered as losses that were in acceptable proportions. A second and intrinsic limitation is the achievement of a totally objective diagnosis to be considered the standard of reference, so the treating physician diagnosis is still adopted by studies validating classification criteria for SLE. In view of the difficulties of diagnosing a complex and variable condition such as JSLE in the real world, we decided to use as our standard of reference the diagnosis consoli- dated during continuous follow-up of all patients through continuous discussion and re-evaluation of clinical scenarios by a group of attending paediatric rheumatologists. Third, we did not include children without a clear diagnosis, which may give an optimistic assessment of any criteria. Finally, the cumulative characteristics of the criteria systems might have influenced the sensitivity and specificity according to follow-up duration.

Conclusion

Our results revealed that there was a sharp increase in the number of patients classified as JSLE by all classification systems in the first year of follow-up. Although the BW criteria had a better sensitivity compared with the 1997 ACR criteria, they are more complex and more time-consuming. Besides they retain many of the limitations of the ACR criteria, especially the non-inclusion of low complement and other neuropsychiatric and cutaneous manifestations.

On the other hand, application of the SLICC criteria resulted in higher sensitivity and accuracy in both follow-up times, maintaining simplicity and includes a wider variety of neuropsychiatric, cutaneous and laboratory findings observed in SLE.

Therefore one should consider using the SLICC criteria as a replacement for the ACR criteria to determine the eligibility of patients with SLE and JSLE for studies on pathogenesis, diagnosis, treatment and prognosis.

Rheumatology key messages

- Boston weighted criteria performed better than the ACR criteria only in terms of sensitivity.
- The application of SLICC criteria resulted in higher sensitivity and accuracy in both observation periods.

Acknowledgements

We acknowledge Drs Fernanda Mariz and Blanca Bica for helpful comments.

Funding: Maria Ignez C. Gaspar-Elsas was supported by CNPq Research Productivity Fellowships.

Disclosure statement: The authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at Rheumatology Online.

References


22 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
